

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНА МЕТАЛУРГІЙНА АКАДЕМІЯ УКРАЇНИ**

**Г. Г. Швачич, О. В. Овсянніков, Л.М.Петречук**

## **Електронне документознавство**

для студентів спеціальності «Документознавство та інформаційна діяльність»

**КОНСПЕКТ ЛЕКЦІЙ**

**Дніпропетровськ НМетАУ 2013**

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНА МЕТАЛУРГІЙНА АКАДЕМІЯ УКРАЇНИ**

**Г. Г. Швачич, О. В. Овсянніков, Л.М.Петречук**

## **Електронне документознавство**

для студентів спеціальності «Документознавство та інформаційна діяльність»

Затверджено на засіданні Вченої ради академії  
як конспект лекцій

**Дніпропетровськ НМетАУ 2013**

УДК 004 (075.8)

Г.Г. Швачич, О.В. Овсянніков, Л.М.Петречук. Електронне документознавство. Конспект лекцій. – Дніпропетровськ: – НМетАУ, 2013. – 90 с.

Викладені основи створення, ведення, зберігання інформації у електронному вигляді. Надана методика створення й ведення електронних документів.

Призначений для студентів спеціальності «Електронне документознавство.».

Іл. 31. Бібліогр.: 5 найм.

Відповідальний за випуск

Г.Г. Швачич, канд. техн. наук, проф.

Рецензенти: Б. И. Мороз д-р техн. наук, проф. (Академія митної Служби України)

Т. И. Пашова канд. техн. наук, доц. (Дніпропетровський державний аграрний університет)

© Національна металургійна академія України, 2012

## Основные понятия электронного документооборота

Потоки информации, циркулирующие в окружающем нас мире, огромны. Во времени они имеют тенденцию к увеличению. В этих потоках имеются документы, содержащие самую разнообразную информацию. Документы сопровождают нас на каждом шагу. Ежедневно в мире создаются миллиарды больших и малых документов и их копий. На производство и воспроизведение документов расходуется огромное количество леса, а на их проверку и хранение требуется большое количество времени. Современное общество не может существовать без документооборота.

Документооборот — движение документов в организации с момента их создания или получения до завершения исполнения или отправления.

Комплекс работ с документами:

- приём,
- регистрация,
- рассылка,
- контроль исполнения,
- формирование дел,
- хранение и повторное использование документации,
- справочная работа.

В настоящее время руководящему звену любой организации постоянно требуется иметь в своем распоряжении обширную и достоверную информацию. Причем, с каждым годом информационные потоки (ИП) все более разрастаются, и могут «задавить» своим давлением.

Эффективная работа с ИП достигается только с помощью средств и методов их автоматизации. Современная вычислительная техника (и технологии) становится удобной базой для решения таких распределенных многопользовательских задач, как электронный документооборот (ЭДО).

Информационный поток — совокупность информации, рассматриваемая в процессе ее движения в пространстве и времени в определенном направлении.

Машиночитаемый документ — документ, пригодный для автоматического считывания содержащейся в нём информации, записанный на магнитных, оптических и других носителях информации.

Электронный документ (ЭД) — документ, созданный с помощью средств компьютерной обработки информации, который может быть подписан электронной цифровой подписью (ЭЦП) и сохранён на машинном носителе в виде файла соответствующего формата. На сегодня в мире существует только один ЗРЕЛЫЙ стандарт на формат электронного – документа – это ODF, который в 2006 году принят под № ISO 26300. Полное описание формата занимает 738 страниц. Компания Microsoft с большим напором продвигала свой формат OOXML и 1 апреля 2008 г было объявлено о принятии спецификации MS

OOXML как стандарта ISO/IEC DIS 29500, однако эта спецификация содержит около 300 нерешенных проблем технического характера.

Электронная цифровая подпись (ЭЦП) — информация в электронной форме, которая присоединена к другой информации в электронной форме (подписываемой информации) или иным образом связана с такой информацией и которая используется для определения лица, подписывающего информацию. ЭЦП аналог собственноручной подписи, являющийся средством защиты информации, обеспечивающим возможность контроля целостности и подтверждения подлинности электронных документов.

Электронная цифровая подпись – реквизит электронного документа, позволяющий установить отсутствие искажения информации в электронном документе с момента формирования ЭП и проверить принадлежность подписи владельцу сертификата ключа электронной подписи (сертификат ключа – цифровой или бумажный документ, содержащий информацию о владельце ключа, название центра сертификации и т.п.). Значение реквизита получается в результате криптографического преобразования информации с использованием закрытого ключа ЭП.

Электронный документооборот (ЭДО) — единый механизм по работе с документами, представленными в электронном виде, с реализацией концепции «безбумажного делопроизводства», - это совокупность технологий, которые не только значительно оптимизируют, но и существенно меняют работу учреждения

Электронный документооборот включает: создание документов, их обработку, передачу, сохранение, вывод (иногда ввод) информации, на основе использования компьютерных сетей.

Системы электронного документооборота кроме процессов создания, управления, доступа и распространения больших объемов документов в компьютерных сетях, также обеспечивают контроль над потоками документов в организации. Часто эти документы хранятся в специальных хранилищах или в иерархии файловой системы. Типы файлов, как правило, поддерживающие системы электронного документооборота включают: текстовые документы, изображения, электронные таблицы, аудио-, видеоданные и Web-документы.

Под управлением электронным документооборотом в общем случае принято понимать организацию движения документов между подразделениями предприятия или организации, группами пользователей или отдельных пользователей. При этом, под движением документов имеется в виду не их физическое перемещение, а передача прав на их приложение с сообщением конкретным пользователям и контролем за их выполнением.

Совместное использование систем электронного делопроизводства и хранилищ информации позволяет систематизировать и объединять информацию, что облегчает ее анализ и составление отчетов. Для поиска скрытых закономерностей в больших массивах данных применяются решения и действия, базирующейся на соответствующих технологиях извлечения информации из данных (мы будем пользоваться процессором Excel).

Хранилище данных ( Data Warehouse) — предметно-ориентированная

информационная база данных, специально разработанная и предназначенная для подготовки отчетов и бизнес-анализа с целью поддержки принятия решений в организации. Строится на базе систем управления базами данных и систем поддержки принятия решений. Данные, поступающие в хранилище данных, как правило, доступны только для чтения. Данные из OLTP-системы копируются в хранилище данных таким образом, чтобы построение отчетов и OLAP-анализ не использовал ресурсы транзакционной системы и не нарушал её стабильность.

OLTP (Online Transaction Processing), транзакционная система — обработка транзакций в реальном времени, выполняет ввод, структурированное хранение и обработку информации (операций, документов) в режиме реального времени.

Реальное время — режим работы автоматизированной системы обработки информации и управления, при котором учитываются ограничения на временные характеристики функционирования.

OLAP (англ. online analytical processing, аналитическая обработка в реальном времени) — технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу.

К общим возможностям систем электронного документооборота относятся:

- создание документов,
- управление доступом,
- конвертация данных,
- обеспечение безопасности данных.

В системах электронного документооборота должны автоматически отслеживаться изменения в документах, сроки исполнения документов, движение документов, а также контролироваться все их версии и подверсии.

Комплексная система электронного документооборота — это система, охватывающая весь цикл делопроизводства организации (от постановки задачи на создание документа до его списания в архив), обеспечивающая централизованное хранение документов в любых форматах. В системах электронного документооборота должно быть реализовано жесткое разграничение доступа пользователей к различным документам в зависимости от их полномочий.

В большинстве систем электронного документооборота реализована иерархическая система хранения документов. Также в ряде систем электронного документооборота реализованы возможности сохранения за счет организации связей между документами.

### **Основные принципы электронного документооборота**

- Однократная регистрация документа, позволяющая однозначно идентифицировать документ.
- Возможность параллельного выполнения операций, позволяющая сократить время движения документов и повышения оперативности их исполнения
- Непрерывность движения документа, позволяющая идентифицировать ответственного за исполнение документа (задачи) в каждый момент времени жизни документа (процесса).

- Единая (или согласованная распределённая) база документной информации, позволяющая исключить возможность дублирования документов.
- Эффективно организованная система поиска документа, позволяющая находить документ, обладая минимальной информацией о нём.
- Развитая система отчётности по различным статусам и атрибутам документов, позволяющая контролировать движение документов по процессам документооборота и принимать управленческие решения, основываясь на данных из отчётов.

## **Системы управления документами**

Система управления документами, СУД, DMS (англ. Document management system) — компьютерная система (или набор компьютерных программ), используемая для отслеживания и хранения электронных документов и/или образов (изображений и иных артефактов) бумажных документов.

Данное понятие тесно связано с концепцией Content Management System (система управления содержимым) и обычно рассматривается как компонент Enterprise Content Management System (CMS уровня предприятия).

В общем случае системы управления документами (DMS) предоставляют хранение, версионирование, пометку метаданными и безопасность по отношению к документам также хорошо, как и индексирование и развитые возможности поиска документа.

### Сервисы DMS

- Метаданные
- Интеграция
- Захват
- Индексирование
- Хранилище
- Развитые возможности поиска
- Известные системы

### **Сервисы DMS**

**Метаданные.** Метаданные обычно хранятся для каждого документа. Метаданные, например, могут включать дату занесения документа в хранилище и идентификатор пользователя, совершившего это действие. DMS также может извлекать метаданные из документа автоматически или запрашивать их у пользователя. Некоторые системы предоставляют сервис оптического распознавания текста сканированных документов, или извлекают текст из электронных документов. Используя извлечённый текст система позволяет производить поиск документа по ключевым словам внутри документа.

**Интеграция.** Многие системы управления документами пытаются интегрировать функцию управления документами непосредственно в различные

приложения, позволяя пользователю получать документ сразу из хранилища системы управления документами, делать какие-либо модификации, и сохранять его обратно в хранилище в качестве новой версии, и всё это проделывать в одном приложении, не выходя из него. Данная интеграция, в основном, доступна для офисных пакетов и почтовых клиентов или для программного обеспечения, предназначенного для групповой/коллективной работы. Интеграция обычно подразумевает использование открытых стандартов та-ких как: ODMA, LDAP, WebDAV и SOAP.

**Захват.** Перевод в цифру бумажных документов, используя сканеры и МФУ. Также часто используется программное обеспечение для оптического распознавания текста, чтобы конвертировать цифровые изображения в машиночитаемый текст.

**Индексирование.** Индексирование предоставляет возможность классифицировать документы посредством метаданных и словарного индекса текста, извлечённого из документа. Индексация существует, главным образом, для поддержки развитых возможностей поиска документов. Одно из главных условий быстрого и качественного поиска — это создание индекса документа.

**Хранилище.** Хранит электронные документы. Хранилище документов также включает в себя и управление тех же самых документов, которое оно хранит; также хранилище обеспечивает миграцию с одного носителя на другой и обеспечивает целостность данных.

Хранилище документов может представлять собой как файловое хранилище, так и хранилище в виде СУБД (базы данных). В свою очередь, хранилище документов в СУБД может производиться как в одной (единой) базе данных, так и в отдельных базах данных.

## **Классификация систем электронного документооборота**

### 1. Универсальные «коробочные» СЭДО:

- стандартный набор функций;
- невозможность полного соответствия потребностям конкретной организации;
- низкие временные затраты на приобретение и установку;
- относительно низкая стоимость;
- необходимость приобретения лицензии на каждое внедряемое рабочее место.

*Пример: Дело*

### 2. Индивидуально разрабатываемые СЭДО:

- максимально персонифицированная система;
- большие временные затраты;
- высокая стоимость разработки;
- сопутствующие расходы: затраты на обучение сотрудников, покупку нового оборудования и программного обеспечения.

