

## Автоматизированные информационные экономические системы

(Курс лекций)

### Информационные процессы в экономике

*Информатизация* – организованный процесс создания оптимальных условий для удовлетворения информационных потребностей граждан, органов государственной власти и производственных организаций на основе формирования и использования информационных ресурсов.

*Информационный ресурс* – это документы и массивы документов в информационных системах (библиотеках, архивах, фондах, банках данных и других информационных системах), т. е. документированные знания. Информационный ресурс для каждой страны является стратегическим ресурсом, аналогичным по важности запасам энергии, ископаемых, сырья и прочим источникам.

Основными задачами информатизации общества являются:

- модернизация информационно-телекоммуникационной инфраструктуры;
- развитие информационных, телекоммуникационных технологий;
- эффективное формирование и использование национальных информационных ресурсов (ИР) и обеспечение широкого, свободного доступа к ним;
- обеспечение граждан общественно значимой информацией и развитие независимых средств массовой информации;
- создание необходимой нормативно-правовой базы построения информационного общества.

Количество, качество и доступность информационных ресурсов уже сейчас во многом определяет уровень развития страны, ее статус в мировом сообществе и бесспорно станут решающим показателем этого статуса в первые десятилетия XXI века.

Современный этап информатизации связан с использованием персональной электронно-вычислительной техники, систем телекоммуникаций, создания сетей ЭВМ. Возрастает потребность в разработке и применении эффективных решений в сфере *информационной индустрии*. Она занимается производством технических и программных средств, информационных технологий для получения новых знаний.

На определенном этапе развития информационной индустрии рождается *информационное общество*, в котором большинство работающих занято производством, хранением, переработкой и реализацией информации,

т. е. творческим трудом, направленным на развитие интеллекта и получение знаний. Создается единое, не разделенное национальными границами информационное сообщество людей.

Переход от индустриального общества к информационному проявляется в следующем:

- все большее перераспределение трудовых ресурсов из сферы материального производства в сферу информационных услуг;
- на смену металлических или бумажных денег приходит электронная информация, заменяя традиционные формы расчетов системой электронных платежей, электронной коммерцией и др.;
- информатизация охватывает все социально значимые области жизни человека: экономику, политику, культуру и пр.

Формирование информационного общества опирается на новейшие информационные, телекоммуникационные технологии и технологии связи. Именно новые технологии привели к бурному распространению глобальных информационных сетей, открывающих принципиально новые возможности международного информационного обмена. Формирование информационного общества концептуально и практически означает формирование мирового информационного пространства.

Информационный ресурс – это документы и массивы документов в информационных системах (библиотеках, архивах, фондах, банках данных и других информационных системах), т. е. документированные знания. Информационный ресурс для каждой страны является стратегическим ресурсом, аналогичным по важности запасам энергии, ископаемых, сырья и прочим источникам.

На первый план выдвигается приоритет информационного ресурса по сравнению с другими материальными ресурсами, развитие информационной индустрии. Вместе с тем наблюдаются и негативные тенденции: увеличивается проникновение в частную жизнь людей, сокращается занятость, ухудшается здоровье пользователей информационных технологий и т. п.

Умение целенаправленно работать с информацией и использовать для ее получения, передачи, обработки и использования компьютерную информационную технологию, современные технические средства и методы, значит обладать *информационной культурой*. Это необходимое профессиональное качество бухгалтеров, финансистов, менеджеров, банковских служащих и т. д.

Информационные ресурсы являются исходной базой для создания *информационных продуктов*. Последние являются результатом интеллектуальной деятельности человека и распространяются с помощью услуг.

Посредством *информационных услуг* осуществляется получение и предоставление в распоряжение пользователя информационных продуктов. Юридической основой этой операции должен быть договор между двумя сторонами – поставщиком и потребителем, а источником информационных услуг – *базы данных*. Они могут существовать в компьютерном и некомпьютерном вариантах, в виде библиографических и небиблиографических взаимосвязанных данных, основанных на общих правилах описания, хранения и манипулирования данными.

Главная тенденция в развитии информации на современном этапе состоит в совершенствовании электронной техники в сочетании с достижениями в области искусственного интеллекта и средств коммуникации.

Слияние компьютеров и средств коммуникации породило *«инфокоммуникационные технологии»*, охватывающие все социально-значимые области человеческой жизни, в том числе:

- электронную коммерцию;
- электронные платежи, платежи в банковских, клиентских, налоговых и других расчетах;
- дистанционное обучение и выполнение других работ.

С развитием инфокоммуникационных технологий и сервиса глобальных, региональных и локальных сетей стал быстро развиваться новый сектор экономики, получивший название сетевого. Сетевая экономика определяется как «среда, в которой любая компания или индивид, находящиеся в любой точке экономической системы, могут контактировать с любой другой компанией или индивидом по поводу совместной работы, торговли или просто для удовольствия» [26].

Достижения в области искусственного интеллекта в сочетании с инфокоммуникационными технологиями породили и прогресс в области экспертных систем, нейросетей и нейрокомпьютеров, информационной безопасности и других областях.

### **Рынок информационных ресурсов, продуктов и услуг, его государственное регулирование**

*Информационный рынок* – система экономических, правовых и организационных отношений по торговле товарами, созданными

информационной индустрией. Как всякий рынок, он характеризуется определенной номенклатурой продуктов и услуг, ценами, спросом и предложением, поставщиками и потребителями.

Но в отличие от торговли обычным товаром, информационные ресурсы, услуги и продукты могут копироваться в неограниченном количестве (например, пакеты автоматизации бухгалтерского учета "1С–Бухгалтерия", "Парус" и др.). Исключение составляют информационные ресурсы, продукты и услуги, которые не могут быть товаром и попадают под действие ФЗ «О государственной тайне», ФЗ «О профессиональной тайне», статьи Уголовного кодекса (УК) №272 «О несанкционированном доступе», статьи УК №273 «Вредоносные программы» и пр.

Составляющими рынка информационных ресурсов, продуктов и услуг являются: аппаратно-программные средства, соответствующие технологии переработки информации, товары информационной индустрии, поставщики и покупатели, соответствующая нормативно-правовая база и справочно-навигационные средства.

### **Автоматизированные информационные системы (АИС), их классификация, структура и этапы развития**

Теоретические аспекты построения АИС исходят из положений кибернетики – науки об управлении в объектах живой и неживой природы и информатики – науки о преобразовании информации с использованием технических средств. Основопологающие понятия кибернетики – это:

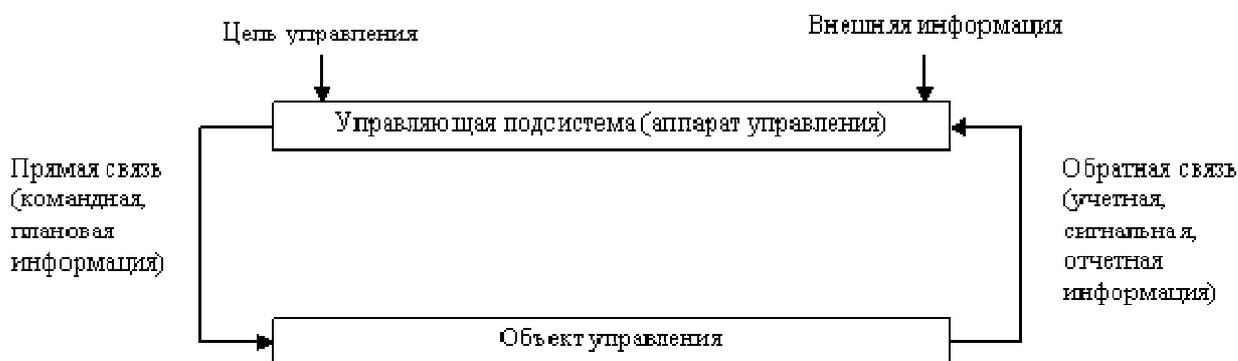
- система;
- системный подход;
- информация;
- прямая и обратная связь.

*Система* – это совокупность взаимосвязанных элементов, подчиненных единой цели. Признаками системы являются следующие:

1. Элементы системы взаимосвязаны и взаимодействуют в рамках системы.
2. Каждый элемент системы может в свою очередь рассматриваться как самостоятельная система (подсистема), но он выполняет только часть функций системы.
3. Система как целое выполняет определенную функцию, которая не может быть сведена к простой совокупности функций подсистем.

4. Подсистемы могут взаимодействовать как между собой, так и с внешней средой и изменять при этом свое содержание или внутреннее строение.

*Система управления* реализует функции управления и состоит из таких подсистем, как прогнозирование, планирование, учет, анализ, контроль и регулирование. Общая схема системы управления представлена на рис.



#### Кибернетическая модель системы управления

Любой системе управления экономическим объектом соответствует *экономическая информационная система (ЭИС)* или совокупность внутренних и внешних потоков прямой и обратной информационной связи экономического объекта, методов, средств, специалистов, участвующих в процессе обработки информации и выработке управляющих решений [8].

Согласно ФЗ «Об информации, информатизации и защите информации», принятому в феврале 1995 г., *информационная система* – это организационно упорядоченная совокупность документов (массивов документов) и информационных технологий, в том числе с использованием средств вычислительной техники и связи, реализующих информационные процессы (процессы сбора, обработки, накопления, хранения, поиска и распространения информации).

Различают ручные и автоматизированные ЭИС. К автоматизированным информационным системам (АИС) относится упорядоченная совокупность информации, экономико-математических методов и моделей, технических и программных средств, организованных на базе новой информационной технологии в решении экономических задач и информационного обслуживания специалистов служб управления.

По объекту управления различают АИС:

- банков;
- финансовых органов;
- фирм или предприятий;
- статистики;

- налоговых органов;
- органов страхования;
- таможенных органов и т. д.

По отраслевому признаку выделяют АИС:

- в промышленности;
- в строительстве;
- на транспорте;
- в торговле и пр.

По виду взаимодействия с объектом управления можно выделить:

- автоматизированные системы управления (АСУ) техническими средствами (АСУ ТС);
- АСУ персоналом (АСУП);
- АСУ организационно-технологическими процессами (АСУ ОТП);
- интегрированные АИС;
- корпоративные АИС;
- АИС научных исследований;
- обучающие АИС.

В АСУ ТС объектом управления являются технические средства (например, станки), отсюда взаимодействие с ЭВМ осуществляется исключительно по каналам связи.

В АСУП объект управления – организационные процессы с участием персонала, а обмен информации осуществляется как по каналам связи, так и документов.

АСУ ОТП являются гибридными системами, объектами управления в которых являются как технологические, так и организационные процессы.

*Интегрированные АИС* предназначены для автоматизации всех функций управления фирмой.

*Корпоративные АИС* используются для автоматизации всех функций управления фирмой или корпорацией, имеющей территориальную разобщенность между подразделениями, филиалами, отделениями, офисами и т. д.

*АИС научных исследований* обеспечивают решение научно-исследовательских задач на базе экономико-математических методов и моделей.

*Обучающие АИС* используются для подготовки специалистов в системе образования, при переподготовке и повышении квалификации работников различных отраслей экономики.

В зависимости от особенностей автоматизированной профессиональной деятельности можно выделить следующие АИС [20]:

- системы поддержки принятия решений (СППР);
- автоматизированные информационные вычислительные системы (АИВС);
- система автоматизации проектирования (САПР);
- проблемно-ориентированные имитационные системы (ПОИС);
- автоматизированные системы обучения (АСО);
- автоматизированные информационно-справочные системы (АИСС);
- автоматизированные системы управления.

*Системы поддержки принятия решений (СППР)* являются достаточно новым классом АИС, теория создания которых в настоящее время интенсивно развивается. СППР называется АИС, предназначенная для предоставления аналитической информации менеджерам высокого уровня для принятия управленческих решений.

*Автоматизированные информационно-вычислительные системы (АИВС)* предназначены для решения сложных в математическом отношении задач, требующих больших объемов самой разнообразной информации. Таким образом, видом деятельности, автоматизируемом АИВС, является проведение различных (сложных и "объемных") расчетов. Эти системы используются для обеспечения научных исследований и разработок, а также как подсистемы АСУ и СППР в тех случаях, когда выработка управленческих решений должна опираться на сложные вычисления.

*Система автоматизации проектирования (САПР)* – это автоматизированная информационная система, предназначенная для автоматизации деятельности подразделений проектной организации или коллектива специалистов в процессе разработки проектов изделий на основе применения единой информационной базы, математических и графических моделей, автоматизированных проектных и конструкторских процедур. САПР является одной из систем интегральной автоматизации производства, обеспечивающих реализацию автоматизированного цикла создания нового изделия от предпроектных научных исследований до выпуска серийного образца.

В области экономики САПР могут использоваться при проектировании экономических информационных систем и их элементов. Кроме того, технология САПР может обеспечить создание автоматизированной системы отображения обстановки на экране в процессе ведения экономических операций в ходе деловых игр различных типов.

*Проблемно-ориентированные имитационные системы (ПОИС)* предназначены для автоматизации разработки имитационных моделей в некоторой предметной области. Например, если в качестве предметной области взять развитие автомобилестроения, то любая модель, создаваемая в этой предметной области, может включать стандартные блоки, моделирующие деятельность предприятий, поставляющих комплектующие; собственно сборочные производства; сбыт, обслуживание и ремонт автомобилей; рекламу и др. Эти стандартные блоки могут строиться с различной детализацией моделируемых процессов и различной оперативностью расчетов. Пользователь, работая с ПОИС, сообщает ей, какая модель ему нужна (т. е. что необходимо учесть при моделировании и с какой степенью точности), а ПОИС автоматически формирует имитационную модель, необходимую пользователю.

*Автоматизированные системы обучения (АСО)* предназначены для автоматизации подготовки специалистов с участием или без участия преподавателя и обеспечивающих обучение, подготовку учебных курсов, управление процессом обучения и оценку его результатов. Основными видами АСО являются автоматизированные Системы программного обучения (АСПО), системы обеспечения деловых игр (АСОДИ), тренажеры и тренажерные комплексы (ТиТК).

Автоматизированная система обеспечения деловых игр предназначена для подготовки и проведения деловых игр, сущность которых заключается в имитации принятия должностными лицами индивидуальных и групповых решений в различных проблемных ситуациях путем игры по заданным правилам.

Автоматизированные системы дистанционного обучения предназначены для подготовки студентов, школьников, специалистов при их удалении от образовательных центров.

*Автоматизированные информационно-справочные системы (АИСС)* – это автоматизированные информационные системы, предназначенные для сбора, хранения, поиска и выдачи в требуемом виде потребителям информации справочного характера.

В зависимости от характера работы с информацией различают следующие виды АИСС:

- автоматизированные архивы (АА);
- автоматизированные системы делопроизводства (АСД);
- автоматизированные справочники (АС) и картотеки (АК);
- автоматизированные системы ведения электронных карт местности (АСВЭКМ) и др.

*Автоматизированная система управления* представляет собой автоматизированную систему, предназначенную для автоматизации всех или большинства задач управления, решаемых коллективным органом управления (министерством, финансовым органом, налоговой службой, страховой компанией и др.).

По уровню в системе управления различают АИС:

- локальные (конкретного коммерческого банка);
- отраслевые и региональные;
- общегосударственные (АИС при Минфине, АИС статистики и пр.).

По отношению к производству выделяются следующие АИС:

- производственные (АСУП, АСУ ТС и др.);
- внепроизводственные (системы, созданные в УВД, в прокуратуре и др.).

В настоящее время существует два подхода к построению АИС: позадачный и процессный.

Первый подход, исторически появившийся ранее, базируется на функциональной модели управления предприятием, отражающей выполнение сотрудниками своих должностных обязанностей согласно целям и функциям управления. В структуре таких ИС выделяют: функциональную часть, отражающую цели и задачи управления, и обеспечивающую часть, содержащую средства решения задач.

В соответствии с данным подходом информационная система создается как инструмент, предназначенный для автоматизации функций управления, типовыми среди которых являются прогнозирование, планирование, учет, анализ, регулирование. Для реализации одной функции или ее части создаются функциональные подсистемы, например, планирования, учета, финансов, оперативного управления и т. д. Функциональные подсистемы состоят из комплексов функциональных задач – например, подсистема административного управления может состоять из следующих комплексов: хозяйственное планирование, управление кадрами и т. д.

Обеспечивающая часть ИС состоит из информационного, программного, технического, организационного обеспечения и т. д.

При позадачном подходе к управлению ИС есть не что иное, как множество связанных между собой АРМ, обслуживающих различные уровни управления. Структура сети АРМ отражает в большинстве случаев организационную структуру управления предприятия.

Однако позадачный подход в управлении обладает рядом недостатков, среди которых в первую очередь можно назвать: во-первых, размытость, а

иногда и отсутствие ответственности на различных стадиях производства и реализации продукции за конечный результат управления; во-вторых, сложность увязки всех функций производства и управления в единую технологию и т. д.

В настоящее время постепенно развивается новый подход к управлению – процессный. Этот подход ориентирует на управление не отдельными структурными подразделениями предприятия, выполняющими свои функциональные обязанности, а сквозными бизнес-процессами.

Под *бизнес-процессом* понимается совокупность действий экономического объекта, выполнение которых позволяет получить конечный результат (товар или услугу).

Бизнес-процессы состоят из бизнес - операций, выполняемых с помощью АРМ. Каждый бизнес-процесс характеризуется определенным во времени началом и концом, интерфейсом с другими процессами. Например, бизнес-процесс «Производство» обслуживают специалисты из производственного, финансового и других отделов.

Какой бы подход к построению АИС не был выбран, функциональная и обеспечивающая части не исчезают.

Содержательную компоненту АИС составляют функциональные подсистемы, включающие комплексы взаимосвязанных задач, реализующих определенную функцию системы управления. При этом под задачей понимается преобразование исходных данных в информацию предназначенную для получения конкретного результирующего документа. (например, составление расчетно-платежной ведомости по учету заработной платы, получение оборотной ведомости по движению материалов и т. д.).

Состав функциональных подсистем во многом определяется особенностями экономической системы, ее отраслевой принадлежностью, формой собственности, размером, характером деятельности предприятия.

Функциональные подсистемы АИС могут строиться по различным принципам:

- предметному;
- функциональному;
- проблемному;
- смешанному (предметно-функциональному).

Так, по смешанному принципу в АИС промышленного предприятия выделяют подсистемы:

1. Стратегическое управление (финансовый менеджмент, анализ финансово-хозяйственной деятельности, маркетинг, управление проектами, управление документооборотом и др).

2. Логистика (управление материальными потоками и сбытом готовой продукции).
3. Бухгалтерский учет (учет денежных средств, основных средств, учет материальных ценностей и пр.).
4. Управление персоналом (создание нормативно-справочной информации, планирование затрат по персоналу, ведение базы данных кадрового состава и др.).
5. Управление производством (технологическая подготовка производства, технико-экономическое планирование, учет затрат на производство, оперативное управление производством).

### **Проектирование АИС**

Проектирование АИС – процесс создания и внедрения проектов комплексного решения экономических задач по новой информационной технологии. Сюда включается детальная разработка проектных решений, внедрение, опытная эксплуатация и анализ проектных решений.

Качественное проектирование и внедрение являются основной предпосылкой эффективного функционирования системы при постоянном совершенствовании ее обеспечивающих и функциональных составляющих. Цель всех этих работ состоит не только в автоматизации манипулирования информационными потоками, но и в совершенствовании управления и организации деятельности экономического объекта. Поэтому первый руководитель должен иметь представление об имеющихся на рынке технических и программных средствах, тенденциях в их развитии, основных принципах проектирования ИС. В каждом подразделении организации должен быть назначен сотрудник, ответственный за проектирование и внедрение ИС, который собирает нужную информацию, подбирает технику и программные средства, ведет обучение персонала, руководит внедрением и анализом функционирования информационных систем.

Современные предприятия (корпорации) относятся к классу больших динамических систем с характерной многопрофильной деятельностью и большим числом кооперативных связей с партнерами. При этом возрастает динамичность бизнес-процессов, связанных с изменяющимися потребностями и сильной конкуренцией. Управление бизнес-процессами предполагает рассмотрение всех материальных, финансовых, трудовых и информационных потоков с системных позиций, т. е. во взаимодействии.

Достижения в области АИС и ИТ дают возможность проведения инжиниринга и реинжиниринга бизнес-процессов .

Целью реинжиниринга бизнес-процессов (РБП) является системная реорганизация управления бизнес процессами на базе новых информационных технологий.

*Реинжиниринг бизнес-процессов* – это создание новых, более эффективных бизнес-процессов без учета предшествующего развития (все начинается заново, подвергается сомнению, проявляется творческое начало во всех действиях).

Технология реинжиниринга основана на том, что в процессе управления пользователь активно использует современные информационные технологии для обучения, стратегического и тактического планирования, анализа возможных путей перестройки и улучшения бизнес-процессов, управления изменениями, реализацию проектов и др.

Работа с информацией и информационная культура в целом является одним из важнейших компонентов попыток компании *управлять изменениями*. Есть три принципиальные причины, в силу которых менеджер сегодня должен заботиться об информационной культуре своей компании.

*Во-первых*, она все больше и больше становится важнейшей частью общей организационной культуры. Все больше компаний понимают необходимость преобразований, ориентированных на удовлетворение ожиданий потребителя. Чтобы сегодня влиять на будущее, нужно представлять себе, на что оно будет похоже. А для этого нужно работать с разнообразнейшей деловой, рыночной, политической, технологической и социальной информацией.

*Во-вторых*, информационные технологии делают возможным создание в компаниях компьютерных сетей, с помощью которых идет общение между менеджерами, - но важно знать, как люди используют эту информацию. Само по себе создание такой сети со всеми ее рабочими станциями и мультимедийными возможностями не гарантирует того, что информация будет использоваться более разумно и более эффективно.

*В-третьих*, для разных функциональных служб, подразделений и рабочих групп информационная культура различна, а это означает различие подходов к процессам осознания, сбора, организации, обработки, распространения и использования информации. Поэтому многие менеджеры согласятся с тем, что корпоративная информационная культура важна для выработки стратегии и осуществления перемен.

*Инжиниринг бизнес-процессов* включает в себя реинжиниринг бизнес-процессов, проводимый с определенной периодичностью, например один раз в 5 - 7 лет, и последующее непрерывное улучшение.

*Обратный инжиниринг* предполагает исследование и анализ функционирующих на предприятии бизнес-процессов. Цель этапа заключается в проведении диагностики «узких мест» в организации существующих бизнес-процессов и формулировании направлений их реорганизации. Задача обратного инжиниринга упрощается, если на предприятии имеется документация о функционирующих процессах после предыдущей реорганизации.

На этапе обратного инжиниринга постановка задач реорганизации бизнес-процессов (РБП) уточняется, сформулированные на этапе идентификации бизнес-процессов в общем виде цели РБП могут быть скорректированы по результатам исследования существующей системы организации бизнес-процессов.

Для оценки эффективности существующих бизнес-процессов используются прежде всего методы и средства для выявления:

- наиболее трудоемких и затратных функций;
- функций, не вносящих вклад в образование прибыли;
- функций с низким коэффициентом использования ресурсов.

Массовое проектирование ИС базируется на использовании нормативно-правовой базы (федеральных законах, ГОСТах и пр.) и таких основополагающих принципах как эффективность, контроль, совместимость, гибкость, системность, развитие, стандартизация и унификация.

Принципы массового проектирования ИС следующие:

*Принцип эффективности* заключается в том, что выгоды от новой автоматизированной системы должны быть равными или больше расходов на нее.

*Принцип контроля* требует, чтобы информационная система обладала механизмами для защиты данных фирмы, ее данные были бы достаточно надежны для принятия управленческих решений.

*Принцип совместимости* предполагает, что проект системы будет совместим с организационной структурой предприятия и особенностями системы управления им.

*Принцип гибкости* требует от системы возможности расширения без проведения больших изменений. Например, в новую автоматизированную систему учета можно легко ввести новые счета в план счетов, если он изменился, новые хозяйственные операции и др.

*Принципы системности* позволяют исследовать объект как единое целое во взаимосвязи всех его элементов. На базе системного подхода применяется и метод моделирования, позволяющий моделировать изучаемые процессы вначале для анализа, а затем и синтеза создаваемых систем.

*Принцип развития* заключается в непрерывном обновлении функциональных и обеспечивающих составляющих системы.

*Принцип стандартизации и унификации* предполагает использование уже накопленного опыта в проектировании и внедрении ИС посредством программирования типовых элементов, что позволяет сократить затраты на создание ИС.

### **Методы проектирования АИС**

Внедрение АИС для управления экономическими объектами включает предпроектное обследование, на результатах которого разрабатывается проект автоматизации управления экономическим объектом: предприятием, территориальным образованием и т.д.

Типовое проектное решение (ТПР) АИС включает следующие разделы:

- состав задач, решение которых позволяет:
  - Осуществлять оперативное управление объектом;
  - Осуществлять планирование деятельности;
  - Проводить анализ экономической деятельности;
- комплекс программных средств и технология для решения этих задач;
- комплекс программно - аппаратных средств включающий:
  - состав и конфигурацию компьютеров – серверы и рабочие станции;
  - состав и характеристики сетевого оборудования;
  - операционные системы(ОС), системы управления базами данных (СУБД)
- информационное обеспечение, включающее исходные документы, справочные материалы, регламент их создания, изменения, ввода, результирующие документы и порядок их использования;
- организационное обеспечение, в котором определяется состав и организационная структура персонала, квалификационные требования к персоналу;

- юридическое обеспечение, позволяющее правильно приобретать программно – аппаратные средства и строить информационные взаимоотношения с внешней средой;
- технология внедрения указанных средств, с тем чтобы обеспечить работу АИС в соответствии с технологией предусмотренной ТПР.

После того, как становится ясным перечень решаемых задач, необходимо выбрать программные средства, которые обеспечат решение этих задач. Существует два подхода к проектированию АИС:

- Разработка уникального проекта АИС
- Использование типового проектного решения.

В первом случае привлекается штат проектировщиков и программистов или заключается договор с специализированной ИТ компанией. Плюсами такого подхода является то, что разрабатываются программные средства позволяющие решать поставленные задачи с максимальной эффективностью. Кроме того, необходимо помнить, что разработка программного комплекса для управления экономическим объектом, это не только совокупность программ решающих задачи управления, но и технологии управления информационной деятельностью объекта. А технология подразумевает определение состава и функциональности АРМ, порядок действий персонала во времени, сетевые средства взаимодействия АРМ, обеспечение хранения, защиты и распределенного доступа к данным и т.д. Поэтому использование этого способа создания ИС упирается в большие финансовые расходы. Их могут позволить себе только крупные компании, которые надеются с помощью ИС обеспечить такую эффективность деятельности, которая в течении приемлемого срока позволит оправдать понесённые расходы. Другой причиной разработки оригинальных программных средств является защита данных от несанкционированного доступа: в уникальных программных комплексах легче построить защиту.

В большинстве случаев используются типовые проектные решения(ТПР). В силу своей тиражированности стоимость программных продуктов ТПР значительно меньше. Однако технологии автоматизированного управления, предоставляемые ими, стандартизированы. С одной стороны, использование ТПР позволяет внедрять их на предприятиях различного типа и размеров. Это облегчает обмен данными между предприятиями и согласование их деятельности. С другой стороны, внедрение ТПР требует подстройки процесса управления объектом под

возможности программного комплекса. Во многих случаях настройка типового программного комплекса на структуру и задачи объекта управления происходит с помощью параметров. Параметры позволяют в сравнительно небольшом диапазоне менять характеристики решаемых задач. Поэтому использование типовых программных комплексов ограничено. Предприятия, внедрившие ТПР, через некоторое время перестают устраивать его возможности. Устранение этой коллизии решается следующим образом. К типовому программному комплексу добавляются средства программирования, которые позволяют значительно расширять функциональность ТПР. Так, например, было сделано в семействе ТПК «1С». Так к программным комплексам «1С: Бухгалтерия», «1С: Склад» и так далее, был добавлен язык программирования с соответствующими средствами разработки программ, с помощью которого можно создавать новые документы и средства их обработки. Таким образом, модификация исходных ТПК требует штата квалифицированных программистов. Существует два способа привлечения программистов для выполнения подобной работы: принять их в штат компании, которая эксплуатирует ТПР и заключить договор с специализированной организацией. Выбор способа привлечения определяется экономическими расчетами. В большинстве случаев первый способ дешевле и возникающие корректировки выполняются более оперативно, так как программисты работающие в штате компании получают задание от менеджеров и вносят требуемые изменения в ТПР. Второй способ дороже и время отклика зависит от места нахождения специализированной фирмы. Кроме того требуется время на согласование изменений. Плюсом второго подхода является то, что работы, как правило, выполняются на более профессиональном уровне. На этом этапе решаются так же большинство юридических вопросов, связанных с приобретением ТПР и его эксплуатацией.

На этапе разработки проекта по выбору аппаратных средств решаются следующие задачи. Исходя из технологии обусловленной ТПР, устанавливается состав, конфигурация и размещение рабочих станций. Выбираются сервера определённой мощности, способные надёжно хранить и резервировать большие объёмы данных. Определяется состав и размещение сетевого оборудования. После этого выполняются работы по размещению, установке и настройке аппаратных средств, а так же разворачиванию на них типовых программных комплексов.

Всю дальнейшую работу невозможно выполнить без подготовки организационного обеспечения. В рамках выполнения этой работы

необходимо подобрать персонал соответствующей квалификации, обучить его, обеспечить материальное и моральное стимулирование.

Следующим по порядку, но не по времени, является этап внедрения информационного обеспечения. По сути, он начинается сразу после выбора и приобретения ТПР и включает. После установки ТПК на аппаратных средствах они вводятся с помощью программных модулей входящих в ТПР в базу данных(БД). В большинстве АИС данные хранятся в базах данных под управлением СУБД. Самым сложным на этом этапе является ввод в БД текущей информации, так как сначала нужно ввести огромный массив ранее накопленных данных. Параллельно нужно вводить текущие данные. Без полной информации в БД о работе предприятия невозможно решать задачи управления с достаточной степенью достоверности.

После ввода АИС в эксплуатацию наступает период опытной эксплуатации, в результате которой проверяется правильность выбранных организационных форм, выявляются недостатки программного обеспечения, эффективность конфигурации аппаратных средств, полнота информационного обеспечения. Как правило, на их устранение уходит время около года. Потом наступает время стационарной эксплуатации АИС. В процессе эксплуатации накапливаются пожелания по совершенствованию системы, возникают предложения по расширению анализа данных, изменению технологии обработки и использования данных. Эта работа возлагается на администратора данных, и является подготовкой к следующему реинжинирингу.

### **Технологии автоматического проектирования АИС**

В области автоматизации проектирования ИС в последние годы сформировалось новое направление CASE-технологии (COMPUTER Aided System / Soft Wore Engineering). Это совокупность методов анализа, проектирования, разработки и сопровождения АИС с максимальной автоматизацией процессов разработки и функционирования систем. Организационно CASE-индустрия включает компании трех типов:

- разработчиков средств анализа и проектирования;
- разработчиков специальных средств с ориентацией на узкие предметные области;
- обучающие, информирующие и консалтинговые фирмы, оказывающие соответствующие услуги при использовании CASE-пакетов.

Компании, предоставляющие такие услуги, получили название системных интеграторов. Следует отметить, что этот термин имеет два понятия. Согласно первому, под термином «системный интегратор»

понимаются как компании, специализирующиеся на сетевых и телекоммуникационных решениях (сетевые интеграторы), имеющие в свою очередь, сеть своих продавцов, так и компании – программные интеграторы. Существует и другая трактовка понятия «системный интегратор», которая закрепляет за компанией комплексное решение задач заказчика при проектировании АИС. При этом имеется в виду, что заказчик полностью доверяет детальную проработку и реализацию проекта системному интегратору, оставляя за собой лишь определение исходных данных и задач, которые должна решать реализуемая ИС.

Фирмы-интеграторы создают, как правило, дилерскую сеть представительств в ряде городов России и в странах СНГ. При этом компании осуществляют техническую и информационную поддержку своих дилеров, проводя совместные семинары и презентации, регулярно рассылая им информационно-рекламные материалы о новых продуктах и перспективных технологиях, осуществляют совместное участие в крупных региональных проектах.

Другим вариантом организации системной интеграции является выполнение проектов от консалтинга до создания прикладной системы, т. е. заказчику сдается готовая к эксплуатации информационная система «под ключ» и допускается привлечение организаций и квалифицированных специалистов в качестве партнеров для реализации.

CASE-технологии проектирования АИС ориентируются на архитектуру готовых программных изделий. Это обусловлено следующими факторами:

- быстрее создавать и внедрять ИС при меньших затратах;
- обеспечить единый простой интерфейс;
- сократить усилия на обслуживание существующих приложений при их адаптации к постоянным изменениям в программно-технической среде.

CASE-технология включает вопросы определения требований к системе и создание проекта на глобальном уровне, так чтобы он наиболее полно отвечал требованиям с учетом заданных экономических и технологических ограничений. CASE-технология содержит средства поддержки всех основных этапах проектирования и внедрения АИС, при этом на этапе анализа целей создания системы обычно используется концепция диаграмм потоков данных. Причем особенно уделяется внимание связям между данными. В результате между входными и выходными данными устанавливаются парные связи

Ядром системы является база данных проекта - репозиторий (словарь данных). Он представляет собой специализированную базу данных,

предназначенную для отображения состояния проектируемой АИС в каждый момент времени.

Репозиторий содержит информацию об объектах проектируемой АИС и взаимосвязях между ними, все подсистемы обмениваются данными с ним. В репозитории хранятся описания следующих объектов:

- проектировщиков и их прав доступа к различным компонентам системы;
- организационных структур экономического объекта;
- диаграмм и пр.

Преимущества CASE-технологии по сравнению с традиционной технологией оригинального проектирования сводятся к следующему:

- улучшение качества разрабатываемого программного приложения за счет средств автоматического контроля и генерации;
- возможность повторного использования компонентов разработки;
- снижение времени создания системы, что позволяет на ранних стадиях проектирования получить прототип будущей системы и оценить его;
- освобождение разработчиков от рутинной работы по документированию проекта, так как при этом используется встроенный документатор;
- возможность коллективной разработки АИС в режиме реального времени.

## **Информационное обеспечение АИС**

Аппаратное и программное обеспечение образуют механизм для автоматизированной обработки информации. Для использования этого механизма необходимо собрать, подготовить в требуемом виде данные, а после обработки надлежащим образом использовать полученные результаты.

Информационное обеспечение ИТ представляет собой среду циркуляции экономической информации в процессе управления экономическими объектами и включает:

1. Совокупность форм, методов и технологий отражающих структуру и организацию подготовки информации для автоматизированной обработки
2. Характеристики и структуру информации используемой для обмена между пользователями в процессе управления экономическими объектами

3. Методы и технологии использования данных, полученных в результате автоматизированной обработки, для принятия управленческих решений

### *Классификаторы*

Информация об экономических объектах имеет выраженную структурированность. При её регистрации и обработке возникает необходимость в идентификации как самих объектов, так и их структурных составляющих. Это необходимо для их однозначного поиска в информационной базе. Например, наименование одного и того же предприятия, например ООО "НПО Витязь", может быть занесено в базу разными способами:

- ООО "НПО витязь"
- НПО "Витязь"
- "Витязь"
- Витязь

С точки зрения компьютера это совершенно разные слова. Поэтому для обозначения лучше использовать цифровой код. Такой, например, как номер телефона. Но лучше в качестве универсального номера для обозначения предприятия использовать уникальный цифровой код не связанный с его реквизитами. Совокупность таких кодов называется классификатором. Классификаторы повсеместно используются для обозначения экономических объектов. Например, почта России традиционно использовала внутреннюю систему кодирования предприятий связи (6 цифр) - известный всем индекс на почтовом конверте. Свой код предприятия введен Министерством по налогам и сборам. Все налогоплательщики пронумерованы в компьютерах МНС РФ специальным кодом - так называемым идентификационным номером налогоплательщика, ИНН.

Сейчас в России принята система ЕСКК (Единой системы классификации и кодирования технико-экономической и социальной информации), частью которой является ОКВЭД (общероссийский классификатор видов экономической деятельности). ОКВЭД построен на основе Статистической классификации видов экономической деятельности в Европейском экономическом сообществе (Statistical classification of economic activities in the European Community).

Существуют международные системы кодирования банков. Например, международный банковский идентификационный код ВИС (Bank Identification

Code) длиной 8-11 знаков, который присваивается всем банкам - участникам международной системы межбанковских переводов S.W.I.F.T.

Таким образом, существует множество банковских, налоговых и других кодов предприятия, которые частично дублируют друг друга.

Очевидно, что в перспективе с учетом роста международного сотрудничества и развития электронной коммерции, которая представляет собой планетарную систему, должен быть введен единый международный код предприятий.

Единая система кодирования предприятий должна отвечать следующим требованиям.

- Система должна быть международной.
- Система должна быть универсальной (т.е. осуществлять возможность кодирования юридических и физических лиц вне зависимости от принадлежности к каким-либо отраслям или сферам деятельности).

Главное требование к коду - его уникальность. Код, присвоенный какому-либо предприятию, не должен повториться нигде.

Примером подобной системы на сегодня является глобальный идентификационный номер GLN EAN/UCC

Стандартом ассоциации EAN International (General EAN/UCC Specifications) введен в действие единый международный Глобальный идентификационный номер EAN/UCC Global Location Number (GLN).

GLN представляет собой 13-разрядный цифровой код (номер), предназначенный для точного и краткого обозначения:

1. юридических лиц, т.е. зарегистрированных предприятий и организаций;
2. функциональных подразделений, т.е. департаментов, управлений, отделов и т.п. внутри организации;
3. физических объектов, т.е. номеров помещений, складских ворот и т.д.

Указанные системы относятся к *глобальным*, так как описывают широкий класс идентифицируемых объектов – на уровне группы предприятий и выше. Часто приходится кодировать экономические объекты внутри предприятия. При этом использование глобальных классификаторов для подобных целей не совсем удобно из-за избыточной сложности и длины кодов. Чаще используют внутренние системы кодов, которые представляют собой либо усечённые глобальные кодировки, либо самостоятельные внутренние классификаторы. Эти классификаторы относятся к группе *локальных*.

При составлении классификаторов используются определённые правила кодирования, которые называют системами кодирования.

### **Системы кодирования**

Под кодированием понимается присвоение информации условных обозначений. Кодирование информации преследует целый ряд целей, в том числе получение более экономного её изображения и большей наглядности, облегчение передачи, хранения, обработки и использования. Совокупность условных обозначений, соответствующих единицам информации, называется кодом. Код строится по определённым системам. *Системой кодирования* называется строго определённый порядок присвоения условных обозначений единицам информации. Для кодирования информации используются цифры, буквы, кодовые знаки. При автоматизированной обработке ЭИ к кодам предъявляются следующие требования:

- должен обеспечивать возможность точной идентификации объекта;
- должен обеспечивать возможность выделения определённых классификационных признаков объекта;
- должен иметь по возможности меньшее число разрядов;
- код одного и того же объекта в различных задачах должен быть одним и тем же;
- система кодирования по возможности должна соответствовать глобальным классификаторам.

Наиболее распространены следующие системы кодирования:

- Порядковая
- Серийная
- Разрядная
- Коды повторений.