*Пример: DIRECTUM*

### 3. Комбинированные СЭДО:

- базовая платформа, к которой разрабатываются необходимые дополнительные модули;
- полное соответствие нуждам предприятия;
- небольшие временные затраты на разработку и внедрение;
- стоимость включает: цену базовой платформы и стоимость индивидуальной доработки, зависящей от сложности заказа;
- передача заказчику прав на продукт;
- простота освоения и использования;
- полная русификация;
- удобный интерфейс;
- взаимодействие с существующими офисными приложениями.

Системы электронного документооборота: ЭлектроОфис, «4К-EDocument» Version 1.3, 1С:Документооборот, Канцлер,

Кодекс: Документооборот, МС-СКЛАД - Приход-расход товара в аптеках

Внедрение систем электронного документооборота позволяет получить следующие результаты:

1. Обеспечения более эффективного управления документами за счет автоматического контроля выполнения, прозрачности деятельности организации на всех уровнях;
2. Поддержка эффективного накопления, управления и доступа к информации;
3. Обеспечение четкой авторизации доступа к информации;
4. Оптимизация процессов анализа и контроля выполнения работы сотрудников.

Основными пользователями систем электронного документооборота, наряду с государственными организациями, предприятиями, банками, промышленными предприятиями, являются также учебные заведения. Поэтому, целесообразно рассмотреть особенности организации документооборота учебного процесса и внедрение новейших технологий в системах документооборота учебного заведения.

### **Организация информации в Microsoft Outlook.**

Рассмотрим в качестве примера *универсальной* системы электронного документа оборота (СЭДО) стандартное приложение пакета Microsoft Office – Outlook.

Microsoft Outlook предназначен для управления перепиской, личными сведениями и организацией рабочей деятельности. Это универсальная записная книжка, электронный ежедневник, средство автоматизации процедур планирования и контроля рабочей деятельности.

Пользуясь Outlook пользователь может:

- принимать, отправлять и пересылать почтовые сообщения,
- вести адресную книгу абонентов,
- сортировать полученную информацию,
- работать с факсимильными сообщениями.

Интерфейс Outlook стандартный как у всех Windows –приложений: строка заголовка, строка главного меню, устанавливаемые панели инструментов).

Действия, совершаемые и сохраняемые с помощью приложения, выполняются в логически соответствующих папках:

- задачи,
- дневник,
- заметки,
- календарь,
- контакты, и др.

Объекты (записи, строки), создаваемые в папках принято называть элементами.

Перечень папок Outlook доступен из меню Вид → Область переходов (или Alt+F1), здесь же осуществляется переход по папкам.

#### ПАПКА КАЛЕНДАРЬ

Внешний вид рабочей области папки Календарь напоминает страницу бумажного ежедневника. Рабочий лист каждой из папок Outlook может иметь несколько представлений. Один из способов изменить текущее представление - меню Вид → Текущее представление или соответствующим элементом на панели инструментов.

Представление – это отображение определенной выборки элементов в специальном графическом формате (формат отображения и предоставления информации). Представления изменяют способ просмотра данных и помогают увидеть информацию так, как это необходимо в данный момент.

В соответствии с поставленной задачей выбирается определенное представление (День/Неделя/Месяц) и заполняется элемент календаря (толи почасовое заполнение, толи каждодневное). Двойной щелчок мышкой по выбранному полю заполнения открывает окно детального заполнения элемента.

Обязательные для заполнения поля:

- Тема
- Место
- Начало : Конец
- Категория.

Категория – это средство Outlook для упорядочивания элементов и структурирования информации

После заполнения окна данными, необходимо выполнить сохранение, нажав на П/И кнопку **Сохранить и Заккрыть**. Удаляется текущий(выбранный) элемент клавишей Delete или кнопкой Удалить в окне детального заполнения.

Любое резервирование времени в расписании в папке Календарь называется Встречей или Событием.

Встреча – это мероприятие, для которого не привлекаются специальные ресурсы.

Событие – то мероприятие, продолжительностью более суток : выставки-ярмарки, отпуск, Олимпийские игры.

*Задание:* Заполнить на ближайшие три месяца календарь своих учебных и оздоровительных мероприятий (минимум три категории). Периодические мероприятия должны быть занесены технически *грамотно!*

### ДОБАВЛЕНИЕ (УДАЛЕНИЕ) ПРАЗДНИКОВ

Сервис→Параметры→Настройки→Параметры календаря→Параметры календаря → Добавить праздники.

Чтобы удалить отображение праздников в календаре необходимо перейти из представления День/Неделя/Месяц в представление По категориям и вызвав контекстное меню на папке Категории : Праздники – удалить её.

*Задание:* установить в календаре праздники России.

### ДОБАВЛЕНИЕ ЧАСОВЫХ ПОЯСОВ

Сервис→Параметры→Настройки→Параметры календаря→Параметры календаря → Часовой пояс. При установке отображения нескольких часовых поясов во избежание неясностей необходимо заполнять поля Метка (напр.: для российского региона - Рос, для Украины – Укр).

Изменение часового пояса можно выполнить через К/М, вызванное на поле отображения часового пояса.

Изменение разбивки часовой сетки также выполняется через К/М, вызванное на поле отображения часового пояса.

### ПАПКА ЗАДАЧИ

Задачей в терминологии Outlook называется поручение личного или служебного характера, выполнение которого можно проследить. Задача может быть разовой или повторяющейся.

После открытия папки Задачи, элементы в ней создаются аналогично элементам папки Календаря.

Поля для заполнения:

- тема,
- дата начала (сегодня, завтра, неделя),
- срок (сегодня, месяц, неделя), имеется в виду конечная дата выполнения задачи,
- состояние, важность, готово,
- категория,

- контакты (визитки), если папка Контакты имеет наполнение, на дополнительной вкладке Подробно
- объем работ, реально затрачено,
- дата завершения (поле содержит фактическую дату завершения задачи, которая никак не связана со сроком и датой начала задачи),
- организация,
- расстояние, расходы.

Постановка любой задачи может предполагать наличие разъяснений, инструкций или другой дополнительной информации, которая вводится в поле Заметки при нажатии кнопки Вложить файл (иконка в виде скрепки).

Повторяющиеся задачи создаются аналогично повторяющимся встречам (событиям), но для задач устанавливаются два принципиально разных режима повторения. В первом режиме повторение происходит в заранее фиксированные дни, а во втором – привязывается к дате завершения предыдущей задачи.

Задание: создать в папке Задачи новые элементы, отображающие Ваши планы и ход их выполнения (напр. каждые 24 дня проходит семинар по направлению «Новые технологии ведения «Электрон.документооборота»», или предстоит поездка в Париж на междунар. студенческий симпозиум, или вы посещаете курс занятий по обучению иностранным языкам и т.п.)

## ПАПКА КОНТАКТЫ

Элемент контакт содержит около 100 стандартных полей со сведениями о корреспонденте. При создании элемента папки Контакт практически все поля должны быть заполнены. Если возле поля установлена кнопка, то заполнение данных желательно проводить после нажатия кнопки.

Адрес эл. почты : user@mail.ru,

Веб-страница: <http://www.borey.ru>

### **Методология построения систем автоматизации документооборота.**

#### **Концепция системы управления содержимым -**

#### **ContentManagementSystem. Сервисы DMS - Document management system.**

Деятельность любой организации всегда можно рассматривать под определенным углом зрения, и ее можно описать как *последовательность порождения, восприятия, модификации и сохранения определенного набора документов*. В зависимости от пристрастий и специфики описываемого программного продукта в прессе можно встретить следующие соответствия между термином “документоведение” и западными терминами:

**DMS (Document Management Systems)** – система управления

документами, западный термин **EDMS - Electronic (Enterprise) Document Management System**. Такое название основывается на утверждении, что и делопроизводство и документооборот являются всего лишь частным случаем более общего понятия "управление документами". Данный класс систем считается универсальным, т.е. автоматизирующим весь комплекс задач, возлагаемых на делопроизводство, от разработки и создания проекта документа, до списания в дело, включая документооборот и хранение документов. EDMS-системы являются неотъемлемой частью мирового рынка ПО управления электронным документооборотом и содержимым информационных систем, так называемого **DCT (document and content technologies)** рынка. На рынке предлагается более 500 систем EDMS. Технологически они различаются по способам индексирования и поиска информации. Основным способом ввода документов в систему – сканирование.

На рынке коммерческие системы EDMS третьего поколения представлены программным продуктом Excalibur EFS (Excalibur Technologies Corp.). Пакет Excalibur EFS базируется на технологии адаптивного распознавания образов APRP (Adaptive Pattern Recognition Processing), разработанной компанией Excalibur Technologies и реализованной в пакете с использованием механизма нейронных сетей.

**DocFlow системы** или системы маршрутизации документов - workflow-системы. Применяется также термин **WMS - Workflow Management System**. В основу функционирования таких систем положено понятие потока работ, именно в том смысле, в котором в BPR (business process reengineering) определяется бизнес-процесс (<http://docprof.uifc.kiev.ua/index.html>). Это другой подход повышения эффективности функционирования организации, подход гораздо более глубокий, нежели автоматизация. BPR, по определению его идеологов – это фундаментальное переосмысление и радикальная реконструкция бизнес-процессов с целью достижения значительно сильных улучшений в критически важных в современных условиях уровнях критериев производительности, таких как стоимость, качество, услуги, скорость. BPR предназначен для реконструирования работы упрощением или удалением, используя компьютеры не для автоматизации, но для реконструкции существующих бизнес-процессов. Под бизнес-процессом (организационно-производственным процессом) понимают логическую серию взаимосвязанных действий, которая использует ресурсы предприятия для создания или получения в обозримом будущем полезного для заказчика выхода, такого как продукт или услуга. BPR подразумевает реорганизацию предприятия. Но не реорганизацию в ее классическом понимании – изменение организационной структуры, а реорганизацию процессов (бизнес-реинжиниринг). Базовым положением в BPR есть постулат о том, что неэффективность функционирования предприятия заключается в заведомо неправильном построении бизнес-процессов, которые необходимо менять и реорганизовывать в корне, с отказом от старых традиций, со сломом старых правил. В результате предполагается получение коренного перелома в работе, обеспечение

"прорыва", радикального повышения эффективности (например, на 80-100%). А автоматизация, в представлении основателей BPR – это всего лишь простой путь эффективнее выполнять неверные вещи. BPR - суть процесс для организации революционный и как любая революция сложен и противоречив.

Базовой единицей в СЭД есть работа, которая должна быть выполнена с определенными условиями в заданной последовательности и заданными исполнителями. Исполнение работы может производиться в различных временных рамках, контролироваться по времени и содержанию, с ней могут связываться документы, задания, резолюции и т.п. Понятие работы в workflow гораздо шире, чем документа, понятие движения работ шире движения документов, т.е. workflow, в сущности, включает в себя документооборот как частный случай. Информацию о работах (карты работ) workflow-система хранит в БД, документы и другие прикрепления к работам либо хранятся на сервере (в файловой системе, в БД), либо передаются физически от исполнителя к исполнителю согласно карте работы. В случае хранения документов на сервере пользователям передаются только права доступа к ним. Неотъемлемой частью таких систем есть наличие графического редактора маршрутов работ. Возможна жесткая, свободная и смешанная маршрутизация, с параллельным и последовательным выполнением работ. Положительно характеризует такого рода системы наличие открытых интерфейсов и механизмов для интеграции с другими приложениями.

Workflow - системы - это системы, организующие распределение сложно устроенных работ по многим взаимодействующим исполнителям, и координацию их деятельности. Типичными понятиями workflow являются Исполнитель, Работа, Список задач, граф работы и др. Администратор с помощью соответствующих средств (обычно графических) описывает бизнес процессы, что должен делать каждый исполнитель, куда от него должна уйти работа, и др. После этого каждый исполнитель получает на своем рабочем месте список того, что он должен сделать в текущий момент времени, и после того, как он делает то, что от него требуется работа, она перенаправляется дальше, дальше - туда, куда нарисовал администратор (жесткая маршрутизация), или туда, куда ее отослал исполнитель (свободная маршрутизация).