*Порядковая система кодирования* предусматривает замену наименований предметов порядковыми номерами в возрастающем порядке. Новым наименованиям обычно присваиваются коды, следующие за последними порядковыми номерами.

Пример. Кодирование списка студентов в группе.

Преимущества: простота и рациональная техника построения, т.к. в ней используются только необходимое количество номеров.

Недостатки: отсутствие возможности расширения наименований без нарушения принятой классификации.

*Серийная система кодирования* предусматривает присвоение серии номеров признакам старшего порядка и самостоятельных кодов признакам

младшего порядка. В отведённых сериях каждой группы признаков старшего порядка оставляются резервные номера для кодирования новых наименований.

Пример. Для кодирования списков студентов нескольких групп можно использовать серийную систему кодирования: каждой группе выделить серию номеров, например, от 1 до 30, от 31 до 60 и т.д.

Преимущества:

- логическое построение,
- небольшая точность,
- возможность резерва номеров для признаков старшего и младшего порядков.

Недостатки:

- сложность построения,
- трудоёмкость при запоминании кодов,
- при получении итогов по сериям необходима предварительная сортировка.

*Десятичная или разрядная система кодирования* характеризуется закреплением одного или нескольких цифровых разрядов за каждым группировочным признаком внутри многопризначной номенклатуры, т.е. закреплением за признаками номеров, кратных десяти. Каждый разряд или несколько цифровых разрядов определяет какой-то признак классификации номенклатуры.

Пример. Структура номера зачётной книжки включает код факультета - один десятичный разряд, номер группы - один десятичный разряд и два десятичных разряда порядковый номер студента в группе.

Преимущества:

- сокращение значности кода за счёт повторения кодов подгрупп и признаков младшего порядка,
- наиболее полно отвечают требованиям учёта:
- стойкость, простота построения, чёткость выделения признаков.
- позволяют более эффективно использовать автоматизированные методы обработки.

Недостатки:

- невозможность рационального использования отведённых номеров для кодирования признаков,
- многозначность, а след., громоздкость и сложность построения.

*Коды повторений* повторяют цифровые обозначения, присвоенных наименованиям номенклатуры или их характеристики.

Пример. Наименование групп для использования в расписании. Так название группы ЭУб-мн-11-3 содержит условные обозначения, которые, с одной стороны, понятны преподавателям и студентам, с другой, однозначно идентифицируют каждую группу университета. В приведенном коде группы первые два знака указывают факультет, следующий знак форма обучения – бакалавр, магистр; затем через тире мнемоническое обозначение специальности – менеджмент; год поступления и номер группы.

#### Преимущества:

- более понятны, быстрее запоминаются и не требуют трудоёмкой работы по кодированию и декодированию,
- рациональна, т.к. уменьшает ошибки в первичных документах.

Недостатки: неудобны при использовании в автоматизированной обработке.

Используются при обозначении даты, разрядов работы, разрядов рабочих, материалов, инструментов, сорта товаров и др.

*Смешанная или комбинированная система кодирования* основана на использовании различных систем, изложенных выше. Смешанные коды применяются для кодирования многозначных номенклатур, содержащих несколько различных признаков. Они широко используются для кодирования производственных затрат, когда балансовые счета, субсчета и статьи расходов обозначаются кодами повторения, а изделия и заказы кодируются с помощью серийной, десятичной или другой системы кодирования. При смешанной системе наиболее рационально используются преимущества каждой из применяемых систем кодирования.

## **Манипулирование экономической информацией**

Информационное обеспечение ИТ решает задачи подготовки данных к обработке с помощью программных средств и использования полученных результатов для эффективного управления экономическим объектом.

Процесс решения этих задач включает следующие этапы:

- Сбор и формализация экономической информации (*рецепция*)
- Организация хранения данных
- Обмен данными для решения задач управления экономическими объектами
- Использование результатов обработки данных для управления экономическими объектами.

Совокупность всех этих действий называется манипулированием экономической информацией.

### *Организация сбора экономической информации*

Начальная стадия информационного процесса – рецепция. В различных информационных системах рецепция выражается в таких конкретных процессах, как сбор и формализация исходной информации.

Процесс рецепции начинается на границе, отделяющей информационную систему от внешнего мира. Здесь, на границе, сигнал внешнего мира преобразуется в форму, удобную для дальнейшей обработки. В экономических системах основной формой представления данных во внешней по отношению к информационной системе является документ. Однако документы предназначенные для ручной обработки, как правило, плохо формализованы, то есть содержат нечёткие, плохо согласованные между собой данные. Поэтому их трудно приспособить к организации автоматизированной обработки данных. Для того, чтобы формализовать документ нужно прежде всего классифицировать входящие в него данные. Поэтому рецепцию всегда можно рассматривать как процесс классификации.

### *Технологии перевода бумажных документов в электронные*

В отличие от бумажных документов, электронные могут обрабатываться более эффективно (тиражироваться, рассылаться, храниться и т.п.). В настоящее время активно развиваются технологии перевода бумажных документов в электронную форму с целью реализации электронного документооборота. Остановимся подробнее на применяемых технологиях и используемой терминологии.

На первом этапе перевода документа в электронную форму производится его сканирование и создается электронная копия документа в виде изображения. Изображение, полученное в результате сканирования, также называют образом документа. Сканирование является начальным этапом любой системы автоматизированного ввода документов.

Обычно процесс сканирования - это промежуточная стадия получения электронного документа. Для того чтобы можно было редактировать документ, осуществлять поиск по нему или использовать его фрагменты при подготовке новых документов и т.д., необходимо перевести полученный образ в текстовый документ, понятный офисным программам. Поэтому следующая задача заключается в распознавании отсканированных документов. Для этого необходим специальный инструмент, способный

перевести изображение в текстовый редактируемый электронный документ. Такие инструменты существуют, их общее название - программы оптического распознавания символов (optical character recognition, OCR). С помощью OCR-программы компьютер сможет "прочитать" на отсканированной странице текст, отделив его от иллюстраций и прочих элементов оформления, найти таблицы и "разобраться" в их содержимом. А затем скомпоновать все это заново, воссоздав внешний вид страницы.

С точки зрения перевода документов в электронный вид (ввода документов в компьютер) их условно делят на формализованные, неформализованные и специальные.

Формализованные документы - это документы, в которых заранее определена форма: расположение обязательных полей, в которые заносятся данные. Например, бланки, накладные, анкеты, картотеки и т.д.

Неформализованные документы - это документы произвольной формы: договоры, письма и т.д.

К специализированным относятся такие документы как, например, карты и отпечатки пальцев.

Перевод каждого из перечисленных видов документов имеет свою специфику. Если вводятся фотографии, то достаточно электронного изображения, если документ содержит текст, его необходимо распознать, если это форматированный текст с рисунками, то нужно не только распознать текст, но и восстановить формат документа, скорее всего, сам документ вообще не нужен, важна только содержащаяся в нем информация.

## Ввод формализованных документов

В подавляющем большинстве случаев электронные документы хранятся в базах данных, структура которых представляет собой совокупность взаимосвязанных таблиц, каждая из которых содержит данные массива однотипных документов. Поэтому чаще всего задача ввода форм состоит в превращении образа документа в строчку таблицы (таблиц) базы данных, содержащих соответствующую информацию. Когда все документы будут введены в базу данных, можно будет их обработать.

Формализованные документы имеют строгий порядок, расположение и смысл структурных единиц, которые называют полями.

При заполнении формализованного документа требуется внести информацию в определенные служебные поля, а задача программы при вводе форм - определять эти специальные поля, распознавать информацию в них и ввести ее в базу данных. При этом особенностью ввода форм в компьютер

является необходимость распознавания текстов, заполненных от руки. Обычно в том случае, если форма должна распознаваться компьютером, заполняющего просят ввести текст отдельными буквами, и такой текст называют рукопечатным. Технологии распознавания рукопечатных символов обозначаются термином ICR (Intelligent Character Recognition). Распознавание рукопечатных символов представляет собой более сложную задачу по сравнению с распознаванием печатных, поскольку требуется распознать символ, вписанный в форму от руки с учетом возможных его отклонений, обусловленных индивидуальными особенностями почерка.

Задачи распознавания при вводе форм не обязательно связаны с распознаванием текста. При вводе форм может потребоваться распознавание различных меток и знаков, для которого тоже существует свой термин: OMR (Optical Mark Recognition). Например, в бюллетенях для голосования голосующего просят поставить крестик (или другой знак) напротив фамилии кандидата, и задача компьютера - распознать, есть в определенном поле какой-нибудь знак или нет.

## Виды систем ввода документов

Помимо технологических аспектов, существует также специфика, связанная с масштабом использования технологии ввода бумажных документов. Действительно, когда говорят о домашних или о персональных системах ввода документов, то обычно подразумевают именно распознавание неформализованных документов, при незначительном количестве ввода документов за единицу времени. В данном случае на первое место выходит именно технология OCR.

Напротив, ввод формализованных документов (форм) - это технология, которая обычно используется в организациях и имеет массовый характер. В данном случае наряду с распознаванием текстов встает масса технологических проблем: организация поточного (массового) сканирования, распределенная обработка, встраивание решения в корпоративные системы документооборота и т.д. Технологии, обеспечивающие решение данного набора задач, называют технологиями Data Capture.

Data Capture (дословно "захват данных") - это комплекс мероприятий по переводу бумажных документов в электронный архив для хранения и обеспечения доступа к ним. Из отечественных компаний на рынке ввода и распознавания документов и форм наиболее активно работают фирмы АBBYY и Cognitive Technologies.

Системы ввода документов с учетом масштаба технологии можно разделить на четыре вида:

- **Продукты для персонального ввода неформализованных документов**

В этой области следует прежде всего выделить продукт FineReader. Программу отличает высокая точность распознавания и оформления документа; большое количество языков распознавания, интеграция с Microsoft Office Word.

- **Продукты для промышленного ввода документов**

Программно-аппаратные решения на основе стандартных домашних сканеров не обеспечивают производительности, необходимой корпоративным заказчикам, которым нужны так называемые технологии потокового сканирования. Только специализированные программные комплексы в совокупности с высокопроизводительным офисным оборудованием позволяют без задержек распространять по организации электронные версии документов, поступивших в бумажном виде.

Здесь можно обратить внимание на систему Ascent Capture , предназначенную для потоковой обработки сканируемых документов и извлечения данных. Максимальный эффект от применения этой системы отмечается при объемах поступления новых документов - от 1000 до 100 000 страниц в сутки.

Среди российских разработок промышленного ввода документов можно выделить систему электронного архива на базе ЕВФРАТ компании Cognitive Technologies. Программа имеет встроенную OCR-систему и обеспечивает потоковое сканирование и распознавание неформализованных документов. Система позволяет автоматизировать сканирование и распознавание бумажных документов, ввод и классификацию документов из разных источников.

- **Продукты для персонального ввода форм**

Наиболее известными решениями являются АBBYY FormReader и система Cognitive Forms.

АBBYY FormReader обеспечивает высокую степень распознавания печатных символов, меток и штрих-кодов. Автоматический контроль результатов распознавания с помощью проверок по словарям и базам данных обеспечивает высокую корректность информации.

Программа легко настраивается на новые формы. Использование АBBYY FormReader, установленной на один компьютер, позволяет ежедневно вводить от 500 до 1000 страниц, в зависимости от сложности формы и аппаратной конфигурации.

Процедура создания шаблона формы в большинстве случаев достаточно проста. FormReader позволяет различить и

идентифицировать в едином потоке около 100 различных шаблонов и форм.

Процедура ввода документа включает следующие этапы: вначале производится сканирование незаполненной формы. Затем система находит такие элементы форм, как линии, текст, повторяющийся на всех формах, штрих-коды, что позволяет избежать их ручного выделения.

Затем пользователь указывает поля, которые должны содержать текст для распознавания, и для каждого из этих полей определяет колонку в таблице базы данных, соответствующую этому полю.

На следующем этапе определяются языки, используемые в распознаваемом тексте, типы данных для полей и правила контроля.

ABBYY FormReader предлагает большое число готовых правил контроля, которые позволяют гарантировать правильность ввода информации

Cognitive Forms обладает основными возможностями ABBYY FormReader, включая способность обрабатывать различные типы форм в одном потоке и осуществлять автоматическую проверку корректности данных.

Эта система позволяет вводить от 2000 до 3000 страниц ежедневно.

В программный комплект Cognitive Forms входит отдельный модуль "Дизайнер форм", с помощью которого пользователь может самостоятельно создавать формы документов, задавать описания полей и варианты контекстных проверок.

- **Продукты для промышленного ввода форм**

В этой области также можно указать решения на базе описанных выше технологий CuneiForm и ABBYY FormReader, которые адаптированы к промышленному использованию, а также ориентированы на системы потокового сканирования и распределенной обработки.

### *Хранение информации. Базы и хранилища данных*

Предметная область какой-либо деятельности - часть реального мира, подлежащая изучению с целью организации управления процессами и объектами. Предметная область может быть разделена (декомпозирована) на фрагменты: например, предприятие - это дирекция, плановые отделы, бухгалтерия, цеха, отделы маркетинга, логистики и продаж, клиенты, поставщики и т. д. Каждый фрагмент предметной области характеризуется множеством объектов и процессов, которые отражаются в виде совокупности фиксируемых данных. Анализ этих данных позволяет установить их структуру и соответствие между ними. В результате можно построить единую схему, которую называют моделью данных. Модели данных составляют основу баз данных.

*База данных, БД (Data Base)* это структурированный организованный набор данных, объединенных в соответствии с некоторой выбранной моделью, хранящихся в электронном виде и предназначенный для коллективного использования.

*Система управления базами данных, СУБД (Data Base Management System)* это специализированная программа или комплекс программ, предназначенные для манипулирования базой данных. Для создания информационной системы и управления ею СУБД необходима в той же степени, как для разработки программы на алгоритмическом языке необходим транслятор.

СУБД часто упрощенно или ошибочно называют "базой данных". Нужно различать набор данных (собственно БД) и программное обеспечение, предназначенное для организации и ведения баз данных (СУБД).

Отличительной чертой баз данных следует считать то, что данные хранятся совместно с их описанием, а в прикладных программах описание данных не содержится. Независимые от программ пользователя данные обычно называются метаданными или данными о данных. В ряде современных систем метаданные, содержащие также информацию о пользователях, форматы отображения, статистику обращения к данным и др. сведения, хранятся в специальном словаре базы данных.

На уровне физической модели электронная БД представляет собой файл или группу файлов в специализированном формате конкретной СУБД. В настоящее время наибольшее распространение получили реляционные базы данных.

*Реляционная база данных* - база данных, основанная на реляционной модели. Слово "реляционный" происходит от английского "relation" (отношение).

Теория реляционных баз данных была разработана доктором Эдгаром Коддом из компании ИВМ в 1970 году. В реляционных базах данных все данные представлены в виде простых таблиц, разбитых на строки и столбцы, на пересечении которых расположены данные. Запросы к таким таблицам возвращают таблицы, которые сами могут становиться предметом дальнейших запросов. Каждая база данных может включать несколько таблиц. Кратко особенности реляционной базы данных можно сформулировать следующим образом:

- данные хранятся в таблицах, состоящих из столбцов ("атрибутов") и строк ("записей");
- на пересечении каждого столбца и строчки стоит в точности одно значение;

- у каждого столбца есть свое имя, которое служит его названием, и все значения в одном столбце имеют один тип;
- между таблицами установлены соответствия, называемые связями;
- запросы к базе данных возвращают результат в виде таблиц, которые тоже могут выступать как объект запросов;
- строки в реляционной базе данных неупорядочены, упорядочивание производится в момент формирования ответа на запрос.

Общепринятым стандартом языка работы с реляционными базами данных в настоящее время является язык структурированных запросов (Structured Query Language - SQL). Это универсальный компьютерный язык, применяемый для создания, модификации и управления данными в реляционных базах данных.

#### *Классификация баз данных.*

По организации хранения данных:

- локальные СУБД (все части локальной СУБД размещаются на одном компьютере);
- распределенные СУБД (части СУБД могут размещаться на двух и более компьютерах).

По способу доступа к БД:

- файл-серверные;
- клиент-серверные;
- трехзвенные;
- встраиваемые.

*Файл-серверные СУБД.* Архитектура "файл-сервер" предполагает организацию вычислений связанных с обработкой данных на клиентском компьютере. На сервере хранится база данных, файлы которой предоставляются клиентским машинам для извлечения данных с помощью запросов. На время выполнения запроса клиентский компьютер плотно занимает сетевой канал, хотя ресурсы сервера расходуются незначительно. Минус - высокая загрузка сети. На данный момент файл-серверные СУБД считаются устаревшими. Примеры: Microsoft Access до 2003 версии, MySQL (до версии 5.0).

*Клиент-серверные СУБД.* Такие СУБД состоят из клиентской части (которая входит в состав прикладной программы) и сервера. Клиент-серверные СУБД, в отличие от файл-серверных, обеспечивают разграничение доступа между пользователями и меньше загружают сеть и клиентские машины. Сервер является внешней по отношению к клиенту программой, и по мере надобности его можно заменить другим. Недостаток клиент-серверных СУБД - в самом факте существования сервера (что плохо для локальных программ - в них удобнее встраиваемые СУБД) и больших

вычислительных ресурсах, потребляемых сервером. Примеры: Interbase, MS SQL Server, Oracle, DB2, PostgreSQL, MySQL (старше версии 5.0).

Существенным недостатком клиент-серверной архитектуры является необходимость установления прямого соединения между клиентским компьютером и базой данных.

При *трехзвенной архитектуре* пользовательское приложение (клиент) соединяется со специально выделенным сервером приложений, и только он уже соединяется с базой данных. Кроме повышения уровня безопасности трехзвенная архитектура позволяет более гибко модернизировать приложения. Как правило, в массовой клиентской части оставляют только минимальный набор функций по доступу и отображению информации, а основную бизнес-логику реализуют в программах, запускаемых на серверах приложений. При этом модернизация обычно затрагивает только сервер приложений, а на массовых клиентских местах переустанавливать ПО не приходится.

*Встраиваемая СУБД* - это, как правило, "библиотека", которая позволяет унифицированным образом хранить большие объемы данных на локальной машине. Доступ к данным может происходить через SQL либо через особые функции СУБД. Встраиваемые СУБД быстрее обычных клиент-серверных и не требуют установки сервера, поэтому востребованы в локальном ПО, которое имеет дело с большими объемами данных - например, геоинформационные системы (Geographic Informational System - GIS). Примеры: SQLite, BerkeleyDB, один из вариантов Firebird, один из вариантов MySQL.

В основе концепции хранилищ данных лежат следующие основополагающие идеи:

- интеграция ранее разъединенных детализированных данных (исторические архивы, данные из традиционных систем обработки документов, разрозненных баз данных, данные из внешних источников) в едином хранилище данных;
- тематическое и временное структурирование, согласование и агрегирование;
- разделение наборов данных, используемых для операционной (производственной) обработки, и наборов данных, используемых для решения задач анализа.

Данные, помещаемые в хранилище, должны отвечать определенным требованиям - предметной ориентированности, интегрированности, поддержки хронологии и неизменяемости

---

Предметная ориентированность	Все данные о некоторой сущности (бизнес-объекте, бизнес-процессе и т. д.) из некоторой предметной области собираются из множества различных источников, очищаются, согласовываются, дополняются, агрегируются и представляются в единой, удобной для их использования в бизнес-анализе форме
Интегрированность	Все данные о различных бизнес-объектах взаимно согласованы и хранятся в едином общекорпоративном хранилище
Поддержка хронологии	Данные хронологически структурированы и отражают историю за период времени, достаточный для выполнения задач бизнес-анализа, прогнозирования и подготовки принятия решения
Неизменяемость	Исходные (исторические) данные, после того как они были согласованы, верифицированы и внесены в общекорпоративное хранилище, остаются неизменными и используются исключительно в режиме чтения

Хранилище данных выполняет множество функций, но его основное предназначение - предоставление точных данных и информации в кратчайшие сроки и с минимумом затрат.