Различные Workflow системы различаются средствами описания графов работ, видами маршрутизации, и собственно, видами работ, которые может выполнять исполнитель в рамках системы. Почти все Workflow системы поддерживают, например, такой вид работы, как заполнение форм (с возможностью, естественно, спроектировать эту форму административными средствами системы и на встроенном языке описать тонкое поведение этой формы при ее заполнении).

Большинство Workflow систем имеют программный интерфейс **API (Application Programming Interface)**, позволяющий внешним задачам, (собственно, тем, которые будут запускать исполнители или сама система)

получать доступ к данным Workflow-системы;

**GroupWare** - системы организации групповой работы. Западный термин - **groupware**. К groupware относят, прежде всего, Microsoft Exchange, Lotus Notes и Novell GroupWise. Системы организации групповой работы являются собой, по сути, расширенные варианты почтовых программ и изначально предназначены для организации обмена информацией в группе - обмена сообщениями, а не документами. Развитие groupware-систем в системы обмена документами обусловлено ростом конкуренции на рынке среди производителей этих систем. Groupware характеризуются ограниченным размером базы данных собственного формата или же имеют возможность ODBC-связи с СУБД, частичной поддержкой SQL, неразвитой системой полнотекстового поиска. Системы, построенные на платформе groupware, перенимают основные недостатки платформы, и автоматизируют, прежде всего, документооборот, а также контрольные и учетные функции, функции хранения и поиска.

Системы groupware и workflow не конкурируют между собой, а скорее дополняют друг друга. Выбор одной из них, а также использование их в комбинации определяются задачами, решаемыми корпорацией. Если вы заботитесь о повышении эффективности работы каждого отдельного сотрудника в коллективе, предпочтение стоит отдать системам класса groupware. На украинно-российском рынке эти системы представлены программным продуктом Lotus Notes (Lotus Development). Если же вы более серьезно обеспокоены повышением эффективности работы всей организации в целом, то вам следует остановить свой выбор на системах класса workflow. На украинно-российском рынке представлены системы Staffware (Staffware plc) и Action Workflow (Action Technologies). В таблице 1 приведена сравнительная характеристика систем класса groupware и workflow.

Таблица 1.

<b>Системы класса GROUPWARE</b>	<b>Системы класса WORKFLOW</b>
Множество работающих	Множество работающих
<i>Системы обоих классов призваны автоматизировать коллективную работу</i>	
Одна выполняемая задача, т.е. поддерживается коллективная работа с одной задачей в данный момент времени (например, с текстовым редактором).	Множество выполняемых задач, т.е. поддерживается многопользовательская работа с несколькими задачами одновременно, как в синхронном, так и в асинхронном режимах.
Отсутствие структуризации в организации работ, т.е. нет никаких правил и предписаний, кто и как должен работать в рамках системы	Строгая структуризация, т.е. выполнение работы четко расписано по ролям, документам, времени обработки документов и т.д.
Реализация на PC и ориентация на небольшие коллективы.	Реализация в среде клиент/сервер и ориентация на масштаб корпорации.

На основе ПО поддержки вполне возможно создать полноценную систему управления документооборотом предприятия, т. к. в некоторых

программных средствах поддержки групповой работы есть развитые среды разработки приложений (например, в Delphi, Lotus Notes). EDMS-системы в большинстве случаев имеют шлюзы в групповое ПО и, при необходимости, могут использовать имеющиеся в нем почтовые программы.

СУБД, системы workflow и системы EDMS - это средства управления информационными потоками. Как же они соотносятся друг с другом? В таблице 2 приведены их краткие характеристики.

Таблица 2

	Назначение	Примечание
<b>СУБД</b>	Ввод, хранение и поиск структурированной информации в	Это только 15% всей информации, с которой
<b>Системы управления документами</b>	Ввод, хранение и поиск неструктурированной информации в электронном виде	Это 85% всей информации, которая проходит через организацию, включая бумажные документы, переведенные в электронную форму
<b>Системы workflow</b>	Управление, маршрутизация и координация передвижением документов в рамках корпоративной системы; контроль за своевременной обработкой документов	Системы могут быть интегрированы как с СУБД, так и системами управления документами

Комбинация технологий СУБД, workflow и управления электронными документами и, соответственно, интеграция программных продуктов, реализующих эти технологии, дает *полное решение* проблемы автоматизации работы с документами любого вида в корпорации любого рода деятельности.

В целом все описанные выше типы систем должны удовлетворять современным требованиям: масштабирование, распределенность, переносимость, поддерживать работу удаленных групп и мобильных пользователей, работу через Интернет, основываясь на трехуровневой модели "клиент - сервер приложений - сервер", иметь модульную архитектуру, быть удобными и простыми в администрировании, иметь интуитивный локализованный интерфейс, локализованные поисковые функции и пр. Системы должны поддерживать такие промышленные стандарты, как OLE Automation, ODMA, ODBC, MAPI, COM/DCOM, CORBA, и т.п., иметь свои собственные открытые API, интегрироваться с офисными приложениями и почтовыми программами. Особое место занимает наличие интеграции с системами ввода документов: сканеры, e-mail-сервера, факс-сервера, OCR-программы, системы голосовой почты.

У специалистов в области автоматизации бухгалтерского учета, финансовой и складской деятельности также сложилась традиция использовать термин "документооборот" еще в одном аспекте. Здесь под документооборотом понимается подсистема программ, обеспечивающая

генерацию “бумажной” отчетности (сводных ведомостей, накладных, справок и пр.) на основе данных, порождаемых в системе.

При рассмотрении вопросов связанных с автоматизацией документооборота обычно возникают смежные темы, такие как *автоматизация перевода документов из бумажного вида в электронный, и связанная с ней задача распознавания текста и содержимого печатных форм, а также задача оптимизации стоимости хранения больших объемов данных;*

в последнее время в прессе появились новые термины, близкие теме автоматизации документооборота - **Document Warehousing (хранилища документов)** и **Knowledge Management (управление знаниями)**. Проблемы, стоящие за этими терминами, находятся в том же отношении к задачам автоматизации документооборота, в каком задачи организации хранилищ данных (**Data Warehousing**) и оперативного анализа данных (**OLAP**) относятся к системам оперативной обработки транзакций (**OLTP**). То есть системы данного класса являются дальнейшим развитием систем документооборота, и необходимы для структуризации и долгосрочного хранения больших массивов накопленных в организации документов, а также оптимального поиска необходимой информации на массиве разнородных накопленных документов.

Прежде чем перейти к собственно структуризации задач и подсистем автоматизации документооборота необходимо отметить некоторые специфические особенности данной отрасли автоматизации предприятия.

Задача автоматизации документооборота лежит на стыке традиционных "**бумажных**" технологий работы с документами и новых компьютерных технологий. Существуют определенные сложности с сертификацией систем электронной подписи, и систем обеспечения секретности доступа к документам.

К тому же, задача документооборота не является изолированной технологической цепочкой в бизнес-процессе организации, документодвижение тесно интегрировано с другими подзадачами, решаемыми информационной системой организации. Таким образом, система автоматизации документооборота *должна обеспечивать прикладные интерфейсы, позволяющие встраивать функции передачи и сохранения документов в прикладные системы, функционирующие в организациях, в которых она внедряется.*

Все перечисленное позволяет сделать вывод - система автоматизации документооборота достаточно сложный механизм. Очевидно, не имеет смысла говорить об отдельном программном продукте, как о системе автоматизации документооборота организации. *Система документооборота предприятия включает в себя множество подсистем, построенных с помощью программных продуктов, как правило, созданных различными производителями.* Система автоматизации документооборота может по-разному интерпретироваться в зависимости от размера организации и специфики ее деятельности. Например, системы автоматизации

документооборота небольшой торговой организации, управленческой структуры, проектной организации и пр. будут выполнять различные функции, строится на различных программных продуктах и вообще иметь мало общего.

**Сложности автоматизации документооборота и процессов управления.** Сегодня существует два альтернативных подхода к автоматизации задач документооборота в организации (<http://www.digdes.ru/docsvision/troubles.phtml>). Первый заключается в последовательной реализации отдельных приложений, автоматизирующих отдельные участки управления документооборота и отдельные процессы управления. Второй - во внедрении платформы для реализации комплексной системы автоматизации документооборота и управления процессами и создание на ее базе приложений, интегрированных в единый информационный комплекс. Оба подхода:

- ❖ внедрение комплекса приложения, реализующие "фиксированную" функциональность,
- ❖ внедрение комплексного решения на базе платформы системы электронного документооборота имеют свои достоинства и недостатки.
- ❖

#### **Приложения, реализующие "фиксированную" функциональность.**

Примерами такого рода приложений могут служить системы автоматизации канцелярии, средства автоматизации контроля исполнения, системы управления проектной документацией и пр. Основным достоинством такого подхода является их относительно невысокая стоимость и достаточно быстрое получение эффекта от внедрения приложений. Приложения данной группы идеально подходят для автоматизации специализированных рабочих мест, ориентированных на выполнение рутинных операций, изолированных от остальных участков автоматизации предприятия, например, таких как рабочее место канцеляриста в крупной организации или система архивной картотеки.

Однако в современной организации все чаще возникает вопрос о необходимости комплексной автоматизации документооборота. Для решения данной задачи приложения должны охватывать большое количество рабочих мест. Причем, один и тот же пользователь может участвовать в разных процессах обработки документов, выступая в различных ролях. Например, может выполнять роль согласующего лица в процессе согласования распорядительных документов, быть ответственным за формирование проектной документации в системе управления проектами, контролировать исполнение заданий в системе контроля исполнения и т. д.

Успешно решая изолированные задачи, конечные приложения не позволяют объединить различные функции в единую систему управления документами и процессами в компании, что существенно уменьшает эффективность от внедрения подобных приложений и ограничивает потенциальные возможности ее использования в дальнейшем. Альтернативой такому подходу является внедрение комплексной системы

автоматизации документооборота на базе единой программной платформы.

### **Комплексные решения на базе платформы системы электронного документооборота**

Внедрение платформы автоматизации документооборота позволяет избежать проблем, возникающих при "кусочной автоматизации" задач документооборота, однако выбор базовой платформы автоматизации является далеко не тривиальной задачей. В настоящее время все системы, представленные на украинно-российском рынке и предлагающие более или менее комплексный подход для решения задач автоматизации документооборота, развиваются на базе одной из трех основных концепций систем автоматизации групповой работы с документами, появившихся еще в конце 80-х - начале 90-х годов. А именно концепций:

а) Groupware приложений и систем поддержки работы с электронными формами (Lotus Notes, Microsoft Exchange);

б) систем управления архивами и маршрутизацией документов (Documentum, PC Docs);

в) систем автоматизации бизнес процессов (Staffware Workflow).

Все эти концепции в той или иной степени могут быть использованы при решении задач, связанных с автоматизацией документооборота, однако реализация их осуществляется различными способами.

Так, например, Groupware-приложения акцентированы на упрощение работы с электронными формами и процессов их создания, а также обеспечение мощных механизмов интерактивной работы с их групповыми представлениями (форм дизайнер Outlook, механизм представлений Lotus).

Системы архивов документов сконцентрированы на всем, что имеет отношение к работе с файлами документов: ввод, фиксация атрибутов, включая организацию справочной информации (поиск, блокировку) и осуществление off-line работы, аудит и разграничение прав доступа, маршрутизация.

WorkFlow-приложения предназначены для моделирования и, собственно, управления бизнес-процессами, а также для организации средств доступа к информации (универсальная очередь заданий), управления и контроля активности пользователей (напоминания, управление приоритетами заданий, средств контроля за ходом процесса и пр.).

Каждый из перечисленных традиционных подходов к построению комплексной системы автоматизации документооборота сконцентрирован на одном из аспектов автоматизации и предоставляет в распоряжение потребителей мощные средства для их реализации. При этом, они либо вовсе не реализуют дополнительные функции, либо реализуют их, но в степени, недостаточной для построения качественных законченных решений.

Так, например, Groupware-приложения не вполне удобны для создания архивов документов - в них недостаточно развиты средства управления жизненным циклом обработки документов, встроенные механизмы маршрутизации ограничиваются пересылкой отдельных сообщений. Кроме

того, в них сложно реализовывать развитые справочники системной информации. Архивы документов совершенно не предназначены для работы с разнородными и сложными структурами учетных картотек документов, средства организации групповой работы в них, как правило, ограничиваются внешними системами пересылки документов и ссылок. Системы Workflow, как правило, поддерживают ограниченные возможности работы с документами, имеют существенные ограничения при работе со сложными данными, не поддерживают функции свободной маршрутизации и не содержат средств для организации архивного хранения информации.

В последнее время помимо выполнения вышеперечисленных функций к системам автоматизации документооборота предъявляются дополнительные требования, связанные с **управлением и извлечением знаний**, накопленных в корпоративной информационной системе - знаний о процессах в системе Workflow, знаний, хранящихся в документах в корпоративных архивах документов, знаний о групповых взаимодействиях и взаимодействиях с внешним окружением компании. Эти требования вносят дополнительные осложнения в выбор того или иного решения в данной области автоматизации.

Из вышесказанного следует, что существуют серьезные затруднения в принятии решения о внедрении той или иной платформы автоматизации документооборота и процессов управления. Перед специалистами отделов автоматизации возникает необходимость решения всего многообразия задач, связанных с обработкой документов и поддержкой бизнес-процессов. Но, выбрав одно из решений, они получают удовлетворительное решение лишь задач определенного типа. Например, получают возможность создавать мощный архив документов с помощью **DMS** системы (Documents Management System — система управления документами. Иногда такие системы называют EDMS (Electronic Documents Management System) — системы управления электронными документами (СЭД)). Остальные же функции, нереализуемые с помощью базовых средств внедренной системы, например, средства группового обслуживания документов приходится решать одним из двух способов:

1) пытаться надстраивать базовые функции системы дополнительными модулями и разрабатывать соответствующие расширения системы. В определенных случаях это бывает крайне затруднительно, а то и невозможно;

2) внедрять систему другого класса, восполняющую недостающие функции, и обеспечивать интеграцию двух систем. При этом происходит частичное дублирование функций различных систем, приводящее к необходимости поддержки нескольких серверов и ряда клиентских приложений, организующих собственное рабочее пространство и средства доставки информации, собственные средства организации справочников и пр.

Из-за этого стоимость внедрения системы и использования приложений существенно возрастает. К тому же организация оптимального взаимодействия приложений, построенных на базе различных платформ,

достаточно сложна и не всегда возможна.

### **Жизненный цикл документа**

Спектр задач и соответственно необходимая система автоматизации определяются стадией жизненного цикла документа, которую необходимо поддерживать. Жизненный цикл документа (рис. 1) состоит из двух основных стадий.

1. Стадия разработки документа:

- а) собственно разработка содержания документа;
- б) оформление документа;
- в) утверждение документа.

В том случае, если документ находится на стадии разработки, то он считается неопубликованным и права на документ определяются правами доступа конкретного пользователя.

2. Стадия опубликованного документа:

- а) активный доступ;
- б) архивный документ;
- в) краткосрочного хранения;
- г) долгосрочного хранения;
- д) уничтожение документа.

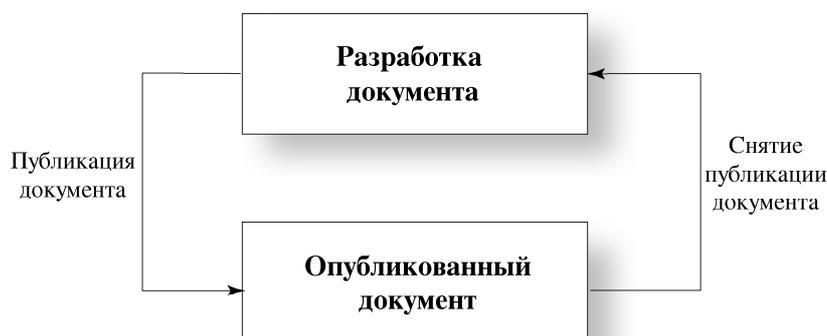


Рис. 1. Жизненный цикл документа

Когда документ переходит на вторую стадию, он считается опубликованным, тогда права на документ остаются только одни - доступ на чтение. В качестве примера опубликованного документа можно привести шаблон стандартного бланка предприятия. Кроме права доступа на чтение могут существовать права на перевод опубликованного документа в стадию разработки.

Любая хорошо спроектированная СЭД позволяет:

- ❖ централизованно отслеживать ход делопроизводственного процесса - вплоть до работы исполнителей над документами на своих рабочих местах, систематизировать результаты контроля и тем самым выявлять «узкие места» в документообороте и работе персонала;
- ❖ при переходе на работу с электронными копиями документов

полностью исключить возможность их утраты, и свести поиск документа к запросу, выполняемому системой в течение считанных секунд. При этом необязательно знать название или регистрационный номер документа - достаточно указать тематику или ключевые слова для поиска. Компания Coopers&Lybrand указывает, что организации делают в среднем 19 копий каждого документа и что почти 7,5% всех документов теряется безвозвратно (<http://eos.ru/eos/21308>);

❖ повысить рост производительности труда сотрудников за счет упорядочения технологии работы с документами, резкого сокращения времени передачи документов по локальной сети или средствами электронной почты. В порядке подтверждения - еще несколько цифр: по оценке Nortan Nolan Institute при переходе к работе с электронными документами возрастает производительность труда сотрудников на 25-50%, сокращается время обработки одного документа более чем на 75%, а уменьшение расходов на оплату площадей для хранения документов составляет до 80% (<http://eos.ru/eos/21308>).

**Подсистемы и их функции.** В данном разделе мы сможем пока лишь коротко остановиться на функциях подсистем, решающих отдельные задачи автоматизации документооборота. Более детальному обсуждению каждой из задач и особенностей программного обеспечения для их реализации будут посвящены дальнейшие рубрики.

Как мы уже убедились, система автоматизации документооборота складывается из нескольких подсистем. Каждая подсистема обладает набором специфических для нее функций. При этом отдельные подсистемы тесно взаимодействуют между собой. Разделение системы документооборота на подсистемы, предпринимаемое нами носит несколько “академический” характер. В реальной практике программные продукты достаточно условно можно отнести к той или иной группе в нашей классификации. Как правило, системы реализуют лишь часть функций, описанных ниже, при этом продукт одного класса может включать в себя часть функций систем другого класса. Поэтому построение систем автоматизации документооборота из существующих на рынке продуктов требует не только хорошего понимания конечной задачи, но и отличного знания рынка программного обеспечения.

Можно выделить следующие подсистемы автоматизации документооборота:

- ❖ системы автоматизации делопроизводства;
- ❖ архивы документов;
- ❖ системы ввода документов и системы обработки образов документов;
- ❖ системы управления стоимостью хранения документов;
- ❖ системы маршрутизации документов;
- ❖ системы комплексной автоматизации бизнес-процессов;
- ❖ системы, основанные на применении Internet/intranet - технологий.

Кратко остановимся на функциях каждой из них.

**Системы автоматизации делопроизводства. Функции**

*автоматизации делопроизводства в том или ином виде представлены в любой системе автоматизации документооборота.* В функции систем автоматизации делопроизводства не входит хранение и перемещение документов в организации. В их функции входит фиксация документов в специальной базе данных, выражающаяся в заполнении специальной карточки документа. Содержимое карточки документа может варьироваться в зависимости от сложившейся в организации ситуации. Структура документов, зафиксированных в базе данных, опирается на так называемую номенклатуру дел, имеющуюся, как правило, в каждой организации, а технология учета и обработки документов опирается на сформулированное в данной организации “Положение о делопроизводстве”. Как мы уже отметили, документы хранятся в бумажном виде, в специальном архиве, но в базе данных отображается их текущее местоположение и статус, включая атрибуты контроля исполнения. Обычно в системах делопроизводства различают входящие и исходящие документы, нормативно-распорядительные документы, документы коллегиальных органов управления, справочные документы и пр. Документы, находящиеся на контроле исполнения, подразделяются по исполнителям, статусу исполнения, срокам исполнения и прочее. Каждый документ в системе представляет собой запись в базе данных, характеризующуюся набором значений атрибутов карточки. Помимо учета и поиска документов в базе данных, система должна обеспечивать генерацию отчетов, позволяющих получить ведомости исполнения документов и прочую сводную информацию.

Для разработки приложений, выполняющих функции автоматизации делопроизводства больше всего подходят стандартные инструменты, используемые для разработки автоматизированных рабочих мест, от настольных баз данных до систем на базе различных **SQL** серверов.

Однако в том случае, если автоматизация документооборота не закончиться данным шагом, то можно подумать и о других инструментах, обеспечивающих более последовательное развитие системы. Так например, при переходе к электронному хранилищу документов, база данных системы делопроизводства должна содержать ссылки на соответствующие объекты электронного архива, при использовании электронных средств маршрутизации документов система должна обеспечивать возможность рассылки документов на рабочие места пользователей, определения текущего местоположения документа и так далее. О конкретных решениях и программных продуктах для реализации данных функций мы поговорим в последующих рубриках.

**Архивы документов.** *Архив документов это то, что собственно хранит электронный документ.* При этом может храниться либо образ документа, либо его содержание, либо и то и другое. В начале 90-х появились два новых подхода к интеграции существующих данных и приложений: технология хранилищ данных и системы планирования ресурсов предприятий (ERP — enterprise resource planning). Каждый из них решал проблему интеграции, но лишь частично. Технология хранилищ данных была

ориентирована на интеграцию данных, их анализ, не обеспечивая при этом интеграции приложений при проведении транзакций. Очевидное преимущество хранилищ данных в том, что они могут быть построены «вокруг» унаследованных приложений. Иными словами, эта технология не требует переписывания старых приложений, а позволяет пользоваться данными, полученными уже существующими программами. Еще одно преимущество хранилищ данных — поэтапность их построения, повышающая отдачу от их внедрения. Следующим шагом после хранилищ данных стало создание механизмов обработки и анализа корпоративной информации. При этом работа может вестись как со структурированными данными (например, производственные и финансовые показатели деятельности компании), так и с неструктурированными (ленты информационных агентств, организационные регламенты и т.д.).

Помимо собственно хранения документов, технология хранилищ данных должна обеспечивать навигацию по иерархии документов и их поиск. В отличие от поиска по атрибутам документов, который имелся и в системах предыдущего класса, архивы документов должны обеспечивать полнотекстовый поиск по содержимому текстовых фрагментов в документе. В предельном случае поисковый механизм должен обладать некоторым интеллектом, то есть обеспечивать поиск близких грамматических конструкций, а также поиск близких по смыслу слов.

В отличие от систем предыдущего класса, в архивах хранятся сами документы, и по этому система должна обеспечивать разграничение прав доступа к документам. Пользователь может идентифицироваться либо посредством сетевого имени, либо с помощью специального имени и пароля определенного в системе управления архивом. Помимо разделения прав доступа на уровне пользователей система должна обеспечивать выделение групп пользователей или ролей.

Следующей функцией архива документов является обеспечение возможности групповой работы с документами, находящимися в стадии создания - это функция блокировок документов или **Check-In/Check-Out контроль**. Если один из пользователей системы начинает редактировать документ, он блокируется для доступа других пользователей до тех пор, пока с ним не закончится работа.

Еще одной функцией архива является поддержка контроля версий. Версии документов могут фиксироваться либо автоматически, либо по инициативе пользователя. В случае необходимости пользователь может вернуться к одной из предыдущих версий документа.