Достоинствами архитектуры классического хранилища данных являются:

- общая семантика;
- централизованная, управляемая среда;
- согласованный набор процессов извлечения и бизнес - логики использования;
- непротиворечивость содержащейся информации;
- легко создаваемые по шаблонам и наполняемые витрины данных;
- единый репозиторий метаданных;
- многообразие механизмов обработки и представления данных.

К недостаткам можно отнести большие затраты по реализации, высокую ресурсоемкость в масштабе всего предприятия, потребность в сложных сервисных системах, рискованный сценарий развития, когда все данные и метаданные находятся в одном репозитории и в неблагоприятном случае могут быть потеряны.

## **Организация обмена данными для управления экономическими объектами**

### ***Использование Интернет для обмена данными в среде управления***

Существует достаточно много толкований термина Internet, однако он имеет два основных качественных значения:

- глобальное сообщество произвольно объединяемых мировых сетей, которые используются для свободного обмена данными, информацией и знаниями;

- совокупность технологий, которые реализуют обмен данными на основе использования семейства протоколов TCP/IP (Transmission Control Protocol / Internet Protocol), называемых Internet-технологиями или технологиями Internet.

В основе создания Internet лежит история развития сети ARPAnet - первой экспериментальной компьютерной сети национального масштаба. Она была создана в конце 1960-х годов в целях поддержки научных исследований Министерства обороны США (Advanced Research Project Agency - ARPA) и объединила сотни компьютеров нескольких крупных научных и университетских центров. Узлы сети были связаны физическими выделенными линиями, а передача и прием данных обеспечивалась специальными программами, работающими на узловых компьютерах.

Сеть изначально предполагалась ненадежной - исследовалась возможность передачи данных в сети, отдельные фрагменты которой могут перестать функционировать в любой произвольный момент. Программные системы, в которые были заложены принципы искусственного интеллекта, должны были отыскивать работающие сегменты сети и "прокладывать" новые маршруты передачи данных. Выход из строя любого канала связи не должен был вывести такую сеть из строя. При этом общий алгоритм был основан на допущении, что любой компьютер мог связаться с любым "ответившим" компьютером как "равный с равным". Реально сеть стала использоваться для обмена сообщениями (E-mail) и файлового обмена (File-oriented Interchange).

Примерно в это же время появились локальные вычислительные сети (Local Area Network - LAN) и компьютеры с операционной системой UNIX, которые, помимо чисто вычислительных задач, стали обслуживать эти сети. Они получили название *рабочие станции*. ОС UNIX была выбрана потому, что в нее была заложена возможность работать с IP-протоколами, которые содержали:

- правила инициализации и поддержания работы в сети;
- описание информационных сетевых пакетов (пакетов данных) семейства IP;
- правила обращения с IP-пакетами (идентификация, проверка целостности, обработка, пересылка, прием и т. д.).

Эти решения оказались успешными, стандартизация протоколов позволила подключать к сети компьютеры с различным базовым программным обеспечением. Появилось понятие "трафик", трактуемое в единицах обмена информацией, которым стали измерять реальную загрузку сети. Технология передачи данных IP-пакетами оказалась чрезвычайно перспективной в техническом отношении, однако в чисто пользовательском

плане ее необходимо было дорабатывать, так как скорость передачи данных не могла компенсировать значительные затраты времени на поиск нужной информации в огромных массивах данных.

В марте 1989 года Тим Бернерс-Ли (Tim Berners-Lee, Conseil Europeen pour la Recherche Nucleaire - CERN, Женева) предложил концепцию распределенной информационной системы с целью "объединения знаний человечества", которую он назвал "*Всемирной паутиной*" (World Wide Web - WWW). Для её создания он объединил две существующие технологии - технологию применения IP-протоколов для передачи данных и технологию гипертекста (Hypertext Technology). Эта технология основана на реализации быстрого перехода от одного фрагмента текста к другому по выделенным ссылкам (Dedicated Links), при этом указанные фрагменты могут располагаться на физически разделенных компьютерных носителях. Информационная система, построенная на этих принципах, могла объединить множество информационных ресурсов, разбросанных по многочисленным открытым базам данных.

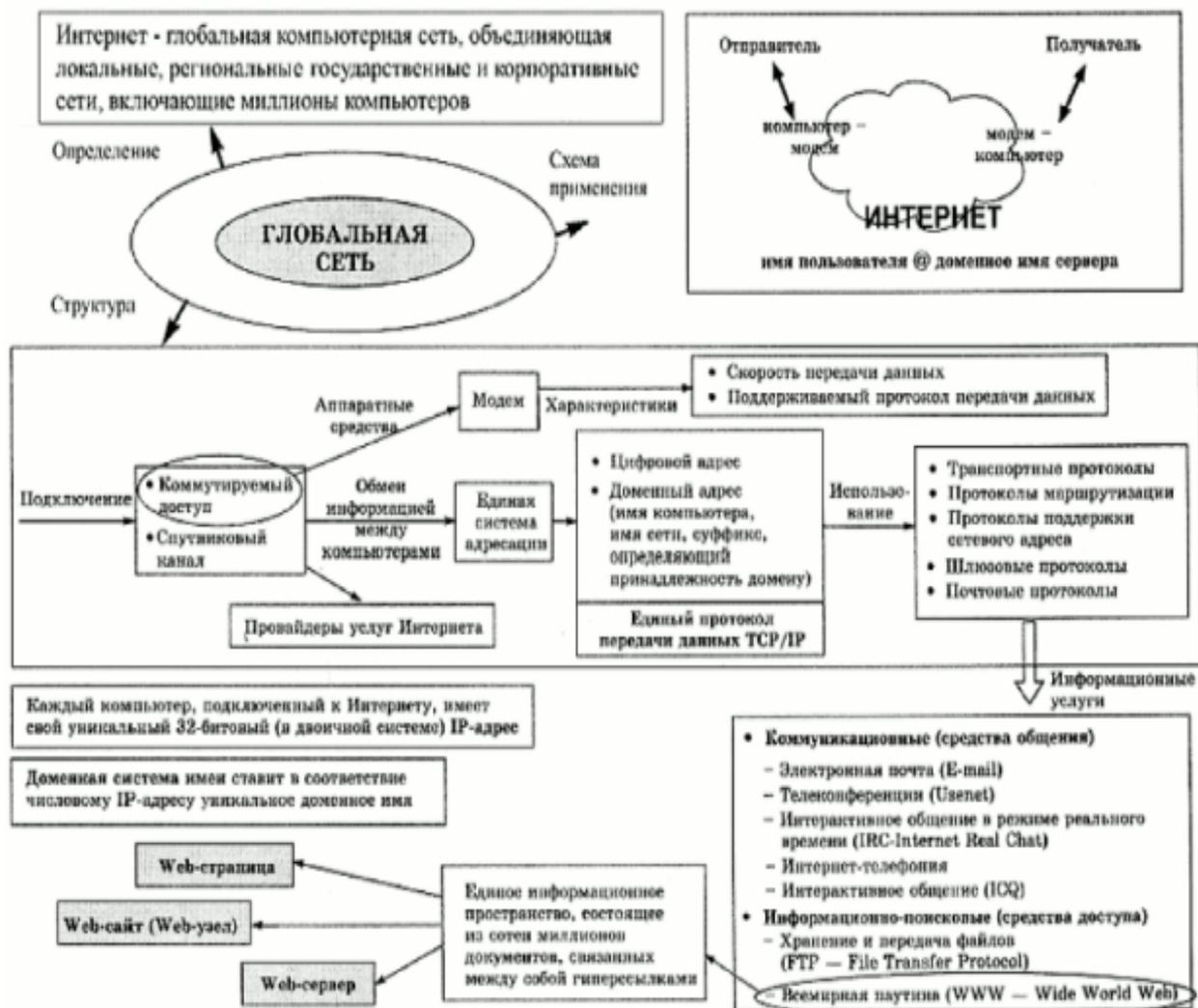
Основная метафора Web-гипертекста - это "*электронная книга*" с автоматически поддерживаемыми мгновенными переходами по ссылкам. Сам же термин *гипертекст* был впервые предложен Тедом Нельсоном в 1965 году, а первую работающую гипертекстовую систему создал в 1968 году Дуг Энгельбард.

В 1991 году был создан первый *браузер* (Browser) - компьютерная программа просмотра гипертекста, - работавший в режиме командной строки. Его применение позволило уже в 1992 году успешно реализовать предложенный проект, направленный в конечном итоге на создание "*бесшовного информационного пространства*" (Seamless Informational Area), охватывающего всю планету.

С точки зрения пользователя, информационное пространство "*всемирной паутины*" состоит из документов различного формата (мультимедиа-документов), предметных указателей и ссылок. Для перехода по ссылке или поиска по указателю пользователь применяет соответствующий браузер, "понимающий" язык разметки гипертекста. Поисковая система отыскивает по ссылке или ключевым словам в "паутине" нужный каталог, читает его структуру, считывает нужный документ и пересылает его пользователю. Web-сервер автоматически генерирует гипертекстовое представление требуемых файлов по запросам пользователя.

В сентябре 1994 года Оливер Мак-Брайан (Oliver McBryan) из Колорадского университета (США) разработал одно из первых автоматических средств составления предметного указателя для WWW,

названное WWW-Worm. За несколько минут Worm формировал базу данных из 300000 мультимедийных объектов, которые можно было находить по ключевым словам. Можно считать, что с этого момента информационное пространство World Wide Web было в принципе сформировано. Дальнейшее развитие шло по линии совершенствования технологий поиска, передачи, обеспечения безопасности, разработки и стандартизации различных Web-интерфейсов, повышающих комфорт использования Web-технологий. С середины 90-х годов эти технологии стали находить все более широкое применение во многих сферах человеческой деятельности.



Основными элементами технологии WWW являются [Артемьев В.И. Разработка INTRANET-приложений. Учебное пособие. Ярославль: изд-во ЯрГПУ, 1998, 233 с.]:

- язык гипертекстовой разметки документов (Hyper Text Markup Language - HTML);
- протокол обмена гипертекстовой информацией (Hyper Text Transfer Protocol - HTTP);

- универсальный способ адресации ресурсов в сети (Universal Resource Identifier - URI, и Universal Resource Locator - URL);
- система доменных имен (Domain Name System - DNS);
- универсальный интерфейс шлюзов (Common Gateway Interface - CGI), добавленный позже сотрудниками Национального Центра Суперкомпьютерных Приложений (National Center for Supercomputing Applications - NCSA).
- расширяемый язык разметки (eXtensible Markup Language - XML), рекомендованный Консорциумом Всемирной паутины.

Язык гипертекстовой разметки HTML создан на опыте использования редактора TeX и системно- и аппаратно-независимых методов представления текста в электронной форме (Standard Generalized Markup Language - SGML, стандарт ISO 8879). Основная идея гипертекста заключается в присутствии внутри ASCII-текста форматирующих полей и ссылок как на части внутри документа, так и на другие документы. Благодаря этому можно просматривать документы в том порядке, в каком требуется, а не последовательно, как при чтении книг. База данных гипертекста является частью файловой системы, которая содержит текстовые файлы в формате HTML и связанные с ними графику, мультимедиа и другие ресурсы.

Текстовый формат XML добавился несколько позже и был предназначен для описания систем хранения структурированных данных. Целью создания формата XML было обеспечение совместимости при передаче структурированных данных между разными системами обработки информации, особенно при передаче таких данных через Internet, а также для создания на его основе более специализированных языков разметки, иногда называемых *словарями*. Словари, основанные на XML, сами по себе формально описаны, что позволяет программно изменять и проверять документы на основе этих словарей, не зная их семантики, то есть не зная смыслового значения элементов. Важной особенностью XML также является применение так называемых пространств имен (Name Space).

Для получения файла из Internet браузеру нужно знать, где находится файл и как общаться с компьютером, на котором этот файл находится. Программа-клиент WWW передает имя необходимого файла, его местоположение в Internet (адрес хоста) и метод доступа (обычно протокол HTTP или FTP). Комбинация этих элементов формирует универсальный идентификатор ресурса (Universal Resource Identifier - URI). URI определяет способ записи адресов различных информационных ресурсов. В основу URI были заложены идеи расширяемости, полноты и читаемости. Реализация URI для WWW является способом адресации в сети (Universal Resource Locator - URL). Общий формат ссылки URL - <протокол://узел/путь/файл /метка>.

Internet является совокупностью эффективных методов коммуникации (на базе современных стандартизированных протоколов связи) и работы с информацией, находящейся на удаленных носителях. Кроме непосредственных функций по транзиту данных любых типов технологии Internet обеспечивают широкий спектр разнообразных информационных услуг, реализуемых различными службами:

- служба пересылки и приема сообщений (E-mail);
- служба гипертекстовой среды (WWW);
- служба передачи файлов (File Transfer Protocol - FTP);
- служба удаленного управления компьютером (Teletype Network - Telnet);
- служба имен доменов (Domain Name System);
- служба телеконференций (Users Network - Usenet) и чат-конференций (Интернет Relay Chat - IRC).

Программная индустрия для Web испытывает сейчас настоящий бум: сотни компаний - разработчиков программного обеспечения для Web создают новые технологии и инструментальные средства для навигации, работы в Сети и разработки пользовательских приложений. К их числу можно отнести:

программы просмотра и навигации (браузеры);  
средства поиска и доставки информации (поисковые машины);  
программное обеспечение Internet и Web-серверов, серверные приложения и расширения;  
средства администрирования в сетях;  
клиентские приложения и расширения (Web-сервисы);  
инструментальные средства разработки;  
средства обеспечения безопасности.

Инструментальные средства разработки Internet-приложений разнообразны и включают:

- редакторы гипертекста и графические редакторы;
- средства разметки карт изображений и конверторы изображений;
- средства мультимедиа (аудио, анимация, видео);
- средства генерации виртуальной реальности;
- средства и языки программирования серверных и клиентских приложений и расширений.

*Редакторы гипертекста* формируют HTML-файлы в режимах программирования или WYSIWYG (What You See Is What You Get). Можно использовать и обычные текстовые редакторы, а также средства, встроенные в браузеры. К этой же группе относятся конверторы, "перегоняющие" офисные документы в гипертекст. Графические редакторы служат для создания изображений, включаемых в гипертекст.

*Средства разметки карт изображений* позволяют разбить изображение на участки и связать гиперссылки с каждым из них. Такие средства могут быть встроены в графический редактор. Конверторы изображений обеспечивают преобразование форматов, размеров и цветов, создание специальных эффектов.

*Средства мультимедиа* предназначены для создания звукового и музыкального сопровождения, анимационных и видеороликов. Часто воспроизведение файлов мультимедиа осуществляется клиентскими расширениями или специальными Helper-программами.

*Средства генерации виртуальной реальности* позволяют запрограммировать трехмерные сцены и управление ими на языке VRML (Virtual Reality Modeling Language). Ввиду того, что процесс воспроизведения виртуальной реальности достаточно сложен, могут потребоваться дополнительные средства автоматизированного проектирования и анимации. Для просмотра Web-страниц с VRML-изображениями необходимо использовать соответствующие браузеры, например: WebSpace от Silicon Graphics или VRML-расширения для Internet Explorer или Netscape Navigator.

*Системы программирования клиентских приложений* предназначены для разработки и отладки сценариев (на языках VBScript или JavaScript) и мобильных приложений (на языке Java), выполняемых на стороне клиента. Наибольшее удобство и производительность разработки дают средства визуального программирования. В качестве средств программирования серверных приложений могут применяться как обычные системы программирования (Visual Basic, C/C++, Java), так и интерпретаторы команд (UNIX-shell, REXX и др.) и интерпретаторы и компиляторы сценариев на JavaScript, VBScript и Perl. Для создания клиентских и серверных расширений используются системы программирования, которые позволяют создавать компоненты с использованием механизмов ActiveX или Plug-in, представленных в виде встроенных или дополнительных библиотек интерфейсов.

*Средства администрирования*, как правило, поставляются в составе программного обеспечения Web-сервера и служат для конфигурирования, активации и мониторинга Web-сервисов, для контроля актуальности гиперссылок и связности гипертекстовой структуры, для учета и протоколирования использования серверов, для настройки и сопровождения системы безопасности.

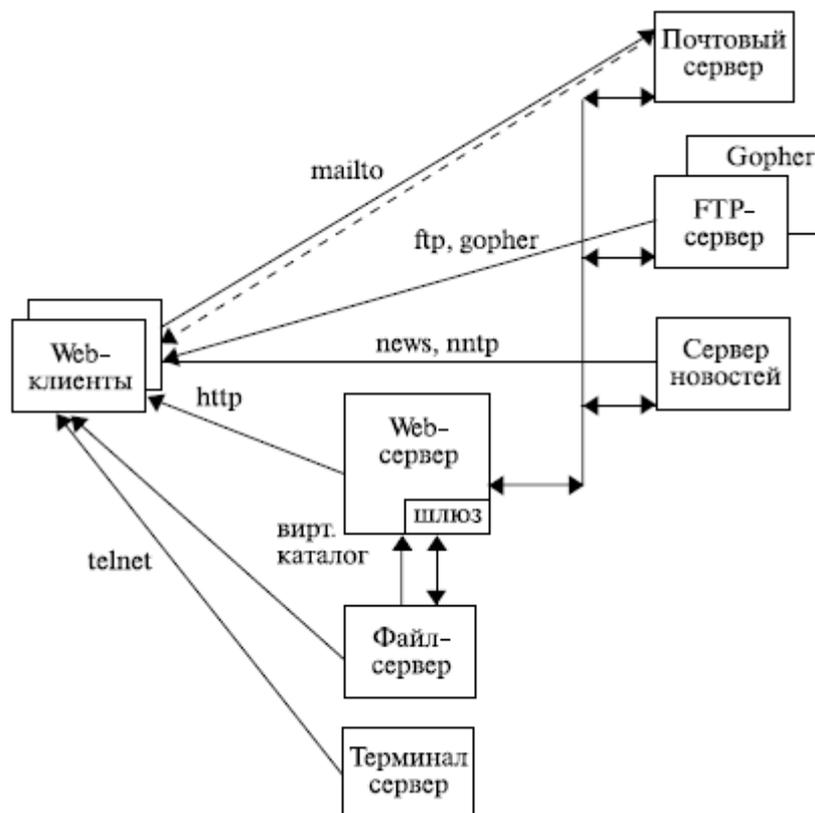
*Средства безопасности* могут быть встроены в программное обеспечение Internet-серверов или представлены в виде дополнительных

компонентов: комплексов Firewall и Proxy-серверов, выполняющих фильтрацию данных на различных уровнях.

На ранних стадиях развития сеть Internet была "улицей с односторонним движением", так как информация с Web-страниц поступала к пользователю от Web-сервера только при наличии запроса пользователя. С появлением в языке HTML диалоговых свойств пользователь получил обратную связь с Web-сервером. Обмен параметров при этом осуществляется через специальный графический интерфейс (Computer Graphical Interface - CGI).

В последнее время все большее распространение получает механизм согласования запускаемых программ через многоцелевые расширения почтовой службы Internet (Multipurpose Internet Mail Extensions - MIME). Современные браузеры, помимо взаимодействия с Web-серверами через протокол http, могут работать с различными типами серверов и служб с использованием протоколов FTP, File, Gopher, Mailto, NNTP, Telnet, WAIS.

*Intranet* - это внутреннее информационное пространство организации, реализуемое либо в локальной сети LAN (Local Area Network), либо в компьютерной сети WAN (Wide Area Network), охватывающей несколько территорий и включающей в себя десятки и/или сотни тысяч компьютеров) и обладающее всеми возможностями Internet.