К сервисным функциям архива документов относятся возможность создания резервных копий документов без прекращения работы системы, интеграция с системами обеспечения оптимальной стоимости хранения данных и прочее. В зависимости от конкретной стадии жизненного цикла документа, с которым имеет дело архивная система, архивы подразделяются на следующие типы.

**Статические архивы документов (либо просто архивы) - системы,**

которые имеют дело только с опубликованными документами.

**Динамические архивы документов (либо системы управления документами)** - системы, имеющие дело как с опубликованными документами, так и с теми, которые находятся в разработке.

### **Системы маршрутизации документов.**

Системы маршрутизации документов занимаются непосредственно пересылкой документов на рабочие места исполнителей, сбор информации о текущем статусе документов, осуществляют консолидацию документов по завершению работы с ними на отдельных этапах, а также обеспечивают средства доступа к информации о текущем состоянии работ с документами.

При построении систем маршрутизации могут применяться два основных подхода.

*Первый носит название документно-ориентированный (режим **Off-Line**)*. Документ является основным объектом системы, и маршрутизируется именно он, а все остальные параметры маршрутизации ассоциированы именно с документом. *Второй подход носит название работо-ориентированный (режим **On-Line**)* и его основным объектом является работа. К работе может быть прикреплен самый разнообразный список объектов, в том числе, и документы. Естественно, работа может существовать и без документов. Второй подход является более общим.

Маршрут движения документа представляет собой ориентированный граф, вершинами которого являются работы, закрепленные за субъектами документооборота. В системе должны быть реализованы три типа работ: ручные, полуавтоматические и автоматические. Примером ручных работ являются подготовка документа с помощью текстового редактора, подписание документа, наложение резолюции и т.д. Под полуавтоматическими работами понимаются такие, в которых часть обработки делается с помощью жесткого алгоритма системы, а вторая часть работы выполняется непосредственно человеком. Типичным примером полуавтоматической работы является процедура сканирования и распознавания текста - сканирование и распознавание выполняются автоматически соответствующими программными системами, а окончательную проверку и корректировку текста выполняет человек. Автоматическими работами являются процедуры, выполняемые в узлах маршрутов движения документов без участия человека, в соответствии с некоторым алгоритмом, программная реализация которого хранится в полях базы данных. Примером автоматической работы являются процедуры импорта/экспорта почтовых сообщений из системы корпоративной электронной почты. Использование таких автоматических работ позволяет интегрировать работу электронной почты и системы документооборота. Система должна быть открытой с точки зрения добавления новых работ различных типов.

Система управления корпоративным документооборотом должна

поддерживать как «жесткую» маршрутизацию движения документов, так и «свободную» маршрутизацию, когда решение о дальнейшем направлении движения документа принимает пользователь документооборота, при наличии у него права на свободную маршрутизацию. В системе должна быть предусмотрена возможность создания шаблонов маршрутов и документов для облегчения работы пользователей.

Важным свойством маршрутизации документов является поддержка вложенных маршрутов движения документов. Поддержка вложенных маршрутов позволит строить систему иерархически. На первом этапе можно построить маршруты движения документов на уровне самых крупных структурных подразделений организации. В дальнейшем, строя подмаршруты, можно детализировать документооборот до конечного пользователя.

Рассмотрим теперь типы систем маршрутизации (рис. 2.)

**1) Свободная маршрутизация.** Выделяется два основных типа маршрутов документов. *Последовательная маршрутизация* - документ последовательно проходит одного исполнителя за другим. Передача документа от одного пользователя к другому может происходить по истечении контрольного времени, либо после завершения работы одним из них. *Параллельная маршрутизация* - документ одновременно поступает всем исполнителям, а завершение маршрута происходит, когда один либо все пользователи завершат работу с документом.

**2) Системы электронной почты.** Минимальной достаточной системой, обеспечивающей маршрутизацию документов является система электронной почты, которая осуществляет параллельное распространение документов (маршрутизация отличается от распространения или рассылки тем, что маршрутизируемый документ возвращается в начало маршрута, например к инициатору, а рассылаемый документ уходит к исполнителю без контроля факта возврата). С помощью дополнительных приложений система электронной почты может обеспечивать последовательную маршрутизацию документов.

**3) Свободная последовательная и параллельная маршрутизация документов с контролем исполнения.** В случае использования последовательной маршрутизации документ последовательно проходит одного исполнителя за другим. Передача документа от одного пользователя к другому может происходить по истечении контрольного времени, либо после завершения работы одним из них. В случае использования параллельной маршрутизации документ одновременно поступает всем исполнителям, а завершение маршрута происходит, когда один либо все пользователи завершат работу с документом.

Под контролем исполнения понимается следующая функциональность:

❖ *контроль доставки задания* - инициатору выдается информация о том, что его задание достигло места назначения (исполнителя);

❖ *контроль прочтения задания* - инициатору выдается информация о том, что с его заданием ознакомились сотрудники для которых это задание

было предназначено;

❖ *контроль выполнения* - инициатору выдается информация о том, что задание выполнено;

❖ *мониторинг задания* - инициатор всегда может посмотреть, кто и что сейчас делает с его заданием.

❖ *извещение о нарушении сроков исполнения* - система документооборота может известить инициатора о том, что посланное им задание просрочено конкретным сотрудником.

❖ *история выполнения заданий*;

❖ *контроль качества исполнения* - означает, что, если пользователь говорит о том, что задание исполнено, это еще не означает, что оно действительно исполнено, инициатор должен проверить качество исполнения, подтвердить или нет исполнение.

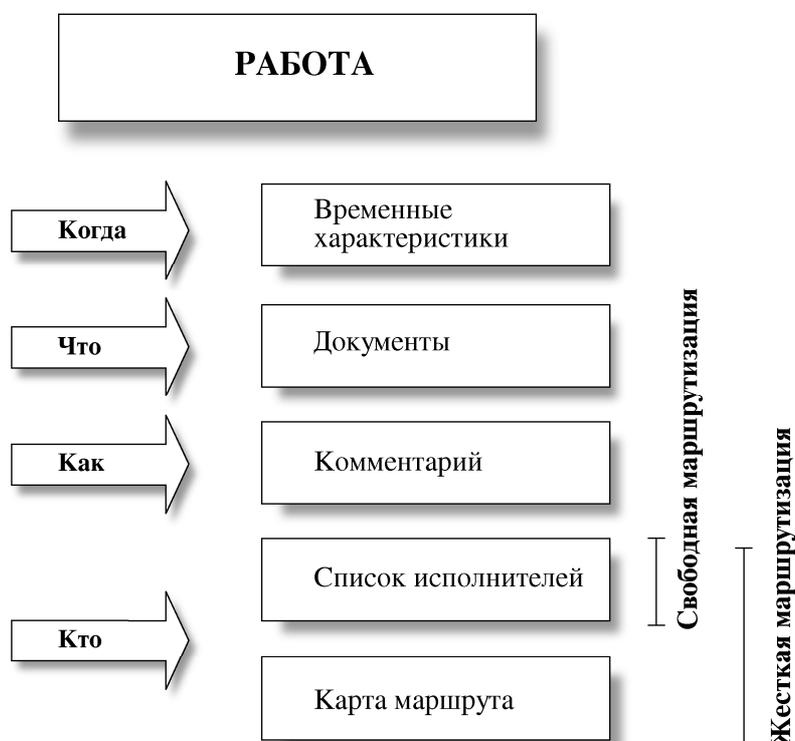


Рис. 2. Объекты системы маршрутизации

Информация может выдаваться в виде изменения статуса задания в окнах входящих и исходящих заданий или в виде нового задания сформированного системой инициатору либо с помощью сообщения по электронной почте.

**4) Свободная последовательно-параллельная маршрутизация документов.** Последовательно-параллельная маршрутизация совмещает в себе особенности последовательной и параллельной маршрутизации документов.

В качестве механизма обеспечивающего параллельную

маршрутизацию документов может быть использована *система корпоративной электронной почты*, которая осуществляет параллельное распространение документов (маршрутизация отличается от распространения или рассылки тем, что маршрутизируемый документ возвращается в начало маршрута, например к инициатору, а рассылаемый документ уходит к исполнителю без контроля факта возврата). С помощью дополнительных приложений система корпоративной электронной почты может обеспечивать последовательную маршрутизацию документов.

#### **5) Маршрутизация документов по заранее определенным маршрутам с контролем исполнения (жесткая маршрутизация).**

Маршруты могут быть более сложными, чем простые последовательные или параллельные:

- ❖ комбинированные из последовательных и параллельных элементов;
- ❖ условные, с переходами в зависимости от состояния тех или иных переменных маршрутов.

Такие маршруты становятся сложными для их задания "на лету", поэтому в этом случае используется специализированный графический редактор, позволяющий создать маршрут. Инициатор вызывает созданный и именованный маршрут и прикрепляет к нему документы - иницирует его. Система маршрутизации должна быть интегрирована с архивной системой, и реальные приложения для работы с документами не могут быть основаны только на файловой системе. И вот почему. Любой процесс маршрутизации документов - это движение одного документа, а не множества его копий, как это происходит в системах электронной почты. Посылать один документ необходимо не только по соображениям экономии пространства, но и в основном для поддержания его целостности - в процессе маршрутизации многие пользователи пытаются вносить изменения в документ. Кроме этого, было бы желательно, чтобы система маршрутизации была интегрирована с архивной системой по следующим параметрам.

*По списку пользователей и системе безопасности.* Это означает, что если вы собираетесь послать кому-то документ, то адресат должен обладать соответствующим набором прав для работы с этим документом. Если прав недостаточно, то система должна попросить инициатора работы или маршрута установить соответствующие права.

*Интеграция с операцией опубликования документа.* Задача состоит в том, что после окончания маршрута документ, ассоциированный с маршрутом, меняет свой статус на опубликованный. В качестве примеров таких маршрутов можно привести процесс утверждения документа.

Обе схемы маршрутизации имеют и другие свои достоинства и недостатки. Подробные их обсуждения мы вынесем в отдельные разделы.

#### **5. Системы комплексной автоматизации бизнес процессов.**

Развитием систем маршрутизации документов являются Workflow системы, или системы комплексной автоматизации бизнес процессов. В отличие от систем маршрутизации документов, объектом маршрутизации в них является

совокупность данных используемых в некотором бизнес процессе. Пользователь получает на рабочее место информацию о том, что он должен сделать и все необходимые для этого данные. WorkFlow приложение определяет, какое приложение должно быть запущено для реализации функций на данном рабочем месте, и загружает в него необходимые данные. Парадигма WorkFlow системы предполагает, что пользователь должен выполнять только необходимые функции, всю рутинную работу – определение последовательности действий, доставку необходимой информации, контроль своевременности исполнения работы и прочее выполняет система WorkFlow.

Функции WorkFlow приложений выходит за рамки функций систем документооборота, однако, технологии, используемые в данных приложениях очень близки технологиям, используемым в системах маршрутизации документов, к тому же маршрутизация документов может рассматриваться как частный случай задачи построения WorkFlow систем, поэтому мы уделим им некоторое внимание в последующих разделах.

**б. Системы управления выводом документов.** Одним из особых сегментов современного рынка СЭД являются так называемые системы управления выводом документов (output management systems — OMS), основным предназначением которых является генерация выходных документов. Ряд OMS-систем отвечает только за распределение и доставку выходных документов (в электронном виде — в форматах HTML, XML и PDF и т. д.). Очень часто OMS-системы интегрированы с программными пакетами сканирования документов и изображений. Полезной возможностью некоторых OMS-систем является и взаимодействие с унаследованными системами.

В некоторых OMS-системах дополнительно реализована возможность архивации и долговременного хранения выходных отчетов и документов. В связи с этим, многие из OMS-систем классифицируются Gartner Group как **интегрированные системы архивации и поиска документов (IDARS — integrated document archive and retrieval systems)**. Однако главной причиной популярности OMS-систем все же является занимаемая ими рыночная ниша — генерация документов и отчетов в информационных системах предприятий и организаций, построенных с использованием ERP-систем. По мнению аналитиков Gartner Group, **одним из слабых мест современных ERP-систем является плохое управление генерацией выходных документов** (разработчики ERP-систем больше сосредоточены на повышении функциональности ключевых модулей своего ПО, чем на «второстепенных» вопросах обеспечения генерации выходных отчетов, не имеющих, по их мнению, хороших рыночных возможностей). Именно этот недостаток ERP-систем и послужил основным фактором появления и быстрого развития рынка OMS-систем.

Для поддержки принятия управленческих решений, система должна обеспечивать формирование различных отчетов, включая:

- ❖ справки, сводки и списки по документам, тематическим рубрикам;

- ❖ отчеты произвольного состава и структуры по сформированной подборке документов;
- ❖ справки по типам документов, по исполнителям, по контролерам;
- ❖ комплексная оценка состояния дел в отношении исполнительской дисциплины, выделения узловых проблем в тематике путем формирования аналитических сводок по ходу исполнения документов за заданный период и др.

**7. Системы, основанные на применении Internet/intranet - технологий.** Неотъемлемой чертой современных стало применение технологий Internet/intranet. Учитывая этот факт, при выборе составляющих СЭД необходимо отдавать предпочтение программам, которые поддерживают полноценную работу из обычного браузера, фактически, имеют так называемый тонкий клиент и специальное серверное программное обеспечение, обеспечивающее функционирование данного клиента. Как правило, такое техническое решение позволяет использовать стандартные хранилища данных (библиотеки документов, базы данных) из локальных, корпоративных и глобальных сетей, не требуя существенных затрат на дополнительное администрирование и поддержание целостности, надежности и безопасности хранения данных.

Рассматривая вопрос применения Internet/intranet - технологий, нельзя не затронуть такую важную проблему, как обеспечение информационной безопасности. Для предотвращения несанкционированного доступа к документам и для исключения возможных диверсий злоумышленников встроенных средств СЭД недостаточно. Поэтому в состав СЭД обязательно должны войти специальные программно-аппаратные средства защиты.

Они, в частности, позволяют шифровать данные, поддерживают электронную цифровую подпись и могут проводить на ее основе аутентификацию пользователей. Все это обеспечивает достоверность и целостность информации внутри СЭД. Эффективность программных средств защиты может быть существенно повышена за счет применения аппаратных и биометрических средств: аппаратных ключей, смарт-карт, устройств распознавания отпечатков пальцев, сетчатки глаза, голоса, лица, оцифрованной подписи. В дополнение к ним на стыке сегментов локальных сетей и Internet желательна установка брандмауэров — средств контроля за внешними (входящими и исходящими) соединениями. Они позволяют отслеживать передачу информации практически всех известных на сегодняшний день протоколов Internet.

В основе системы документооборота лежит подсистема хранения. В простейшем случае это может быть база данных учета документов в системах автоматизации делопроизводства. В том случае, если система документооборота хранит документы или их образы необходима специальная среда их хранения. В простейшем случае для этого может использоваться файловая система. В более сложном случае это может быть специализированная система построения архива документов. База данных

должна содержать информацию о правах доступа к документам, блокировках и пр. В этом случае возможно присутствие систем оптимизации стоимости хранения документов и системы поддержки полнотекстового индексирования.

Клиентские приложения, соответственно, должны включать такие дополнительные функции, как доступ к документам в архиве, поддержка блокировки и изменения атрибутов документов, а также возможность полнотекстового поиска.

Для ввода документов в систему может использоваться система распознавания текстов и печатных форм.

В случае наличия в системе средств маршрутизации, база данных хранит дополнительную информацию о маршрутах движения документов, текущем состоянии маршрутов и пр. Модуль маршрутизации осуществляет непосредственную доставку документов на рабочие места пользователей и внесение информации об изменении статуса документов в базу данных. Клиентское рабочее место системы маршрутизации может непосредственно не взаимодействовать с базой данных и архивом документов, а получать доступ к документам посредством среды маршрутизации, в качестве которой может использоваться электронная почта.

**Общие требования к системе электронного документооборота.** Они заключаются в следующем:

**Масштабируемость.** Желательно, чтобы система документооборота могла поддерживать как пять, так и пять тысяч пользователей, и способность системы наращивать свою мощность определялось только мощностью соответствующего аппаратного обеспечения. Выполнение такого требования может быть обеспечено с помощью поддержки промышленных серверов баз данных производства таких компаний, как Sybase, Oracle, Informix и др., которые существуют практически на всех возможных программно-аппаратных платформах, тем самым обеспечивая самый широкий спектр производительности.

Система на базе сервисов абсолютно масштабируема. Минимальный вариант: все сервисы и клиенты располагаются на одном компьютере, разделяя процессорное время. Максимальный вариант: каждый сервис имеет в своем монопольном владении не только свой ресурс, но и целый процессор. Промежуточные варианты: часть системы находится на одном компьютере, а часть на другом. В результате система может быть масштабирована в зависимости от количества информации, которую нужно обрабатывать и необходимой скорости обработки. Можно также использовать разнокалиберные ПК в одной системе, распределяя нагрузку соразмерно мощности. Последнее свойство особо ценно, если уже имеется некоторый парк унаследованных машин.

Система позволяет размещать хранилище электронных документов на нескольких территориально разнесенных серверах. Предусмотрена поддержка различных носителей данных - жестких дисков, магнитных лент, библиотек магнитооптических дисков, CD- и DVD-библиотек. Имеются

штатные средства кэширования, сетевой оптимизации, сжатия данных, автоматического перемещения редко используемых документов на более медленные носители. Доступ к документам, хранящимся в системе, возможен из любого места, где есть Internet-соединение. Система имеет открытую архитектуру и предоставляет разработчикам полный комплект библиотек для создания собственных интерфейсов прикладного программирования (API) для клиент-серверной и Web-архитектуры.

**Распределенность.** Основные проблемы при работе с документами возникают в территориально-распределенных организациях, поэтому архитектура систем документооборота должна поддерживать взаимодействие распределенных площадок. Причем распределенные площадки могут объединяться самыми разнообразными по скорости и качеству каналами связи. Архитектура системы должна поддерживать также взаимодействие с удаленными пользователями. Распределенное, расширяемое управление документами приводит к резкому повышению продуктивности работы сотрудников, усилению общей конкурентоспособности организации, обеспечивая оптимизацию любого количества междисциплинарных процессов, вместо автоматизации отдельных вертикальных. Корпоративное управление документами является существенным шагом на пути к воплощению в жизнь инициатив по управлению корпоративными знаниями. Также должна быть обеспечена автоматическая поддержка распределенного управления различными информационными материалами на протяжении всего их жизненного цикла, от создания до рецензирования, утверждения, распространения и архивирования.

*Распределенность* предполагает способность архитектуры системы поддерживать удаленных (мобильных) пользователей и территориально-распределенные площадки предприятия (удаленные филиалы, управления, дочерние компании) в независимости от скорости и качества каналов связи. Система не должна накладывать жесткие требования на каналы передачи данных. Оптимальным способом выполнения указанного требования является построение системы на основе трехуровневой архитектуры с тонким клиентом по технологии Intranet/Internet, что позволит сократить затраты на модернизацию (или полное перепроектирование) корпоративной сети передачи данных на предприятии.

**Безопасность и защищенность.** Все действия, связанные как с целыми документами, так и их частями, могут протоколироваться. Используется понятие жизненного цикла хранимых документов и его мониторинга. Система безопасности имеет необходимые средства аутентификации и шифрования. Функции электронной подписи бывают встроенными или подключаемыми как внешний модуль от сторонних производителей. Есть полноценные средства администрирования и обслуживания рабочих групп.

*Защищенность* предполагает использование надежных технологий авторизации и защиты информации от несанкционированного доступа. Защита информации должна быть обеспечена на нескольких уровнях (многоуровневая защита):

❖ уровень защиты корпоративной сети (например, построение виртуальных частных сетей – VPN);

❖ уровень защиты операционной системы. Каждый пользователь системы является пользователем операционной системы, и для него определяются доступные и недоступные виды деятельности;

❖ уровень защиты СУБД (уровень протокола используемой СУБД). Любая деятельность пользователя с базой данных должна быть зарегистрирована средствами используемой СУБД;

❖ уровень защиты информационных объектов с помощью электронной подписи;

❖ уровень защиты информационных объектов с помощью криптографических алгоритмов

❖ уровень защиты при авторизации пользователя. Каждый субъект документооборота имеет определенную структуру доступа к различным приложениям и данным. Авторизация доступа может проводиться как с помощью обычной парольной защиты, так и с использованием дополнительных технических средств. При авторизации настраиваются доступные пользователю части меню, протоколируются попытки запрещенных действий. Пользователи могут быть объединены в группы, каждой из которых доступны определенные виды деятельности;

❖ уровень протокола пользовательского интерфейса. Любая деятельность пользователя регистрируется средствами приложений системы.

СЭД должна обеспечивать разграничение доступа работников отдельных подразделений как непосредственно к текстам документов или их карточкам, так и ко всем этапам документооборота, включая наложение резолюций руководителями и подготовку отчетов исполнителями. Помимо аутентификации пользователей требуется поддержка нескольких уровней доступа (разрешающих, например, просмотр, редактирование, создание, удаление, печать, и прочих), определяемых в отношении каждого документа (и даже полей в регистрационной карточке), а также ролевых групп с различными полномочиями (допустим, группы начальников отделов). В случае, если СЭД является территориально распределенной системой, желательно, чтобы для обмена документами по открытым каналам связи она имела встроенные и сертифицированные ФАПСИ средства цифровой подписи и шифрования.

**Модульность.** Вполне возможно, что заказчику может не потребоваться сразу внедрение всех компонентов системы документооборота, а иногда спектр решаемых заказчиком задач меньше, чем весь спектр задач документооборота. Тогда очевидно, что система документооборота должна состоять из отдельных модулей, интегрированных между собой, причем обеспечивать гибкость управления доступом ко всему спектру документов, от электронной почты до дискуссионных баз данных, от видео клипов до формализованных документов всех типов. Должна быть обеспечена возможность обеспечения мгновенного доступа к документам через Web-браузеры, настольные приложения и другие общедоступные типы

клиентов.

**Администрирование и обеспечение надежности. Выборочное резервирование.** Средства администрирования СЭД, позволяющие назначать и модифицировать ролевые полномочия пользователей по отношению к документам и функциям системы, должны уметь настраивать ее в соответствии с постоянно меняющейся организационной структурой. Желательно, чтобы действие таких настроек распространялось не только на отдельные документы, но и на их группы, имеющие иногда сложную иерархическую структуру. Для аналитической работы с документами не обойтись без инструментов, позволяющих относить документы к различным темам и классифицировать по тем или иным признакам. Важнейшими характеристиками устойчивости СЭД к неизбежным сбоям и отказам являются: независимость функционирования рабочих мест друг от друга, время восстановления сеанса после его аварийного завершения, наличие средств обеспечения целостности данных и их резервного копирования.

Выборочное резервирование является одним из самых ресурсоемких, но обязательных требований для систем повышенной ответственности. Однако на практике не вся информация требует резервирования, поэтому резервирование должно быть выборочным, настраиваемым. Надежность имеет непосредственное отношение к функциональной модели. Например, выход из строя сервиса архивирования данных (ABS) по какой-либо причине (ошибки программирования, разрушение ресурса и т. д.) не приведет к выходу из строя интерфейсов пользователей-операторов и сервисов связи с контроллерами. В результате оперативное отображение текущего состояния технологического объекта и управление будут доступны пользователю и начнут нормально функционировать.