*Intranet* ориентирован, как правило, на применение в рамках одного компактного или распределенного предприятия и отличается высокой безопасностью и скоростью работы. Используется для решения задач по автоматизации документооборота, информационному сопровождению бизнес-процессов, поиска и совместного доступа к данным и документам организации и имеет шлюзы для подключения в Internet. Для примера можно привести Intranet-сети, реализованные на основе технологий Microsoft. Пользователь работает с данными в привычном интерфейсе, пользуясь средствами Microsoft Office для доступа к сетевым данным.

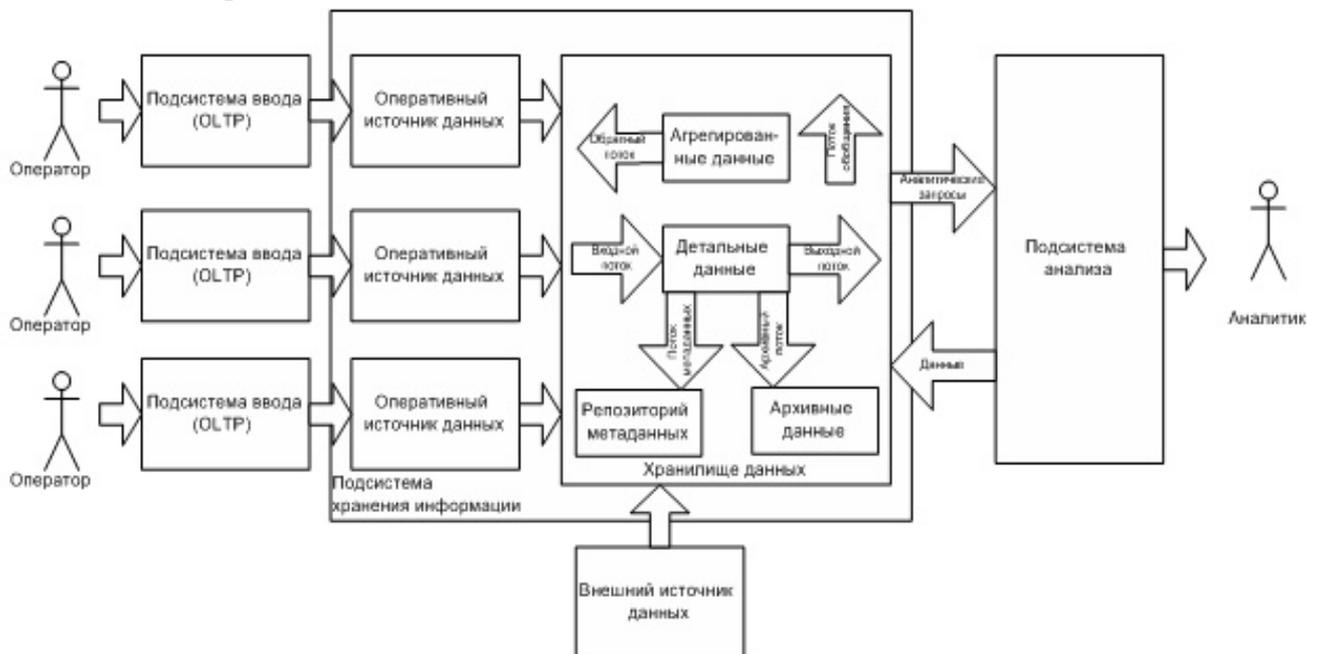
## Понятие и модель данных OLAP

### Понятие OLAP

OLAP (Online Analytical Processing) - технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях бизнес-анализа (BI, Business Intelligence) или, в русскоязычной терминологии, поддержки процессов принятия решений (СППР).

Основное назначение OLAP-систем - поддержка аналитической деятельности, произвольных запросов пользователей - аналитиков. Цель OLAP-анализа - проверка возникающих гипотез.

### Категории данных в ХД



Сбор и хранение информации, а также решение задач информационно-поискового запроса эффективно реализуются средствами систем управления базами данных (СУБД). В OLTP (*Online Transaction Processing*)-подсистемах реализуется *транзакционная обработка* данных. Непосредственно OLTP-системы не подходят для полноценного анализа информации в силу противоречивости требований, предъявляемых к OLTP-системам и *СППР*.

Для предоставления необходимой для принятия решений информации обычно приходится собирать данные из нескольких *транзакционных баз данных* различной структуры и содержания. Основная проблема при этом состоит в несогласованности и противоречивости этих баз-источников, отсутствии единого логического взгляда на корпоративные данные.

Поэтому для объединения в одной системе OLTP и *СППР* для реализации подсистемы хранения используются концепция хранилищ данных (ХД). В основе концепции ХД лежит идея разделения данных, используемых для оперативной обработки и для решения задач анализа, что позволяет оптимизировать структуры хранения. ХД позволяет интегрировать ранее разъединенные детализированные данные, содержащиеся в исторических архивах, накапливаемых в традиционных OLTP-системах, поступающих из внешних источников, в единую базу данных, осуществляя их предварительное согласование и, возможно, агрегацию.

Подсистема анализа может быть построена на основе:

- подсистемы информационно-поискового анализа на базе реляционных СУБД и статических запросов с использованием языка SQL;
- подсистемы оперативного анализа. Для реализации таких подсистем применяется технология оперативной аналитической обработки данных OLAP, использующая концепцию многомерного представления данных;
- подсистемы интеллектуального анализа, реализующие методы и алгоритмы Data Mining.

Все данные в ХД делятся на следующие категории :

1. детальные данные - данные, переносимые непосредственно из OLTP-подсистем. Соответствуют *элементарным событиям*, фиксируемым в OLTP-системах. Подразделяются на:
  - измерения - наборы данных, необходимые для описания событий (товар, продавец, покупатель, магазин, ... );
  - факты - данные, отражающие сущность события (количество проданного товара, сумма продаж, ...);

2. агрегированные (обобщенные) данные - данные, получаемые на основании детальных путем суммирования по определенным измерениям;
3. метаданные - данные о данных, содержащихся в ХД. Могут описывать:
  - объекты предметной области, информация о которых содержится в ХД;
  - категории пользователей, использующих данные в ХД;
  - места и способы хранения данных;
  - действия, выполняемые над данными;
  - время выполнения различных действий над данными;
  - причины выполнения различных действий над данными.

### *Информационные потоки в ХД*

Данные в ХД образуют следующие информационные потоки:

входной поток - образуется данными, копируемыми из OLTP-систем в ХД; данные при этом часто очищаются и обогащаются путем добавления новых атрибутов;

поток обобщения - образуется агрегированием детальных данных и их сохранением в ХД;

архивный поток - образуется перемещением детальных данных, количество обращений к которым снизилось;

поток метаданных - образуется потоком информации о данных в репозиторий данных;

выходной поток - образуется данными, извлекаемыми пользователями;

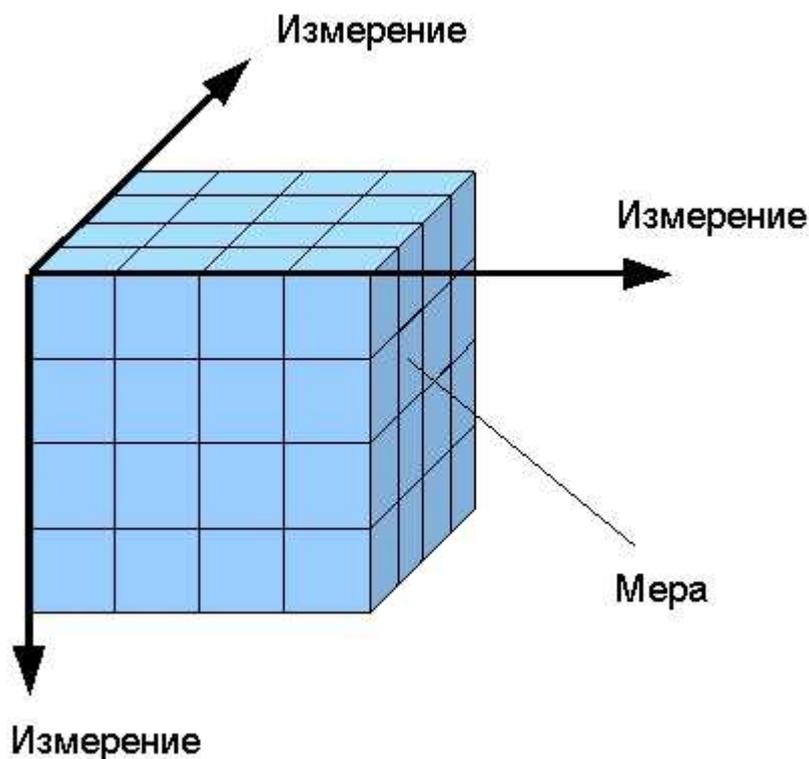
обратный поток - образуется очищенными данными, записываемыми обратно в OLTP-системы.

### *Структура OLAP-куба*

В процессе анализа данных часто возникает необходимость построения зависимостей между различными параметрами, число которых может быть значительным.

Под измерением будем понимать последовательность значений одного из анализируемых параметров. Например, для параметра "время" это - последовательность дней, месяцев, кварталов, лет.

Возможность анализа зависимостей между различными параметрами предполагает возможность представления данных в виде многомерной модели - гиперкуба или OLAP-куба.



### Гиперкуб

Оси куба представляют собой измерения, по которым откладывают параметры, относящиеся к анализируемой предметной области, например, названия товаров и названия месяцев года.

На пересечении осей измерений располагаются данные, количественно характеризующие анализируемые факты - меры, например, объемы продаж, выраженные в единицах продукции.

В простейшем случае двумерного куба получается таблица, показывающая значения уровней продаж по товарам и месяцам.

Дальнейшее усложнение модели данных возможно по нескольким направлениям:

- увеличение числа измерений данные о продажах не только по месяцам и товарам, но и по регионам. В этом случае куб становится трехмерным;
- усложнение содержимого ячейки например, нас может интересовать не только уровень продаж, но и *чистая прибыль* или остаток на складе. В этом случае в ячейке будет несколько значений;
- введение иерархии в пределах одного измерения общее понятие "время" связано с иерархией значений: год состоит из кварталов, квартал из месяцев и т.д.

## *Иерархия измерений OLAP-кубов*

Каждое из измерений OLAP-куба может быть представлено в виде иерархической структуры. Например, измерение "Регион" может иметь следующие уровни иерархии: "страна - федеральный округ - область - город - район".

Некоторые измерения могут иметь несколько уровней иерархического представления, например измерение "время" - представление "год - квартал - месяц - день" и представление "год - неделя - день".

Точно так же в рамках измерения "География" можно ввести уровни "Страна", "Регион", "Область" и "Город".

## *Таблица фактов*

Таблица фактов - является основной таблицей хранилища данных. Как правило, она содержит сведения об объектах или событиях, совокупность которых будет в дальнейшем анализироваться. Обычно говорят о четырех наиболее часто встречающихся типах фактов. К ним относятся:

- факты, связанные с транзакциями (Transaction facts). Они основаны на отдельных событиях (типичными примерами которых являются телефонный звонок или снятие денег со счета с помощью банкомата);
- факты, связанные с "моментальными снимками" (Snapshot facts). Основаны на состоянии объекта (например, банковского счета) в определенные моменты времени, например на конец дня или месяца. Типичными примерами таких фактов являются объем продаж за день или дневная выручка;
- факты, связанные с элементами документа (Line-item facts). Основаны на том или ином документе (например, счете за товар или услуги) и содержат подробную информацию об элементах этого документа (например, количестве, цене, проценте скидки);
- факты, связанные с событиями или состоянием объекта (Event or state facts). Представляют возникновение события без подробностей о нем (например, просто факт продажи или факт отсутствия таковой без иных подробностей).

Таблица фактов, как правило, содержит уникальный составной ключ, объединяющий первичные ключи таблиц измерений. Чаще всего это целочисленные значения либо значения типа "дата/время" - ведь таблица фактов может содержать сотни тысяч или даже миллионы записей, и хранить в ней повторяющиеся текстовые описания, как правило, невыгодно - лучше поместить их в меньшие по объему таблицы измерений. При этом как ключевые, так и некоторые неключевые поля должны соответствовать будущим измерениям OLAP-куба. Помимо этого таблица фактов содержит

одно или несколько числовых полей, на основании которых в дальнейшем будут получены агрегатные данные.

Для многомерного анализа пригодны таблицы фактов, содержащие как можно более подробные данные (то есть соответствующие членам нижних уровней иерархии соответствующих измерений). В данном случае предпочтительнее взять за основу факты продажи товаров отдельным заказчикам, а не суммы продаж для разных стран - последние все равно будут вычислены OLAP-средством.

В таблице фактов нет никаких сведений о том, как группировать записи при вычислении агрегатных данных. Например, в ней есть идентификаторы продуктов или клиентов, но отсутствует информация о том, к какой категории относится данный продукт или в каком городе находится данный клиент. Эти сведения, в дальнейшем используемые для построения иерархий в измерениях куба, содержатся в таблицах измерений.

### *Таблицы измерений*

Таблицы измерений содержат неизменяемые либо редко изменяемые данные. В подавляющем большинстве случаев эти данные представляют собой по одной записи для каждого члена нижнего уровня иерархии в измерении. Таблицы измерений также содержат как минимум одно описательное поле (обычно с именем члена измерения) и, как правило, целочисленное ключевое поле (обычно это суррогатный ключ) для однозначной идентификации члена измерения. Если будущее измерение, основанное на данной таблице измерений, содержит иерархию, то таблица измерений также может содержать поля, указывающие на "родителя" данного члена в этой иерархии. Нередко (но не всегда) таблица измерений может содержать и поля, указывающие на "прародителей", и иных "предков" в данной иерархии (это обычно характерно для сбалансированных иерархий), а также дополнительные атрибуты членов измерений, содержащиеся в исходной оперативной базе данных (например, адреса и телефоны клиентов).

Каждая таблица измерений должна находиться в отношении "один ко многим" с таблицей фактов.

### *Архитектура OLAP-систем*

Полномасштабная OLAP-система должна выполнять сложные и разнообразные функции, включающие сбор данных из различных источников, их согласование, преобразование и загрузку в хранилище, хранение аналитической информации, регламентную отчетность, поддержку произвольных запросов, многомерный анализ и др.

В настоящее время существуют фактические стандарты построения OLAP-систем, основанных на концепции ХД. Эти стандарты опираются на современные исследования и общемировую практику создания хранилищ данных и аналитических систем.

В общем виде архитектура корпоративной OLAP-системы описывается схемой с тремя выделенными слоями :

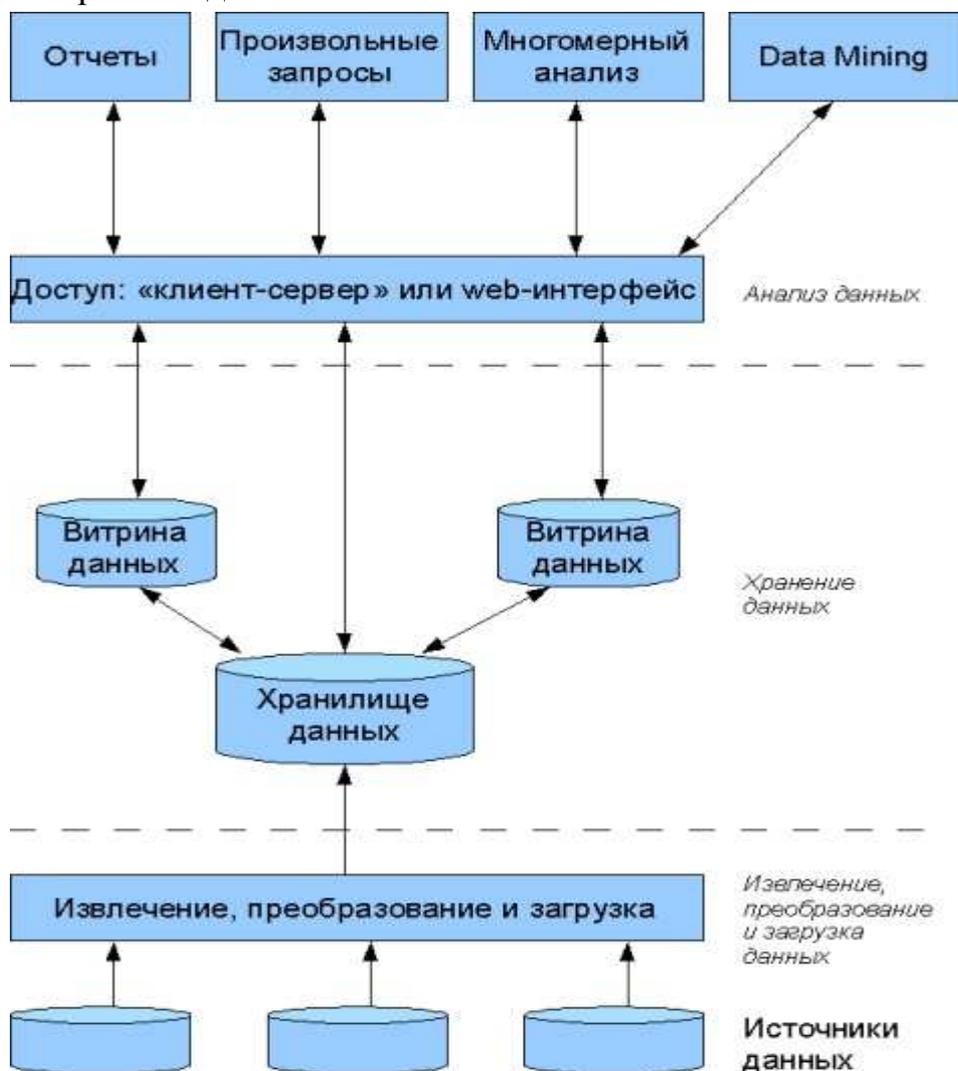


Рис. Архитектура корпоративной OLAP-системы

- извлечение, преобразование и загрузка данных;
- хранение данных;
- анализ данных.

Данные поступают из различных внутренних OLTP-систем, от подчиненных структур, от внешних организаций в соответствии с установленным регламентом, формами и макетами отчетности. Вся эта информация проверяется, согласуется, преобразуется и помещается в хранилище и витрины данных. После этого пользователи с помощью специализированных инструментальных средств получают необходимую им

информацию для построения различных табличных и графических представлений, прогнозирования, моделирования и выполнения других аналитических задач.

### Слой извлечения, преобразования и загрузки данных

С организационной точки зрения, данный слой включает подразделения и структуры организации всех уровней, поддерживающие базы данных оперативного доступа. Он представляет собой низовой уровень генерации информации, уровень внутренних и внешних информационных источников, вырабатывающих "сырую" информацию. Эта информация является рабочей для повседневной деятельности различных подразделений, которые ее вырабатывают и используют.

С системно-технической точки зрения данный слой представлен ЛВС всех подразделений всех уровней, к которым подключены специализированные технические комплексы, хранящие информацию, чаще всего реализованные в виде реляционных СУБД.

Из источников данных информация перемещается на основе некоторого регламента в *централизованное хранилище*. Как правило, необходимые для хранилища данные не хранятся в окончательном виде ни в одной из OLTP-систем. Эти данные обычно можно получить из исходных баз данных путем специальных преобразований, вычислений и агрегирования.

Кроме того, несмотря на различную функциональную направленность, исходные транзакционные системы часто "пересекаются" по данным, т.е. их локальные базы данных содержат однотипную по смыслу информацию. Это, прежде всего, касается нормативно-справочной информации, которая используется в том или ином виде в любой OLTP-системе. При этом существенно, что одинаковые по смыслу данные обычно имеют в разных системах различный формат, вид представления, идентификацию, единицы измерения и т.п. Перед загрузкой в хранилище вся эта информация должна быть согласована, чтобы обеспечить целостность и непротиворечивость аналитических данных.

Согласование данных необходимо и при загрузке данных из одного источника. Дело в том, что в хранилище хранятся исторические данные, т.е. данные за достаточно большой промежуток времени. В оперативной системе данные хранятся в целостном виде за ограниченный промежуток, после чего они отправляются в архив. При изменениях в структуре или собственно данных архивы не подвергаются никакой дополнительной обработке, а хранятся в исходном виде. Следовательно, при необходимости иметь данные за достаточно большой период времени необходимо согласовывать архивную информацию с текущей.

Таким образом, загрузка данных из источников в хранилище осуществляется специальными процедурами, позволяющими:

- извлекать данные из различных баз данных, текстовых файлов;
- выполнять различные типы согласования и очистки данных;
- преобразовывать данные при перемещении их от источников к хранилищу;
- загружать согласованные и "очищенные" данные в структуры хранилища.

### Слой хранения данных

Слой хранения данных предназначен непосредственно для хранения значимой, проверенной, согласованной, непротиворечивой и хронологически целостной информации, которую с достаточно высокой степенью уверенности можно считать достоверной.

Собственно ХД не ориентировано на решение какой-либо определенной функциональной аналитической задачи. Цель ХД - обеспечить целостность и поддерживать хронологию всевозможных корпоративных данных, и с этой точки зрения оно нейтрально по отношению к приложениям. В связи с этим в большинстве случаев для выполнения определенного комплекса функционально замкнутых аналитических задач рационально создавать витрины данных, в основе которых может быть как многомерная, так и реляционная модель данных. По существу витрина представляет собой относительно небольшое, но что самое важное, функционально-ориентированное ХД, в котором информация хранится специальным образом, оптимизированным с точки зрения решения конкретных аналитических задач некоторого подразделения или группы аналитиков.

ХД чаще всего реализуется в виде реляционной БД, работающей под управлением достаточно мощной реляционной СУБД. Такая СУБД должна поддерживать эффективную работу с терабайтными объемами информации, иметь развитые средства ограничения доступа, обеспечивать повышенный уровень надежности и безопасности, соответствовать необходимым требованиям по восстановлению и архивации.

### Слой анализа данных

Для организации доступа аналитиков к данным ХД и ВД используются специализированные рабочие места, поддерживающие необходимые технологии как оперативного, так и долговременного анализа. Результаты работы аналитиков оформляются в виде отчетов, графиков, рекомендаций и сохраняются как на локальном компьютере, так и в общедоступном узле локальной сети.

Аналитическая деятельность в рамках корпорации достаточно разнообразна и определяется характером решаемых задач, организационными особенностями компании, уровнем и степенью подготовленности аналитиков.

В связи с этим современный подход к инструментальным средствам анализа не ограничивается использованием какой-то одной технологии. В настоящее время принято различать следующие основные виды аналитической деятельности:

- стандартная отчетность;
- нерегламентированные запросы;
- многомерный анализ (OLAP);
- извлечение знаний (data mining).

Каждая из этих технологий имеет свои особенности, определенный набор типовых задач и должна поддерживаться специализированной инструментальной средой.

### *Клиентские OLAP-средства*

Клиентские OLAP-средства представляют собой приложения, осуществляющие вычисление агрегатных данных (сумм, средних величин, максимальных или минимальных значений) и их отображение, при этом сами агрегатные данные содержатся в кэше внутри адресного пространства такого OLAP-средства.

Если исходные данные содержатся в настольной СУБД, вычисление агрегатных данных производится самим OLAP-средством. Если же источник исходных данных - серверная СУБД, многие из клиентских OLAP-средств посылают на сервер SQL-запросы, содержащие оператор GROUP BY, и в результате получают агрегатные данные, вычисленные на сервере.

Как правило, OLAP-функциональность реализована в средствах статистической обработки данных (из продуктов этого класса на российском рынке широко распространены продукты компаний StatSoft и SPSS) и в некоторых электронных таблицах. В частности, средствами многомерного анализа обладает Microsoft Excel. С помощью этого продукта можно создать и сохранить в виде файла небольшой локальный многомерный OLAP-куб и отобразить его двух- или трехмерные сечения.

Надстройки к пакету приложений Microsoft Office для извлечения и обработки данных представляют собой ряд функций, обеспечивающих доступ к возможностям извлечения и обработки данных из приложений Microsoft Office, и тем самым позволяющих осуществлять прогностический анализ на локальном компьютере. Благодаря тому, что встроенные в службы платформы Microsoft SQL Server алгоритмы извлечения и обработки данных

доступны из среды приложений Microsoft Office, бизнес-пользователи могут легко извлекать ценную информацию из сложных наборов данных всего несколькими щелчками мыши. Надстройки к пакету приложений Office для извлечения и обработки данных дают конечным пользователям возможность выполнять анализ непосредственно в приложениях Microsoft Excel и Microsoft Visio.

В состав Microsoft Office 2007 входят три отдельных OLAP-компонента:

- клиент извлечения и обработки данных для Excel позволяет создавать проекты извлечения и обработки данных на базе служб SSAS и управлять ими из Excel 2007;
- средства анализа таблиц для приложения Excel позволяют использовать встроенные в службы SSAS функции извлечения и обработки информации для анализа данных, хранящихся в таблицах Excel;
- шаблоны извлечения и обработки данных для приложения Visio позволяют визуализировать деревья решений, деревья регрессии, кластерные диаграммы и сети зависимостей на диаграммах Visio.

Ниже изображена сводная таблица Excel, используемая для доступа клиентов к данным служб аналитики.

	A	B	C	D
1	Total Sales Amount	Column Labels		
2	Row Labels	CY 2001	CY 2002	CY 2003
3	Components		85604.4568	303505.9056
4	Mountain	5506331.462	12520092.59	15501696.35
5	Road	5774299.691	17568248.73	17090770.85
6	Accessory	51177.8098	500827.408	1202658.024
7	Touring			7895098.596
8	Grand Total	11331808.96	30674773.18	41993729.72
9				
10				
11				
12				
13				

Рис. Сводная таблица Excel 2007

С помощью приложения Microsoft Office Visio можно аннотировать, дополнять и отображать графические представления результатов извлечения и обработки данных. Платформа SQL Server 2008 в сочетании с приложением Visio 2007 позволяет:

- визуализировать деревья решений, деревья регрессии, кластерные диаграммы и сети зависимостей;
- сохранять модели извлечения и обработки данных в виде документов Visio, внедренных в другие документы приложений Office или сохраненных в виде веб-страниц.

Клиентские OLAP-средства применяются, как правило, при малом числе измерений (обычно рекомендуется не более шести) и небольшом разнообразии значений этих параметров, - ведь полученные агрегатные данные должны уместиться в адресном пространстве подобного средства, а их количество растет экспоненциально при увеличении числа измерений. Поэтому даже самые примитивные клиентские OLAP-средства, как правило, позволяют произвести предварительный подсчет объема требуемой оперативной памяти для создания в ней многомерного куба.

### **Методы OLAP анализа**

OLAP-сервис представляет собой инструмент для анализа больших объемов данных в режиме реального времени. Взаимодействуя с OLAP-системой, пользователь сможет осуществлять гибкий просмотр информации, получать произвольные срезы данных и выполнять аналитические операции детализации, свертки, сквозного распределения, сравнения во времени одновременно по многим параметрам. Вся работа с OLAP-системой происходит в терминах предметной области и позволяет строить статистически обоснованные модели деловой ситуации.

Программные средства OLAP - это инструмент оперативного анализа данных, содержащихся в хранилище. Главной особенностью является то, что эти средства ориентированы на использование не специалистом в области информационных технологий, не экспертом-статистиком, а профессионалом в прикладной области управления - менеджером отдела, департамента, управления, и, наконец, директором. Средства предназначены для общения аналитика с проблемой, а не с компьютером. На рисунке показан элементарный OLAP-куб, позволяющий производить оценки данных по трем измерениям.

Многомерный OLAP-куб и система соответствующих математических алгоритмов статистической обработки позволяет анализировать данные любой сложности на любых временных интервалах.

Имея в своем распоряжении гибкие механизмы манипулирования данными и визуального отображения, менеджер сначала рассматривает с разных сторон данные, которые могут быть (а могут и не быть) связаны с решаемой проблемой.

Далее он сопоставляет различные показатели бизнеса между собой, стараясь выявить скрытые взаимосвязи; может рассмотреть данные более пристально, детализировав их, например, разложив на составляющие по времени, по регионам или по клиентам, или, наоборот, еще более обобщить представление информации, чтобы убрать отвлекающие подробности. После этого с помощью модуля *статистического оценивания и имитационного моделирования* строится несколько вариантов развития событий, и из них выбирается наиболее приемлемый вариант.

У управляющего компанией, например, может зародиться гипотеза о том, что разброс роста активов в различных филиалах компании зависит от соотношения в них специалистов с техническим и экономическим образованием. Чтобы проверить эту гипотезу, менеджер может запросить из хранилища и отобразить на графике интересующее его соотношение для тех филиалов, у которых за текущий квартал рост активов снизился по сравнению с прошлым годом более чем на 10%, и для тех, у которых повысился более чем на 25%. Он должен иметь возможность использовать простой выбор из предлагаемого меню. Если полученные результаты ощутимо распадутся на две соответствующие группы, то это должно стать стимулом для дальнейшей проверки выдвинутой гипотезы.

В настоящее время быстрое развитие получило направление, называемое *динамическим моделированием (Dynamic Simulation)*, в полной мере реализующее указанный выше принцип FASMI.

Используя динамическое моделирование, аналитик строит модель деловой ситуации, развивающуюся во времени, по некоторому сценарию. При этом результатом такого моделирования могут быть несколько новых бизнес - ситуаций, порождающих дерево возможных решений с оценкой вероятности и перспективности каждого.

Практически всегда задача построения аналитической системы для многомерного анализа данных - это задача построения *единой, согласованно функционирующей информационной системы, на основе неоднородных программных средств и решений*. И уже сам выбор средств для реализации ИС становится

чрезвычайно сложной задачей. Здесь должно учитываться множество факторов, включая взаимную совместимость различных программных компонент, легкость их освоения, использования и интеграции, эффективность функционирования, стабильность и даже формы, уровень и потенциальную перспективность взаимоотношений различных фирм производителей.

OLAP применим везде, где есть задача анализа многофакторных данных. Вообще, при наличии некоторой таблицы с данными, в которой есть хотя бы одна описательная колонка и одна колонка с цифрами, OLAP-инструмент будет эффективным средством анализа и генерации отчетов. В качестве примера применения OLAP-технологии рассмотрим исследование результатов процесса продаж.

Ключевые вопросы "Сколько продано?", "На какую сумму продано?" расширяются по мере усложнения бизнеса и накопления исторических данных до некоторого множества факторов, или разрезов: "..в Санкт-Петербурге, в Москве, на Урале, в Сибири...", "..в прошлом квартале, по сравнению с нынешним", "..от поставщика А по сравнению с поставщиком Б..." и т. д.

Ответы на подобные вопросы необходимы для принятия управленческих решений: об изменении ассортимента, цен, закрытии и открытии магазинов, филиалов, расторжении и подписании договоров с дилерами, проведения или прекращения рекламных кампаний и т. д.

Если попытаться выделить основные цифры (факты) и разрезы (аргументы измерений), которыми манипулирует аналитик, стараясь расширить или оптимизировать бизнес компании, то получится таблица, подходящая для анализа продаж как некий шаблон, требующий соответствующей корректировки для каждого конкретного предприятия.

Важный вопрос - наличие данных. Если они есть в каком-либо виде (Excel- или Access-таблица, данные из базы учетной системы, в виде структурированных отчетов филиалов), ИТ - специалист сможет передать их OLAP-системе напрямую или с промежуточным преобразованием. Для этого OLAP-системы имеют специальные инструменты конвертации данных.

После настройки OLAP-системы на данные пользователь получит возможность быстро получать ответы на ключевые вопросы путем простых манипуляций мышью над OLAP-таблицей и соответствующими меню. При этом будут доступны некоторые стандартные методы анализа, логически следующие из природы OLAP-технологии. К наиболее распространённым инструментам анализа можно отнести:

- Факторный (структурный) анализ

- Анализ динамики
- Анализ зависимостей
- Сопоставление
- Дисперсионный анализ

*Факторный (структурный) анализ.* Анализ структуры продаж для выявления важнейших составляющих в интересующем разрезе. Для этого удобно использовать, например, диаграмму типа "Пирог" в сложных случаях, когда исследуется сразу 3 измерения - "Столбцы". Например, в магазине "Компьютерная техника" за квартал продажи компьютеров составили \$100000, фототехники -\$10000, расходных материалов - \$4500. Вывод: оборот магазина зависит в большой степени от продажи компьютеров (на самом деле, быть может, расходные материалы необходимы для продажи компьютеров, но это уже анализ внутренних зависимостей).

*Анализ динамики (регрессионный анализ - выявление трендов).* Выявление тенденций, сезонных колебаний. Наглядно динамику отображает график типа "Линия". Например, объемы продаж продуктов компании Intel в течение года падали, а объемы продаж Microsoft росли. Возможно, улучшилось благосостояние среднего покупателя, или изменился имидж магазина, а с ним и состав покупателей. Требуется провести корректировку ассортимента. Другой пример: в течение 3 лет зимой снижается объем продаж видеокамер.

*Анализ зависимостей (корреляционный анализ).* Сравнение объемов продаж разных товаров во времени для выявления необходимого ассортимента - "корзины". Для этого также удобно использовать график типа "Линия". Например, при удалении из ассортимента принтеров в течение первых двух месяцев обнаружилось падение продаж картриджей с порошком.

*Сопоставление (сравнительный анализ).* Сравнение результатов продаж во времени, или за заданный период, или для заданной группы товаров. В зависимости от количества анализируемых факторов (от 1 до 3-х) используется диаграмма типа "Пирог" или "Столбцы". Пример: сравнение результатов продаж однотипных магазинов для оценки качества работы менеджеров.

*Дисперсионный анализ.* Исследование распределения вероятностей и доверительных интервалов рассматриваемых показателей. Применяется для прогнозирования и оценки рисков.

Этими видами анализа возможности OLAP не исчерпываются. Например, применяя в качестве алгоритма вычисления промежуточных и

окончательных итогов функции статистического анализа - дисперсию, среднее отклонение, моды более высоких порядков, - можно получить самые изощренные виды аналитических отчетов.

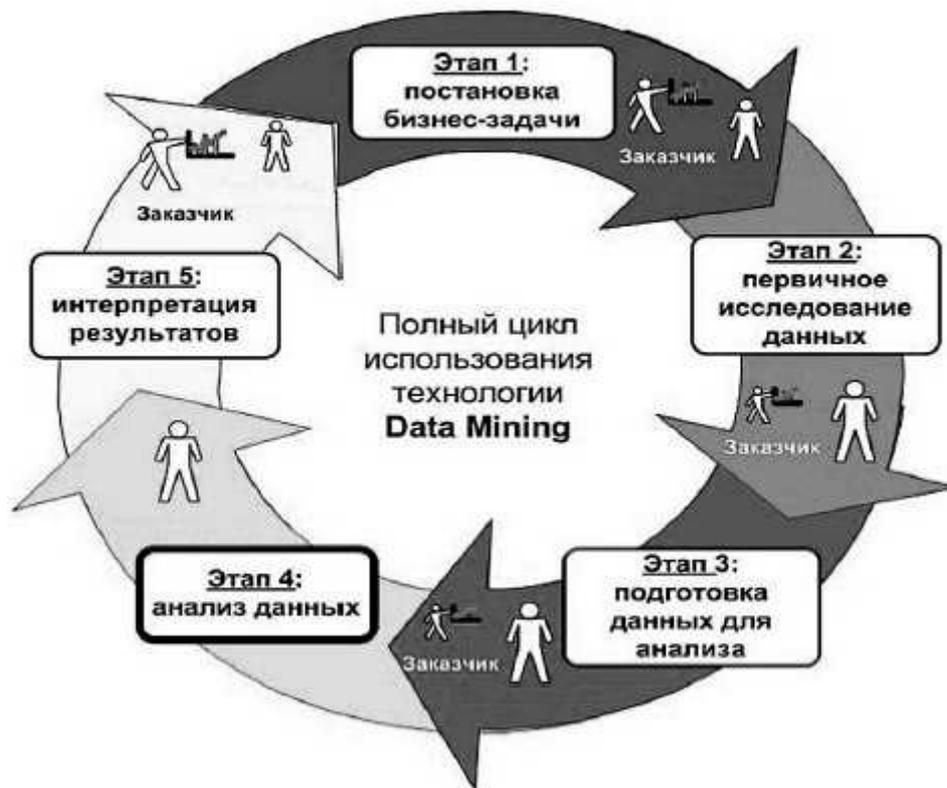
OLAP-системы являются частью более общего понятия "интеллектуальные ресурсы предприятия" или "средства интеллектуального бизнес-анализа" (Business Intelligence - BI), которое включает в себя помимо традиционного OLAP-сервиса средства организации совместного использования данных и информации, возникающих в процессе работы пользователей хранилища. Технология Business Intelligence обеспечивает электронный обмен отчетными документами, разграничение прав пользователей, доступ к аналитической информации из Internet и Intranet.

### *Технологии Data Mining*

В настоящее время элементы *искусственного интеллекта* активно внедряются в практическую деятельность менеджера. В отличие от традиционных систем искусственного интеллекта, технология интеллектуального поиска и анализа данных или "*добыча данных*" (Data Mining - DM), не пытается *моделировать* естественный интеллект, а *усиливает его возможности* мощностью современных вычислительных серверов, поисковых систем и хранилищ данных.

Data Mining - это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Data Mining представляют большую ценность для руководителей и аналитиков в их повседневной деятельности. Деловые люди осознали, что с помощью методов Data Mining они могут получить ощутимые преимущества в конкурентной борьбе.

В основу современной технологии Data Mining (Discovery-driven Data Mining) положена концепция шаблонов (Patterns), отражающих фрагменты многоаспектных взаимоотношений в данных. Эти шаблоны представляют собой закономерности, свойственные выборкам данных, которые могут быть компактно выражены в понятной человеку форме. Поиск шаблонов производится методами, не ограниченными рамками априорных предположений о структуре выборки и виде распределений значений анализируемых показателей. На рисунке показана схема преобразования данных с использованием технологии Data Mining.



Основой для всевозможных систем *прогнозирования* служит историческая информация, хранящаяся в БД в виде временных рядов. Если удастся построить шаблоны, адекватно отражающие динамику поведения целевых показателей, есть вероятность, что с их помощью можно предсказать и поведение системы в будущем. На рисунке показан полный цикл применения технологии Data Mining.

Важное положение Data Mining - нетривиальность разыскиваемых шаблонов. Это означает, что найденные шаблоны должны отражать *неочевидные, неожиданные* (Unexpected) регулярности в данных, составляющие так называемые скрытые знания (Hidden Knowledge). К деловым людям пришло понимание, что "сырые" данные (Raw Data) содержат глубинный пласт знаний, и при грамотной его раскопке могут быть обнаружены настоящие самородки, которые можно использовать в конкурентной борьбе.

Сфера применения Data Mining *ничем не ограничена* - технологию можно применять всюду, где имеются огромные количества каких-либо "сырых" данных!

В первую очередь методы Data Mining заинтересовали коммерческие предприятия, развертывающие проекты на основе *информационных хранилищ данных* (Data Warehousing). Опыт многих таких предприятий показывает, что отдача от использования Data Mining может достигать 1000%. Известны

сообщения об экономическом эффекте, в 10-70 раз превысившем первоначальные затраты от 350 до 750 тыс. долларов. Есть сведения о проекте в 20 млн долларов, который окупился всего за 4 месяца. Другой пример - годовая экономия 700 тыс. долларов за счет внедрения Data Mining в одной из сетей универсамов в Великобритании.