Горячее резервирование реализуется на основе способности клиента соединяться одновременно с несколькими сервисами любого типа. Для горячего резервирования используются сервисы одного типа. Например, у нас имеются два компьютера, на каждом из которых установлен сервис архивирования данных. Клиент отправляет данные истории на оба сервиса. При выходе из строя одного из них клиент распознает это и запрашивает информацию (графики истории, протоколы) из другого, работающего сервиса. При восстановлении вышедшего из строя сервиса клиент распознает его, конфигурирует и продолжает работу с ним. Параметры, подлежащие резервированию, определяются в конфигурации.

**Протоколирование работы пользователей.** Протоколирование позволяет отследить всю историю документа (кто и когда его создал, редактировал, просматривал, печатал и т. д.) и дает возможность разрешать конфликтные ситуации при совместной работе над ним нескольких пользователей.

**Открытость.** *Открытость* предполагает возможность тесной (бесшовной) интеграции системы с другими системами, эксплуатируемыми на предприятии (к таким системам можно отнести: корпоративный web-сайт, систему корпоративной электронной почты, офисные приложения, системы

управления предприятия, в том числе системы класса ERP/MRP, системы управления проектами и т.д.). Система документооборота не может и не должна существовать в отрыве от других систем, если, например, необходимо интегрировать систему с прикладной бухгалтерской программой. Тогда система документооборота должна иметь открытые интерфейсы для возможной доработки и интеграции с другими системами. Все EDMS-системы построены по модульному принципу, а их API-интерфейсы являются открытыми. Это позволяет добавить к EDMS-системам (при необходимости) новые функции или расширять уже имеющиеся. В настоящее время разработка приложений, хорошо интегрируемых с EDMS-системами, стала отдельным видом бизнеса в отрасли промышленного производства ПО, и множество третьих фирм готовы предложить здесь свои услуги. Возможность относительно простого добавления к EDMS-системам множества модулей разработки третьих фирм значительно расширяет их функциональные возможности. Например, для EDMS-систем разработаны модули ввода документов со сканера, связи с электронной почтой, с программами пересылки факсов и др. Открытая, расширяемая архитектура, позволяет организациям, во-первых, быстро расширять платформу управления документами в ответ на появление новых бизнес целей, таких, как управление записями и, во-вторых, интегрировать управление документами с более широкими стратегическими инициативами, такими как Управление знаниями.

Интеграция системы необходима для создания единого информационного пространства на предприятии с возможностью сквозного управления деятельностью предприятия и обеспечения эффективного контроля исполнения управленческих решений. Система должна иметь открытые интерфейсы, включая открытую структуру баз данных, исходных текстов, описание структуры и логики работы системы для ее возможной доработки (развития) и интеграции с другими корпоративными приложениями.

**Совместимость.** При создании системы используются информационно-технологические интерфейсы, благодаря которым она может взаимодействовать с другими системами. Важной особенностью EDMS-систем является также высокая степень их интеграции с прикладными программами за счет использования технологий OLE Automation, DDE, ActiveX, ODMA, MAPI и др. А непосредственно при работе с документами вообще нет необходимости пользоваться утилитами EDMS-системы. Пользователи имеют дело только с обычными прикладными программами: в момент инсталляции клиентской части EDMS-системы прикладные программы дополняются новыми функциями и элементами меню. Например, пользователь текстового процессора MS Word, открывая файл, сразу видит библиотеки и папки с документами EDMS-системы (откуда он и выбирает необходимый ему документ). При сохранении документ автоматически помещается в базу данных EDMS-системы. То же относится и к другим офисным и специализированным программам.

**Гибкость и интегрируемость.** Платформа позволяет настраивать среду пользователей в соответствии с бизнес процессами, в которых они принимают участие. Собственный клиент системы обеспечивает работу и навигацию пользователей по всей локальной и распределенной структуре хранения, а также по гипертексту. Клиентская часть легко конфигурируется в редакторе форм, без специального программирования. Интерфейсом может быть также стандартный Internet Explorer или даже офисные приложения (такие, как MS Word или Excel), дополненные встроенными функциями системы документооборота. Система работает на всех основных платформах (Windows, Mac OS, Unix, Linux), поддерживает различные СУБД (MS SQL Server, Oracle, DB2 и др.) и способна функционировать в гетерогенной распределенной среде.

**Интеграция с офисными приложениями.** Платформа обеспечивает «бесшовную» интеграцию с программными продуктами MS Office и другими ODMA-совместимыми приложениями. Имеется также опция «жесткой» интеграции с офисными приложениями. В случае ее использования документы могут сохраняться только в заданном месте, локальное сохранение не допускается.

**Интеграция с почтовыми системами.** Платформа предусматривает использование как существующей системы доставки электронной почты (по крайней мере, MS Exchange и Lotus Notes), так и встроенной. Информационным содержанием управляют через привычный интерфейс почтовых клиентов, например MS Outlook. Все сообщения содержатся в едином централизованном хранилище, возможны их обработка, просмотр и повторное применение.

**Интеграция с ERP-системами.** Важнейшими особенностями ERP-систем являются следующие:

- ❖ охват практически всех видов деятельности, всех бизнес процессов предприятия, управление всеми ресурсами;

- ❖ первичным для ERP-систем является оптимизация производственных процессов, повышение эффективности производственной деятельности. Автоматизация таких участков, как бухгалтерский учет - вторична. При этом необходимая бухгалтерская отчетность может формироваться в разных модулях системы;

- ❖ возможность управлять не только отдельным предприятием, но и корпорациями, холдингами, объединениями;

- ❖ наличие в последних версиях полноценного доступа ко всем функциям через Интернет, присутствие подсистем управления цепочками поставок (SCM), управления отношениями с клиентами (CRM) и др.

ERP-системы, обеспечивающие автоматизацию жестко структурированных производственных и коммерческих процессов, порождают массу документов, которые также должны участвовать в общем документообороте организации. ERP-документы находятся в корпоративном хранилище данных СЭД, а отчеты на их основе могут автоматически рассылаться по списку адресатов средствами СЭД.

Схемы принятия решений на уровнях бизнес системы и бизнес процессов в ERP-системе примерно одинаковы ([http://ccc.ru/magazine/depot/98\\_04/0206.htm](http://ccc.ru/magazine/depot/98_04/0206.htm)). Каждое решение на предприятии проходит ряд стандартных фаз (см. рис.3). Понятно, что это некий универсальный алгоритм решения любой задачи, в том числе и в производственной сфере. Все прикладные алгоритмы, которые имеются в ERP-системе, выдают некоторые цифры или рекомендуемые действия. Это хорошо, но этого мало. Производство - это процесс непрерывной выработки решений. Система же ERP пока лишь инициирует процессы планирования, раздачи заданий и последующего контроля на основе договоров, поступающих на ее вход.

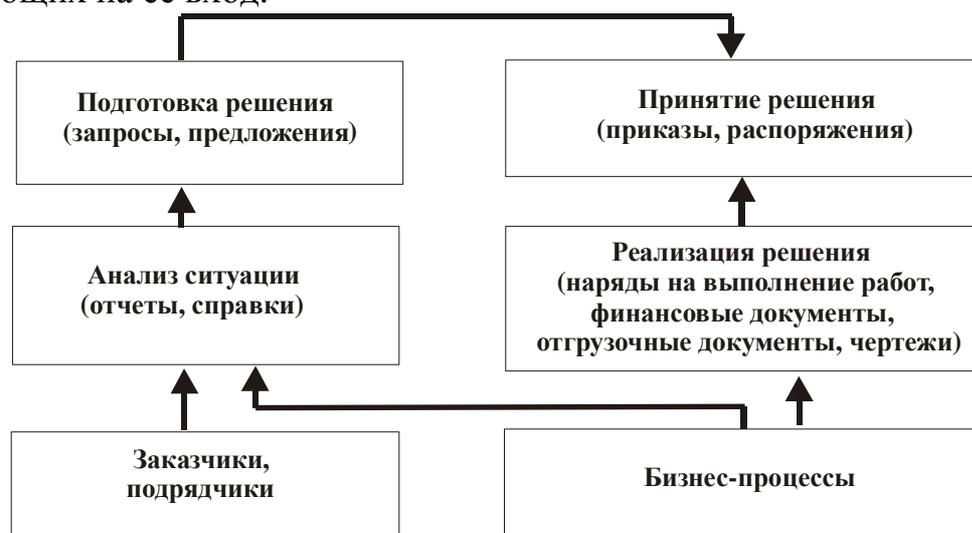


Рис. 3. Общий бизнес-цикл главного бизнес процесса производственного предприятия

Обеспечить выполнение алгоритма выработки решения (сбор информации - анализ - подготовка решения - принятие решения - реализация) может система документооборота. Речь идет даже не о реализации этого алгоритма, а о придании необходимой динамики процессу управления средствами системы документооборота.

ERP-система претендует скорее на роль поставщика информации для принятия решений. Конечно, пропускать абсолютно всю необходимую информацию через ERP-систему просто нереально, потому что решения принимаются на основе отчетов и сводок, показаний датчиков, сведений из газет, разговоров с коллегами, начальством и пр. Не всегда все источники информации вербальные. Таким образом, ERP-система - это не средство поддержки принятия решений, а инструмент реализации готовых договоров и заказов (все решения принимаются вне системы). Но поскольку через ERP-систему проходит экономически значимая информация и на нее замкнута практически вся экономическая жизнь предприятия, то она становится поставщиком информации для поддержки принятия решений.

Какими же функциями можно "нагрузить" СЭД в отношении поддержки принятия решений? Какая составляющая реально доминирует при принятии решения? Неизвестно. Если человек принимает решения не на основе информации, а на основе интуиции, то никакая компьютерная система ему помочь не сможет. Чем меньше доля интуитивности, тем более объективно проверяемым является решение. Автоматизация документооборота позволяет получать в едином потоке и вербализованную информацию, и давать выход своей интуиции. Коллегиальное обсуждение, которое должно завершиться к указанному сроку, опять-таки реализуется в рамках СЭД. На практике приходится сталкиваться с ситуацией противопоставления СЭД уровня предприятия и ERP-системы. Представители предприятий иногда не до конца понимают, что брать за основу при построении корпоративной информационной системы - СЭД или ERP-систему, ведь предприятие обычно в состоянии осилить (в финансовом смысле) только одну покупку такого масштаба. Перед руководителем встает дилемма: что выбрать? Поэтому и возникла гипотеза, кстати никем пока не проверенная, что возложение на СЭД функций бизнес приложений обойдется предприятию не дороже, чем переписывание модулей западной ERP-системы, зачастую малоподходящих к нашей реальности. Все строится на предположении, что дистанция (в смысле трудоемкости адаптации) от зарубежных ERP-систем и от СЭД до необходимого на предприятии ПО примерно одна и та же. Но, вообще говоря, это зависит от конкретных СЭД, ERP-системы и предприятия. Если на предприятии накоплен багаж хорошо работающих, но автономных программ, то логично вызывать их в узлах маршрутизации документов, заданных деловыми процедурами (workflow). Организация, контроль и передача информации в рамках деловых процедур, имеющих формализованные средства документирования - это сильная сторона СЭД. В этом случае разумно считать ее основой для построения корпоративной информационной системы, тем более что в СЭД корпоративного уровня обычно есть программные интерфейсы к ERP-системам. Если же предприятию надо приобрести сами алгоритмы автоматизации цепочки снабжение-производство-сбыт, то, естественно, надо отталкиваться от ERP-системы.

**Интеграция с системами проектирования.** Чертежи и другая информация, которая готовится в системах PDM, CAD и им подобных, также размещаются в общем корпоративном хранилище в формате CAD или одном из растровых, получаемых при оцифровке бумажных архивов. СЭД обеспечивает для них высокий уровень безопасности и другие преимущества, часто отсутствующие в системах проектирования.

**Интеграция с корпоративными информационными порталами.** Реализует Web-интерфейс, который обеспечивает сотрудникам предприятия персонализированное представление данных и бизнес приложений. Платформа имеет возможности интеграции с порталными приложениями (например, Citrix, Plumtree, SAP Portals и подобными им).