Можно назвать пять стандартных типов закономерностей, выявляемых с помощью методов Data Mining:

- ассоциация,
- последовательность,
- классификация,
- кластеризация
- прогнозирование.

*Ассоциация* имеет место в том случае, если несколько событий связаны друг с другом. Например, исследование, проведенное в компьютерном супермаркете, может показать, что 55% купивших компьютер берут также и принтер или сканер, а при наличии скидки за такой комплект принтер приобретают в 80% случаев. Располагая сведениями о подобной ассоциации, менеджерам легко оценить, насколько действенна предоставляемая скидка.

Если существует цепочка связанных во времени событий, то говорят о *последовательности*. Так, например, после покупки дома в 45% случаев в течение месяца приобретается и новая кухонная плита, а в пределах двух недель 60% новоселов обзаводятся холодильником.

С помощью *классификации* выявляются признаки, характеризующие группу, к которой принадлежит тот или иной объект. Это делается посредством анализа уже классифицированных объектов и формулирования некоторого набора правил.

*Кластеризация* отличается от классификации тем, что сами группы заранее не заданы. С помощью кластеризации средства Data Mining самостоятельно выделяют различные однородные группы данных.

## **Защита информации**

### ***Вредоносные программы***

Сегодня Интернет является самым популярным источником информации. В нем можно найти массу полезных программ. Многие компании предоставляют свои услуги через Интернет. Но у пользования Всемирной сетью есть и обратная сторона. Вирусы и другие компьютерные паразиты, нежелательная и рекламная информация, программы-шпионы и сетевые атаки - все это угрожает компьютеру, подсоединенному к Интернету.

Прежде всего необходимо рассмотреть вредоносные программы (или коды). По способу распространения их можно условно разделить на компьютерные вирусы, сетевые черви и троянские программы.

Компьютерные вирусы умеют размножаться и внедрять свои копии в другие файлы; сетевые черви распространяются по различным сетевым ресурсам (чаще всего по электронной почте), но не внедряют свои копии в другие программы; троянские программы не распространяются сами по себе, но выполняют на зараженных компьютерах вредоносные действия.

### ***Вирусы***

Компьютерные вирусы - это программы, способные размножаться самостоятельно, дописывая свой код к другим файлам или в служебные области диска. Каждый вирус способен выполнять деструктивные или нежелательные действия на зараженном компьютере. Он может:

- демонстрировать видеоэффекты,
- замедлять работу системы,
- похищать и уничтожать личную информацию пользователя,
- многое другое.

В любом случае вирус мешает другим программам и самому пользователю работать на компьютере.

Существует множество разновидностей вирусов. Самыми старыми являются файловые вирусы. Они размножаются, используя файловую систему. Почти столь же древними являются загрузочные вирусы. Они так названы потому, что заражают загрузочный сектор (boot sector) жесткого диска. Загрузочные вирусы замещают код программы, получающей управление при запуске системы. Таким образом, после перезагрузки системы управление передается вирусу. Следует отметить, что сегодня загрузочные вирусы встречаются очень редко. С середины 90-х годов получили распространение макровирусы. Эти вредители представляют собой программу на макроязыке. Макроязык - это средство создания программ, которое поддерживается некоторой системой обработки данных (например, программами семейства Microsoft Office, такими как Microsoft Word и Excel, а также другими программами, например продуктами компании "1С"). Для размножения макровирусы используют встроенные возможности, например, текстового или табличного редактора. Таким способом эти вредители переносят себя из одного зараженного файла в другой.

### ***Сетевые черви***

"Червей" часто называют вирусами, хотя, строго говоря, это не совсем верно. Сетевые черви - это программы, которые не изменяют файлы на

дисках, а распространяются в компьютерной сети, проникают в операционную систему компьютера, находят адреса других компьютеров или пользователей и рассылают по этим адресам свои копии. Сетевые черви могут вообще не обращаться к ресурсам компьютера (за исключением оперативной памяти). По среде распространения червей можно условно разделить на следующие типы: Интернет-черви (распространяются по Интернету), LAN-черви (распространяются по локальной сети), IRC-черви (распространяются через чаты Internet Relay Chat).

Есть еще одна крайне опасная разновидность червей - бестелесные. Эти паразиты не используют для своего размножения ни временных, ни постоянных файлов. Бестелесные черви вне компьютера существуют в виде сетевых пакетов, а внутри зараженного компьютера - в виде программного кода прямо в оперативной памяти. Такие черви распространяются автоматически, используя уязвимости в прикладном и системном программном обеспечении. Для этого не требуется вмешательство со стороны пользователя. Бестелесные черви распространяются очень быстро. Одному из них в конце января 2003 года удалось в течение нескольких часов вывести из строя 25% всех серверов в Интернете, т.е. каждый четвертый узел, запрашиваемый пользователем, был недоступен. Этот бестелесный червь известен под двумя названиями: Slammer и Helkern (более подробную информацию о червях можно найти по адресу [www.viruslist.ru](http://www.viruslist.ru))

### ***Троянские программы***

Троянские программы, "тройанские кони" и просто "тройанцы" - это вредоносные программы, которые сами не размножаются. Подобно знаменитому Троянскому коню из "Илиады" Гомера, программа-троянец выдает себя за что-то полезное. Чаще всего троянский конь маскируется под новую версию бесплатной утилиты, какую-то популярную прикладную программу или игру.

Таким способом "троянец" пытается заинтересовать пользователя и побудить его переписать и установить на свой компьютер вредителя самостоятельно.

По выполняемым вредоносным действиям троянские программы можно условно разделить на следующие виды:

- утилиты несанкционированного удаленного администрирования (позволяют злоумышленнику удаленно управлять зараженным компьютером);
- утилиты для проведения DDoS-атак (Distributed Denial of Service - распределенные атаки типа отказ в обслуживании);
- шпионские и рекламные программы, а также программы дозвона;

- серверы рассылки спама;
- многокомпонентные "троянцы"-загрузчики (переписывают из Интернета и внедряют в систему другие вредоносные коды или вредоносные дополнительные компоненты).

Следует отметить, что на практике часто встречаются программы-"троянцы", относящиеся сразу к нескольким перечисленным выше видам.

### ***Утилиты несанкционированного удаленного администрирования***

Удаленное управление компьютером часто используется в крупных и средних компаниях, а также в тех случаях, когда необходимо оказать техническую помощь пользователю, находящемуся на значительном расстоянии. С помощью средств удаленного управления системный администратор может настроить каждый компьютер в организации, не вставая со своего рабочего места. Однако эта полезная функциональность в руках злоумышленника превращается в грозное оружие. "Троянские кони" часто представляют собой вполне легальные утилиты удаленного управления, адаптированные под нужды хакеров. Если злоумышленнику удастся внедрить такого "троянца" в чужую систему, он сможет незаметно управлять этим компьютером втайне от его настоящего владельца.

Управление зараженным компьютером обычно осуществляется через Интернет. Вот лишь небольшая часть того, что может сделать злоумышленник на инфицированном ПК: выкрасть любую информацию с компьютера-жертвы (файлы, пароли, реквизиты и т.д.), провести любую файловую операцию (отформатировать жесткий диск, стереть или переименовать какие-то файлы и т.д.), перезагрузить компьютер, подключиться к сетевым ресурсам, использовать зараженный компьютер для атаки на какой-то третий компьютер или сервер в Интернете.

Самыми известными утилитами несанкционированного удаленного администрирования являются "троянцы" Back Orifice и NetBus.

### ***Утилиты для проведения DDoS-атак***

Цель DoS-атаки, или атаки типа отказ в обслуживании, - исчерпать ресурсы информационной системы. В случае успешного проведения DoS-атаки система перестает выполнять свои функции, становится недоступной и иногда непредсказуемой. Чаще всего объектом атаки типа отказ в обслуживании является web-сервер, например Интернет-магазин.

DDoS-атака, или распределенная атака типа отказ в обслуживании, отличается от DoS-атаки тем, что в ней один и тот же узел атакуют сразу несколько компьютеров. Для того чтобы исчерпать ресурсы web-сервера,

злоумышленник должен искусственно создать повышенную нагрузку на него. Каждый web-сервер тратит определенные ресурсы (память, вычислительные мощности и т.д.) на обработку входящих запросов. Если большое число компьютеров, на которых установлена утилита для проведения DDoS-атак, одновременно начнут посылать свои запросы web-серверу, то велика вероятность, что ресурсы web-сервера быстро исчерпаются, а сам сервер не сможет обслуживать легальных пользователей. При проведении масштабной распределенной атаки типа отказ в обслуживании злоумышленник чаще всего контролирует и координирует действия зараженных троянской программой компьютеров.

### ***Шпионское и рекламное ПО, программы дозвона***

Шпионские программы втайне наблюдают за действиями пользователя и записывают в свой журнал интересующие злоумышленника события. Существует класс программ - клавиатурные шпионы. Эти вредители следят за пользователем и записывают каждое нажатие клавиши. По команде хакера или через определенное время клавиатурный шпион отправляет собранные сведения на компьютер злоумышленника. Существуют также "троянцы"-шпионы, которые отправляют на удаленный компьютер пароли и другую личную информацию пользователя. За последние несколько лет получили распространение шпионские программы, которые собирают сведения о предпочтениях пользователя, часто посещающего Интернет. Такие вредители записывают адреса посещенных сайтов и имена загруженных файлов, а потом передают эти данные на чужой компьютер. Эта информация представляет большую ценность для компаний, проводящих маркетинговые, аналитические и статистические исследования. Поэтому злоумышленники идут на все, чтобы добыть ее. Даже на нарушение закона. Следует отметить, что чаще всего встречаются шпионские программы, которые передают не только сведения о предпочтениях пользователя в Интернете, но и его пароли, а также некоторые важные личные файлы.

Суть рекламных программ вытекает из их названия: эти компьютерные паразиты любым способом пытаются рекламировать продукты или услуги каких-то третьих компаний. Чтобы добиться своего, рекламные программы могут встроить рекламные объявления в какое-нибудь наиболее часто используемое приложение, например в web-браузер. Рекламные программы также используют всплывающие окна (pop-ups), в которых либо показывается объявление, либо сразу же загружается рекламная страница из Интернета. Часто таких окон открывается сразу несколько. Однако наиболее недобросовестным и трудноопределяемым способом рекламы является

подтасовывание результатов поиска в Интернете. Когда пользователь ищет что-то в поисковой машине, рекламная программа изменяет результаты поиска таким способом, чтобы наверху оказались ее собственные рекламные ссылки и объявления.

Программы дозвона - это компьютерные паразиты, которые пытаются с помощью модема и телефонной линии дозвониться до платного сервера, находящегося чаще всего в другой стране. Такие серверы обычно предоставляют различные услуги порнографического характера и берут со своих пользователей поминутную оплату. Если вредоносной программе дозвона удастся соединиться с таким удаленным сервером и продержат соединение несколько минут, то пользователю потом придет счет на оплату услуг порнографического сервера и международные переговоры. И этот счет придется оплатить, даже если предоставленные услуги оказаны не были.

### ***Серверы рассылки спама***

Спам это нежелательные электронные сообщения. Чтобы избежать ответственности за рассылку спама, злоумышленники не рассылают письма со своего компьютера. Они предпочитают заразить компьютеры других пользователей Интернета специальным "троянцем", который превратит чужой ПК в сервер рассылки спама. Злоумышленнику останется лишь указать троянской программе, какое письмо и по каким адресам следует рассылать. Ответственность за эти незаконные действия будет нести легальный пользователь зараженного компьютера.

На практике одного сервера рассылки спама преступнику не хватает, ему требуется целая сеть зараженных компьютеров. Такие сети управляемых злоумышленником компьютеров называют сетями зомби-машин или просто зомби-сетями. В типичную зомби-сеть входит около 3 тыс. компьютеров, находящихся по всему миру.

Чтобы заразить большое число компьютеров и превратить их в серверы рассылки спама, преступники комбинируют различные виды вредоносных кодов. Чаще всего объединяют сетевых червей и "троянцев". Сетевые черви распространяются очень быстро, что позволяет заразить довольно много компьютеров, а при попадании на машину-жертву червь не только рассылает повсюду свои копии, но и устанавливает "троянца", который, собственно, и превращает зараженный компьютер в зомби.

### ***Многокомпонентные "троянцы"-загрузчики***

Многокомпонентность таких вредоносных кодов заключается в том, что после заражения нового компьютера они переписывают из Интернета другой вредоносный код и внедряют его в систему. Например, червь из

предыдущего примера, который превращает компьютер в зомби, может не нести на себе "тройнца", а переписать его из Интернета уже после того, как ему самому удастся заразить компьютер.

Различают компоненты разного уровня. Например, если "тройнец" переписшет из сети другой вредоносный код, то это уже будет компонент второго уровня. Иногда бывает, что компонент второго уровня сам скачивает и внедряет еще одного паразита. Это уже компонент третьего уровня.

### *Административные меры борьбы с вирусами*

Говоря о степени ответственности антивирусной защиты, требуется разделять корпоративные и частные системы. Если речь идет об информационной безопасности организации, то необходимо позаботиться не только о технических (программных и аппаратных) средствах, но и об административных.

Если в некоторой компании есть сеть, не связанная с Интернетом, то вирус извне туда не проникнет, а чтобы вирус случайно не попал в корпоративную сеть изнутри, можно просто не давать пользователям возможности самостоятельно считывать носители информации, такие как CD-диски, USB-флэш или выходящие из употребления дискеты. Например, если кому-то из сотрудников необходимо считать что-либо с CD, он должен обратиться к администратору, который имеет право установить CD и считать данные. При этом за проникновение вирусов с этого CD уже несет ответственность администратор.

При нормальной организации безопасности в офисе именно администратор контролирует установку любого ПО; там же, где сотрудники бесконтрольно устанавливают софт, в сети рано или поздно появляются вирусы.

Большинство случаев проникновения вирусов в корпоративную сеть связано с выходом в Интернет с рабочей станции. Существуют режимные организации, где доступ к Интернету имеют только неподключенные к корпоративной сети станции. В коммерческих организациях такая система неоправданна. Там Интернет-канал защищается межсетевым экраном и прокси-сервером, о принципах работы которых будет сказано позднее. Во многих организациях разрабатывается политика, при которой пользователи имеют доступ лишь к тем ресурсам Интернета, которые нужны им для работы.

Конечно, поддержка политики жесткого разграничения прав доступа требует дополнительных инвестиций, а в ряде случаев приводит к замедлению выполнения некоторых работ. Поэтому каждая компания должна

искать для себя разумный компромисс, сопоставляя финансовые потери от порчи информации и замедления бизнес-процессов. В ситуации, когда документы содержат важные стратегические данные, например государственную тайну, именно степень ущерба в случае разглашения тайны определяет бюджет на меры безопасности.

Помимо антивирусной защиты важно не забывать о таком важном средстве защиты данных, как система резервного копирования. Резервное копирование является стратегическим компонентом защиты данных. Если данные уничтожены вирусом, но у администратора есть вовремя сделанная резервная копия, потери будут минимальными.

### *Принципы работы антивирусных программ*

В зависимости от технологии распространения данных менялись вирусы и принципы их работы. Различают следующие типы вирусного инфицирования:

- Заражение и распространение с помощью файлов на дискетах
- Использование в качестве носителя вируса файлов офисных пакетов в локальных сетях
- Использование почтовых вложений в сети Интернет для распространения вирусов.

В России первые вирусы появились в 1988 году, но до середины 90-х годов они не получили широкого распространения. В то время персональных компьютеров, имеющих доступ в Интернет, было очень мало. Пользователи обменивались друг с другом данными и программами с помощью дискет. Объемы информации были невелики, и емкости обычных дискет вполне хватало.

Таким же способом распространялись и вирусы. Попав на персональный компьютер, вирус быстро размножался. Но чтобы заразить еще один компьютер, вредителю приходилось долго ждать, пока инфицированный файл запишут на дискету и передадут другому пользователю. Борьба с вирусами тоже велась очень просто: пользователи защищали каждый персональный компьютер отдельно с помощью антивируса, который мог находить вирусы и лечить инфицированные файлы. Каждый пользователь знал, что перед тем, как переписать новые файлы с какой-либо дискеты, саму дискету надо обязательно проверить на вирусы.

Эти вирусы прикреплялись как правило к исполняемым файлам, но с появлением офисных пакетов сфера заражения файлов расширилась.

Благодаря операционной системе Microsoft Windows 95 и офисному пакету Microsoft Office значительно расширилась функциональность программ и возрос объем данных, например мультимедиа. Пользователям приходилось распространять информацию уже на компакт-дисках. К этому времени также возросло и число компьютеров, объединенных локальной сетью или имеющих доступ в Интернет. Пользуясь особенностями офисных документов вирусы стали прикрепляться к текстовым и мультимедийным файлам. Первая серьезная эпидемия, которая привлекла к проблеме вирусов внимание всего мирового сообщества, случилась в 1996-м году. Это был макровирус Car, распространявшийся в документах Microsoft Word. Ему удалось парализовать работу десятков компаний по всему миру.

Поворотной точкой стал 2000 год, когда появились эпидемии сетевых червей. В отличие от всех остальных компьютерных паразитов, они очень быстро распространялись в почтовых сообщениях. Число домашних пользователей и компаний, подключенных к Интернету, к 2000 году сильно возросло. Люди обменивались между собой электронными почтовыми сообщениями и еще не подозревали, какую это таит опасность. Первый современный сетевой червь был создан в самом начале 1999 года (это был Гарри99), а первый сетевой червь, вызвавший небывалую по тем временам эпидемию, появился в мае 2000 года. Это был червь, известный под двумя названиями: I love you и LoveLetter.

"Любовные письма" только в мае заразили более 40 млн. компьютеров по всему миру.

Исследовательский центр Computer Economics оценил убытки мировой экономики за первые пять дней эпидемии этого паразита в 6,7 млрд. долл. Червь рассылал свои копии сразу же после заражения системы по всем адресам электронной почты, найденным в адресной книге почтовой программы Microsoft Outlook. В качестве адреса отправителя червь указывал адрес электронной почты владельца зараженного компьютера и адресной книги Outlook. К каждому сообщению червь прикреплял вредоносный файл с расширением "txt.vbs". Такой ход дезориентировал пользователей, так как расширение "txt" соответствует текстовым файлам, а "vbs" - программам на языке Visual Basic Script. Текст электронного письма представлял собой любовное послание, приглашающее пользователя открыть прикрепленный инфицированный файл.

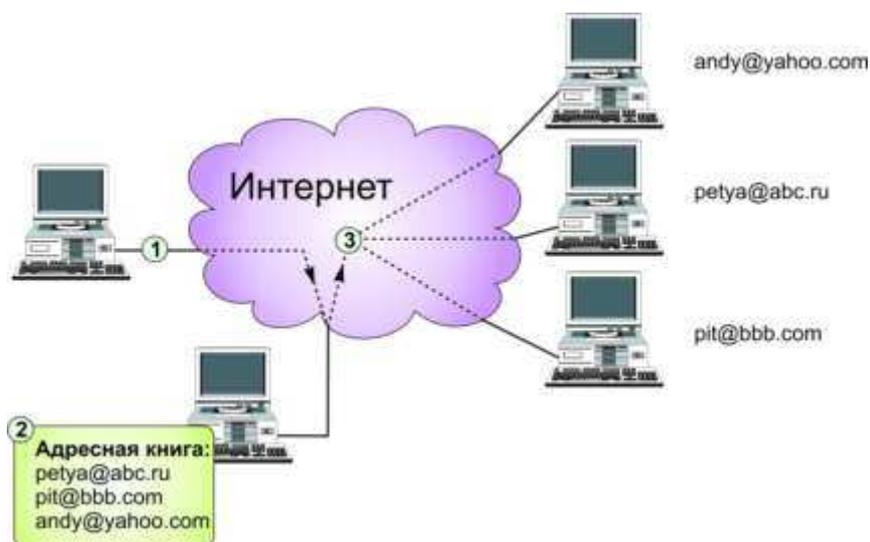


Схема рассылки вирусов с использованием адресной книги зараженного компьютера. С зараженного компьютера (пункт 1) посредством электронной почты червь попадает на компьютер (пункт 2) и использует данные его адресной книги. В результате (пункт 3) будет атаковано сразу несколько компьютеров - заражение идет в геометрической прогрессии.

Электронная почта и Интернет оказались идеальной средой для передачи вредоносных кодов. Сегодня 98% всех вредителей попадает на компьютер именно через электронную почту. Через электронную почту распространяются не только сетевые черви, но также "трояницы" и файловые вирусы.

### ***Брандмауэр как средство защиты частных сетей***

Межсетевые экраны, называют также брандмауэрами или файерволами (от англ. firewall). Интересно происхождение данного термина. Брандмауэрами называли специальные устройства в поездах. В машинном отделении паровозов топливо находилось поблизости от топки. Кочегар бросал лопатой уголь в топку, и легковоспламеняющаяся угольная пыль часто вспыхивала. Из-за этого в машинном отделении нередко возникали пожары. Чтобы огонь не распространялся на пассажирские вагоны, позади машинного отделения начали устанавливать железные загородки. Именно этот щит получил название "брандмауэр".

По прошествии времени значение термина "брандмауэр" изменилось. Первые компьютерные брандмауэры были созданы для того, чтобы препятствовать распространению сетевого программного обеспечения, содержащего множество ошибок, на всю сеть с одного ее участка. Подобно своим железнодорожным прототипам, они были средством локализации "пожара".

Сегодня брандмауэры выступают в роли защитников границ между локальными сетями и Интернетом. Персональные брандмауэры выполняют те же функции, но на границе между домашним компьютером и Интернетом.

Брандмауэр, или межсетевой экран,- это система, предотвращающая несанкционированный доступ извне во внутреннюю сеть. Брандмауэры бывают аппаратными или программными. Аппаратный брандмауэр - это устройство, которое подключается к сети физически, фильтрует входящий и исходящий трафик и защищает от нежелательных проникновений во внутреннюю сеть или на персональный компьютер. Программный брандмауэр выполняет те же функции, но является не внешним аппаратным устройством, а программой, установленной на компьютере. В роли параметров фильтрации выступают адреса получателя и отправителя каждого сетевого пакета, протокол передачи данных (например, HTTP, FTP и т.д.), приложение, которое отправляет или принимает сетевой пакет и т.д.

Принцип работы брандмауэра проиллюстрирован на рисунке ниже. Несанкционированный пользователь не сможет получить доступ в локальную сеть, если ее защищает брандмауэр.



Брандмауэр запрещает доступ в локальную сеть из Интернета пользователям, не имеющим соответствующих полномочий

Брандмауэр защищает частную сеть и отфильтровывает те данные, обмен которыми запрещен. Если в компании есть, например, 100 персональных компьютеров, объединенных в локальную сеть и имеющих выход в Интернет, но нет брандмауэра, то злоумышленник сможет проникнуть на каждый из этих компьютеров из Интернета.

Брандмауэры, предназначенные для защиты корпоративной сети, часто имеют встроенные проху-сервер и систему обнаружения вторжений.

Проху-сервер играет роль посредника между внутренней сетью организации и Интернетом. Сервер-посредник кэширует (сохраняет) часто запрашиваемые web-страницы в своей памяти. Когда пользователь запрашивает какую-нибудь страницу из Интернета, проху-сервер проверяет, есть ли она в его базе данных. Если есть, то страница сразу же отправляется к

пользователю. Если нет, то проху-сервер запрашивает оригинальный сервер, где размещена страница, и, получив ее, отправляет пользователю. Механизм сохранения часто запрашиваемой информации позволяет значительно сэкономить время доступа к наиболее важным данным.

Еще одной функцией проху-сервера часто является трансляция сетевых адресов (NAT - Network Address Translation). Ее суть в том, чтобы сделать компьютеры внутренней сети организации невидимыми для внешних запросов. Если злоумышленник попытается "заглянуть" во внутреннюю сеть компании, он увидит один лишь проху-сервер (в данном случае еще и брандмауэр). Он не сможет узнать внутренние адреса компьютеров, а следовательно, вторгнуться в корпоративную сеть будет значительно сложнее. Естественно, механизм трансляции адресов немного замедляет работу всей защищаемой сети.

Системы обнаружения вторжений, даже являясь составной частью крупного брандмауэра, дополняют другие системы информационной безопасности. Они не только определяют сам факт проникновения в сеть, но и выявляют подозрительные действия.

Если брандмауэры рассматривать как забор с калиткой, через которую могут пройти те, кто наделен соответствующими полномочиями, система обнаружения будет выступать здесь в роли устройств внешнего видеонаблюдения и охранной сигнализации. Охранная система включается, когда злоумышленник перелез через забор или сломал калитку и теперь намеревается захватить центральный пульт управления. То есть когда хакер уже проник внутрь и готовится поразить жизненно важную систему.

Работа системы обнаружения вторжений строится на законах математической статистики. Каждое действие, происходящее в системе, подвергается анализу на соответствие сценарию сетевой атаки. Так как действия злоумышленника разнятся от случая к случаю, системе обнаружения вторжений приходится учитывать отклонения реально происходящих событий от сценария нападения.

### ***Персональные брандмауэры***

Персональные брандмауэры защищают отдельные автономные компьютеры, подсоединенные к Интернету. Чаще всего персональные брандмауэры используются на домашних ПК. Основная задача этого средства защиты - фильтровать входящий и исходящий сетевой трафик (поток данных), контролировать сетевую активность приложений и блокировать любые опасные действия.

Персональный брандмауэр умеет фильтровать входящие и исходящие соединения по целому ряду признаков. Это прежде всего адреса абонентов, используемый для соединения порт (число, которое идентифицирует процесс или приложение внутри компьютера) и полномочия приложения, осуществляющего обмен информацией. Есть и более сложные способы фильтрации. Например, при анализе входящих соединений брандмауэр всегда может проверить, запрашивало ли какое-нибудь приложение соединение с данным узлом. Если нет, то входящее соединение нужно запретить, а если да, значит, на персональный компьютер просто пришел ответ на посланный ранее запрос. Для того чтобы эффективно фильтровать трафик, в брандмауэре должна быть реализована поддержка большого числа протоколов и технологий.

В операционной системе Microsoft Windows XP имеется свой встроенный брандмауэр.

### *Брандмауэр, встроенный в Microsoft Windows XP*

Брандмауэр Windows представляет собой фильтр сетевых пакетов (а информация в сети передается именно в пакетах), способный отразить стандартные сетевые атаки и не допустить низкоуровневого сетевого подключения к защищенному компьютеру. Также поставляемый по умолчанию с Microsoft Windows XP брандмауэр умеет корректно обрабатывать диагностические и служебные пакеты, приходящие из Интернета. Между тем служебные пакеты чаще всего используются для организации сетевых атак. Благодаря брандмауэру Windows компьютер пользователя намного сложнее обнаружить в Интернете и атаковать.

В операционную систему Microsoft Windows XP Service Pack 2 (пакет обновлений 2) включен "Центр обеспечения безопасности", который следит за тем, чтобы встроенный в ОС брандмауэр был включен и правильно настроен.

Однако для защиты компьютера под управлением Microsoft Windows необязательно использовать встроенный в операционную систему брандмауэр. Сами разработчики Microsoft рекомендуют попробовать и другие решения, а потом сделать выбор.

### *Банковские АИС*

Современные БИТ охватывают следующие направления банковской деятельности:

- клиринговые операции (взаиморасчеты банков);

- торговые операции и маркетинг;
- управление деятельностью банка;
- кредитные операции, включая анализ заявок клиентов на их кредитоспособность;
- использование банковских автоматов и др.

Автоматизация банковских операций позволяет:

- осуществлять безбумажные платежные операции с минимальным привлечением труда людей и сокращением операционных расходов;
- осуществлять обработку платежей преимущественно в реальном времени;
- ускорять обмен информацией между банками, банками и клиентами, банками и их отделениями с помощью телекоммуникационных линий связи;
- минимизировать типичные виды банковского риска – потери документов, ошибочную адресацию, фальсификацию платежных документов;
- обеспечивать руководителей стратегическими оценками положения банка.

Для реализации этих функций в настоящее время разработано множество аппаратно – программных комплексов. Наиболее известными из них являются :система электронного перевода средств SWIFT, представляющая собой защищенную международную коммуникационную сеть. Система платежей для клиринговых операций (CNAPS) – позволяет производить межбанковские расчеты день в день.

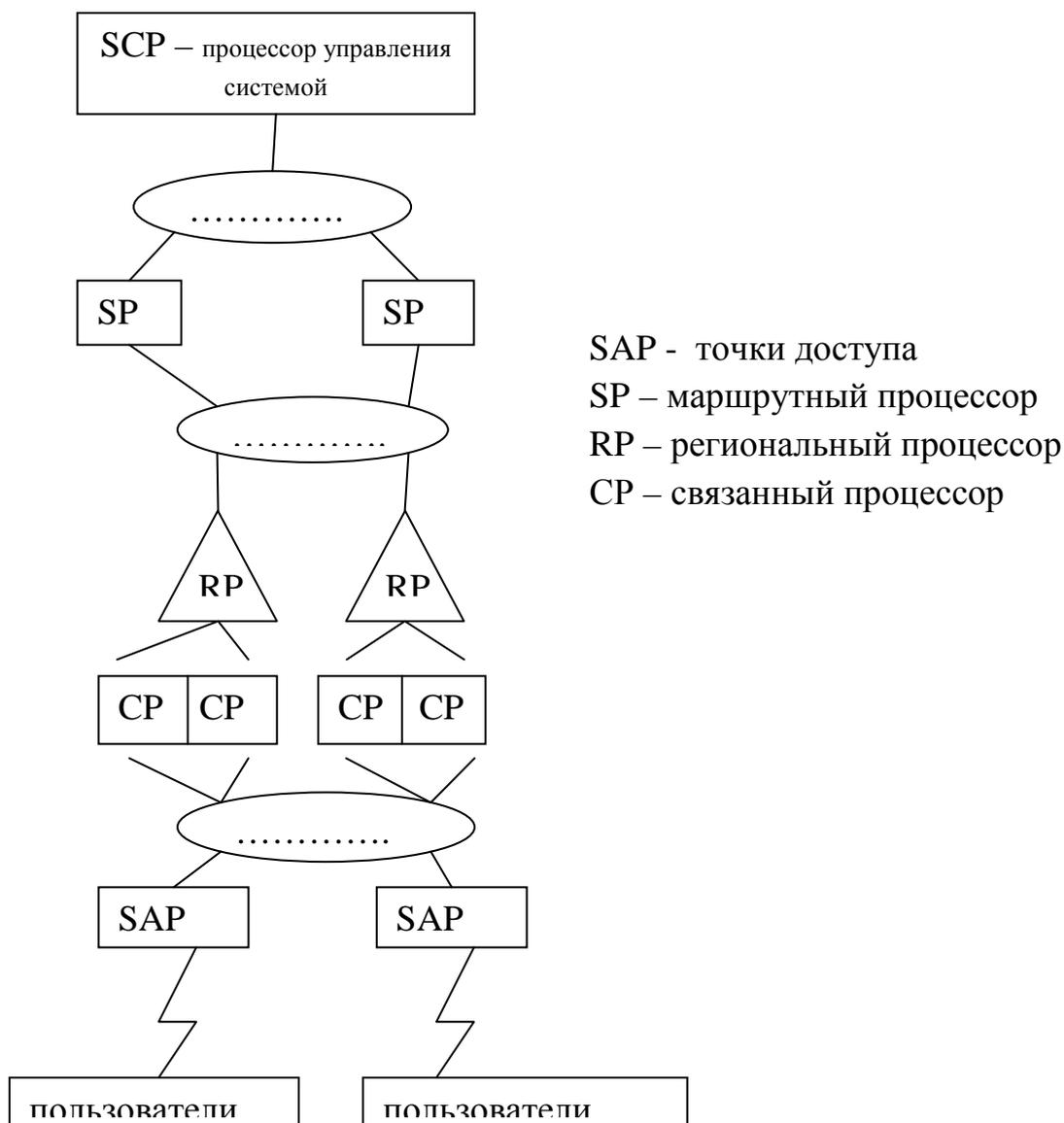
### ***Международная банковская сеть SWIFT***

В настоящее время ее членами являются более 3000 финансовых организаций из более чем 80-ти стран.

- Повышение эффективности работы банков за счет стандартизации и современных способов передачи информации.
- Обеспечение надежности при передаче сообщения ( кодирование и специальный порядок передачи и приема).
- Прямой доступ банков, участников SWIFT к своим корреспондентам, отделениям и филиалам.

Обмен информацией в сети ведется стандартизированными сообщениями. При поступлении сообщения, система проверяет его правильность и соответствие стандарта. Это гарантирует правильность его передач. Система сохраняет копию каждого сообщения, что обеспечивает большую безопасность.

Архитектура SWIFT . Сеть имеет следующую структуру:



Абоненты подключаются к SAP либо по выделенной линии, либо по телефонному каналу. Местная сеть связанных процессоров обеспечивает соединение региональных процессоров с сетью точек доступа. Кроме того,

СР связывают RP с другими узлами сети. Это дает возможность RP связываться со своим маршрутным процессором и в тоже время дает возможность получать сообщения от других SP.

Региональные процессоры выполняют функции точек входа в SWIFT. Весь доступ пользователей к системе осуществляется при помощи этих узлов. Программное обеспечение, работающее на RP, связывается с обеспечением пользователя для правильности и безопасного подключения к сети.

Все входящие сообщения перед передачей на следующий уровень проверяются на правильность и в соответствии со стандартами SWIFT.

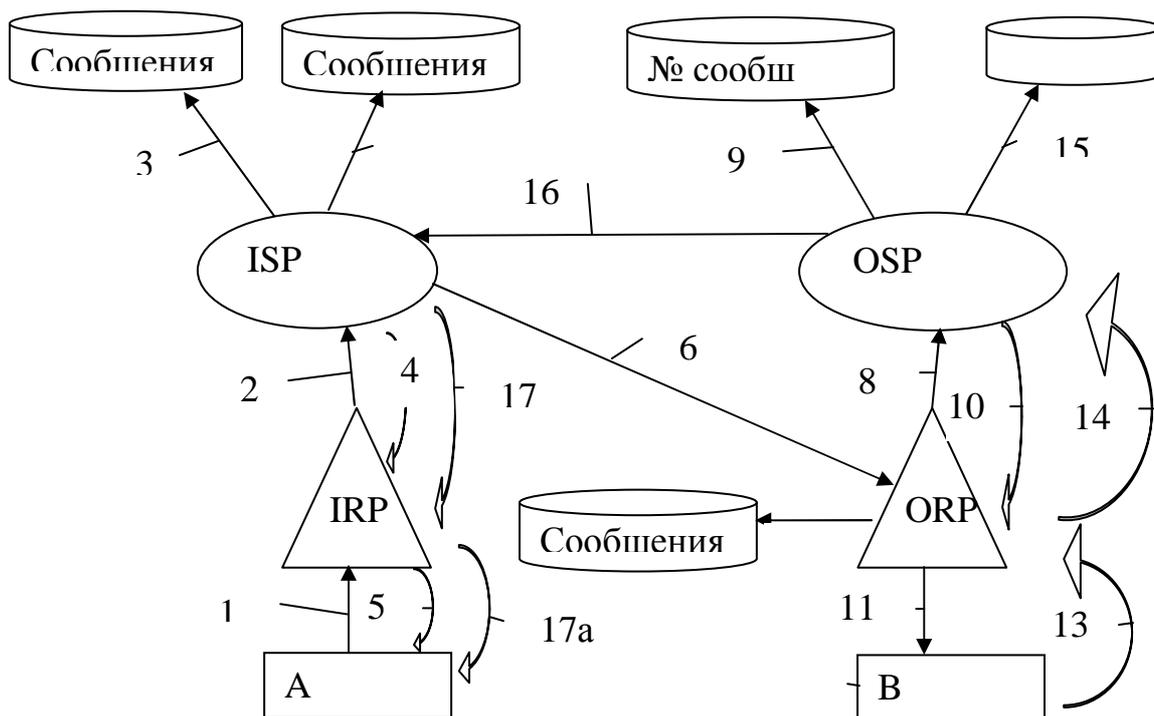
Выходящие сообщения хранятся на соответствующем RP, который обеспечивает их передачу пользователям.

Все пользователи представлены своими адресами, хранящимися в своем SP. Маршрутные процессоры управляют передачей и хранением сообщений. На них лежит основная вычислительная нагрузка сети.

Процессор SCP управляет всеми компонентами системы и следит за всеми попытками доступа к ней.

### *Маршрутизация сообщений в сети.*

Весь процесс передачи сообщения от пользователя А к пользователю В по сети SWIFT можно представить в виде последовательности шагов, указанных в следующей схеме:



Весь процесс передачи сообщений от пользователя А к пользователю В по сети S. W. I. F. T. можно представить в виде последовательностей шагов:

1. получив доступ к сети S. W. I. F. T. пользователь А посылает сообщение для пользователя В на свой RP (входной)
2. Входной RP проверяет правильность заголовка и текста сообщения. После этого результаты проверки и само сообщение пересылаются на соответствующий SP.
3. SP записывает полученное сообщение на диск.
4. SP направляется в адрес RP подтверждение получения и записи сообщения.
5. Получив подтверждение от SP, RP отправляет пользователю А положительное или отрицательное подтверждение, в зависимости от того принято или отвергнуто сообщение. Положительное подтверждение свидетельствует о том, что система в W. I. F. T. взяла на себя ответственность за доставку сообщения, отрицательное – что сообщение не может быть доставлено.
6. Получив сообщение входной SP определяет на основании своей базы данных, какой RP является основным для пользователя В, и отправляет копию сообщения на этот RP.
7. RP получатель временно записывая сообщения на диск и помещает его в одну из очередей доставки пользователю В. Здесь сообщение находится до тех пор пока пользователь В не подключится к сети и не пошлёт запрос на получение сообщений из этой очереди.
8. Перед попыткой доставки сообщения RP- получатель присваивает ему номер и запрашивает у своего SP разрешение на доставку.
9. SP – получатель проверяет правильность номера сообщения для доставки и записывает его на диск.
- 10.SP – получатель разрешает доставку сообщений от RP к получателю В.
- 11.RP – получатель отправляет сообщение пользователю В.
- 12.Пользователь получает сообщение и записывает его для дальнейшего использования
- 13.Если терминальное оборудование пользователя подтверждает правильность полученного сообщения (совпадает контрольная сумма), на основной для пользователя В RP передаётся положительное подтверждение приёма. Если контрольные суммы не совпадают, на RP подаётся отрицательное подтверждение, и сообщение считается не доставленным.
- 14.RP – получатель создаёт из положительных и отрицательных подтверждений историю доставки сообщений и передаёт её своему SP.
- 15.SP – дополняет общую историю доставки сообщения и сохраняет её на диске.
- 16.SP – приёмник посылает копию истории доставки SP – передатчику, который сохраняет её на диске.

17. Если пользователь А запрашивает подтверждение доставки сообщения, то после получения от SP приёмника истории доставки SP – передатчик посылает соответствующее подтверждение входному RP для передачи пользователю.