**Эффективность.** В результате разработки системы обеспечивается

рациональное соотношение между затратами на создание системы и целевыми эффектами, включая конечные результаты автоматизации документооборота. Необходима доступность широкого спектра дополнительных технологий для повышения уровня возврата от инвестиций.

**Проблемы стандартизации.** Как и любая область человеческой деятельности, сфера документооборота не могла избежать всеобщего веяния стандартизации и имеет свои проблемы.

**Проблема 1.** Архивная система должна быть интегрирована с приложениями, в которых порождаются различные электронные документы. Желательно, чтобы эта интеграция была прозрачной для пользователя, который работал бы с архивной системой напрямую, минуя обращения к файловой системе. Следовательно, диалоги операций с файловой системой должны быть заменены на диалоги работы с архивной системой. Единственным решением удовлетворить как производителей приложений, так и производителей архивных систем является выработка единого стандарта взаимодействия между системами такого класса. Этой цели достигла первая версия стандарта ODMA (Open Document Management API). Однако, как все универсальное, ODMA содержит определенные ограничения и всегда оказывается «запасным» вариантом взаимодействия с СЭД, когда, по какой-то причине, нет возможности реализовать более полноценную интеграцию. К примеру, несмотря на то, что начиная с версии Office 97 в продуктах Word и PowerPoint поддерживается ODMA, практически все производители СЭД поставляют специальные макросы для интеграции с MS Office. Правда, есть случаи, когда протокол ODMA оказывается вполне эффективным. К примеру, система ABBYY FineReader, начиная с версии 4.0, поддерживает ODMA, что позволяет пользователю, не обращаясь к производителю СЭД или к услугам интегратора, вводить бумажные документы в хранилище систем, поддерживающих этот протокол. К сожалению, ODMA не затрагивает вопросов взаимодействия между различными СЭД, однако другие стандарты имеют существенно меньшее применение.

На сегодняшний день данный интерфейс поддерживается следующими производителями архивных систем: PC DOCS, Saros, Novell (Soft Solutions), Watermark, Documentum и со стороны производителей приложений компаниями Corel (Corel WordPerfect Suite) и Microsoft (Office 2000).

**Проблема 2.** Иногда предприятие использует одновременно несколько систем управления документами. В качестве примера можно привести транснациональную и многопрофильную корпорацию DuPont. В подразделениях, которые ведут разработку новых химических продуктов, исторически используют Documentum; новые подразделения остановили свой выбор на DOCS Open, как на более дешевом решении в расчете на одного пользователя. Соответственно возникает проблема, как пользователю с одного рабочего места иметь доступ к нескольким архивным серверам для поиска документов. Для обеспечения совместной работы нескольких

архивных серверов предназначен стандарт ODMA.

**Проблема 3.** Аналогичная проблеме 2, но для систем класса workflow. Выработкой стандарта для совместной работы workflow-систем от различных производителей занимается некоммерческая организация WorkFlow Coalition, а выработанная ею спецификация носит название Workflow Coalition API.

**Проблема 4.** При работе с образами документов важна унификация используемых форматов. В качестве единого формата для черно-белых образов документов был принят формат TIFF GROUP IV. Для распространения электронных документов принимается формат, разработанный компанией Adobe - PDF.

**Основные задачи решаемые СЭД:**

- ❖ обеспечение более эффективного управления за счет автоматического контроля выполнения, прозрачности деятельности всей организации на всех уровнях;
- ❖ разработка бизнес процессов компании и представление их в виде графических схем;
- ❖ разработка методологии технологического учёта компании, технологических планов счетов и типовых проводок (учетной политики);
- ❖ разработка и настройка соответствующей прикладной части системы, подключение её к ядру СЭД и "наполнение" заготовок бизнес процессов прикладными процедурами;
- ❖ поддержка системы контроля качества, соответствующей международным нормам;
- ❖ протоколирование деятельности предприятия в целом (внутренние служебные расследования, анализ деятельности подразделений, выявление "горячих точек" в деятельности);
- ❖ оптимизация бизнес процессов и автоматизация механизма их выполнения и контроля;
- ❖ исключение или максимально возможное сокращение оборота бумажных документов на предприятии. Экономия ресурсов за счет сокращения издержек на управление потоками документов в организации;
- ❖ исключение необходимости или существенное упрощение и удешевление хранения бумажных документов за счет наличия оперативного электронного архива;
- ❖ быстрое создание отчётности в Генераторе Отчётов.

**Технологичность управления и обработки данных. Структуры компьютерных сетей и распределенные базы данных в виде справочников для формирования системы документооборота**

*Технологичность — это возможность изменять параметры системы при ее внедрении или эксплуатации с целью получения модели, наиболее адекватно реализующей требования заказчика.*

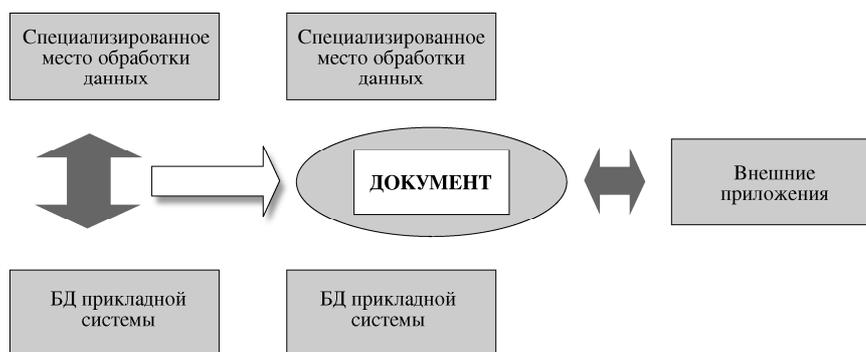


Рис. 4. От работы с базой данных к работе с объектами

С одной стороны вопросы технологичности определяют требования к разработке — несмотря на возрастающий функционал системы, она не должна становиться тяжелой в работе и сложной в понимании.

С другой — необходимо при внедрении предоставить максимальные возможности по настройке как самих объектов (различные атрибуты, формы представления объектов перед пользователями в зависимости от различных контекстов состояния системы) прикладной системы, так и правил их поведения (бизнес-логика обработки отдельных объектов системы и связанное изменение параметров нескольких объектов).

Можно выделить три основных момента, которые присущи любому предприятию вне зависимости от его рода деятельности.

1. *Информация* — обобщенное определение для различных документов, обрабатываемых сотрудниками, от специализированных, скажем финансовых (счета, накладные и т. п.), хранение и обработка которых осуществляется специализированными программами, до общих организационно-распорядительных документов (приказы, распоряжения, указания), которые имеют форматы различных офисных приложений, посредством которых они создаются и редактируются.

2. *Маршрут прохождения информации* — список лиц, групп и последовательность, в которой они должны обработать эти данные, а также временные сроки выполнения работ.

3. *Операции над документом* — действия, которые необходимо совершить над документом на различных этапах его обработки: редактирование, визирование, утверждение, выполнение проводки и т. п. Операции регламентируются, с одной стороны, правами доступа выполняющих отдельные этапы сотрудников, а с другой — различными аналитическими операциями преобразования и обработки документов.

**Цель создания приложений для управления деловыми процессами.** Часто, наряду с обработкой данных, необходимо организовать очень четкое взаимодействие сотрудников организации, участвующих в формировании и анализе информации — кто, когда и что должен предпринять. Наиболее оптимальный способ наделить свой продукт такими способностями состоит в использовании специализированного ядра workflow-системы.

Чтобы выполнить возложенные поручения, сотрудник вынужден работать с офисными приложениями (скажем, для оформления договоров), с почтой (дабы вести переписку с респондентами) и Web-навигатором (допустим, чтобы получать самую актуальную информацию о конкурентах). Итого, в простейшем случае получается 3 приложения, к которым обычно прибавляется еще и планировщик (как правило, средство ведения дел от компании Symantec или Lotus). Итак, для того чтобы успешно выполнять свою работу, пользователь переключается между приложениями с помощью комбинации клавиш <ALT+TAB>. Даже отсюда видно, что чем большее количество приложений участвует в формировании данных для выполнения предполагаемого полезного действия, тем более сложную комбинацию управления прикладными средствами приходится осваивать сидящему за компьютером, что неизбежно повышает вероятность ошибок.

Использование универсального ящика заданий создает универсальную оболочку, которая позволяет унифицировано интегрировать между собой различные прикладные системы, что непременно высоко оценят, к примеру, экономисты, конструкторы, юристы или другие сотрудники, которые обычно хорошие специалисты в своей узкой области, но непрофессиональные пользователи компьютеров.

Изменение схемы работы пользователя в различных архитектурных решениях приводится на рис. 5 и 6.

На рис.5 приведена привычная, но крайне неудобная схема взаимодействия.



Рис.5. Стандартная схема работа с приложениями

1. *Пользователь — данные — приложение обработки данных*, что выражается в изолированности пользователя от данных, которые ему необходимо обработать (проще говоря, суть заключается в том, каким образом сотрудник узнает, что именно и к какому сроку он должен сделать). Этот вопрос отчасти упрощается, когда пользователь работает с документами

одного типа, что может выражаться в едином канале доставки информации (например, документы приходят по электронной почте, или финансовый документ всегда можно загрузить, находясь с специализированном рабочем месте, периодически выполняя запрос на обновление информации).

Но картина резко меняется, когда типов документов, с которыми приходится работать сотруднику, несколько. Допустим, он участвует в обработке финансовых документов, ведет переписку с клиентами и, вероятнее всего, знакомится с распоряжениями и приказами по своей компании. (Бегло прикинуть количество приложений, между которыми должен будет переключаться пользователь, дабы аккуратно выполнять свою работу, довольно сложно). Заметьте, что при этом о необходимости выполнить поручение он узнает только после того, как дополнительно нажмет на магическую комбинацию клавиш, специфическую для каждого приложения.

2. Не учитывается очень важный фактор контроля за ходом обработки данных (здесь рассматриваем плоскость *Данные — срок обработки — исполнитель — контролирующие лица*). Зачастую это выражается просто в различных выборках данных, имеющих определенные атрибуты, что сопряжено с дополнительными манипуляциями пользователя.

Обработка любого объекта (данных) подразумевает, наряду с механизмами изменения состояния (атрибутов) самого объекта, также наличие маршрута его обработки сотрудниками компании (кто, в какой последовательности, в течение какого времени и что именно должен сделать), для чего, вполне естественно, целесообразно использовать специализированную систему управления деловыми процессами, где они досконально определены.

Кроме того, для того чтобы избежать невыполнения по причине забывчивости сотрудников, система должна в соответствии с регламентом обработки данных выполнять напоминание как лицам, которым эта работа поручена, так и лицам, которые отвечают в целом за обработку конкретного типа данных и отслеживают все нюансы, связанные с их прохождением.

3. Гибкость в изменении маршрутов (или схем) обработки данных (плоскость *Данные — сценарий обработки*). Устоявшаяся схема создания приложений подразумевала ввод ограничений на возможные значения и порядок изменения одного или нескольких связанных атрибутов объекта с дополнительным запрещением выполнения зависящих от них операций. Все эти вещи задавались жестко в таблицах настройки на этапе внедрения системы или же вообще являлись неизменными. Нетрудно представить себе проблемы людей, отвечающих за функционирование такого комплекса, когда по ряду причин (структурная реорганизация и т. п.) приходилось изменять схемы обработки данных.

Совсем другая картина получается, когда на все происходящее взглянуть через призму списка заданий на выполнение (см. рис. 6).

Для системы управления деловыми процессами совершенно неважно, каково содержание данных, это возлагается на механизмы прикладной

системы (вот здесь как раз оптимальна работа с объектами прикладной системы, так как именно они инкапсулируют всю внутреннюю бизнес-логику взаимодействия объектов, а посредством методов можно гибко извне управлять их поведением).

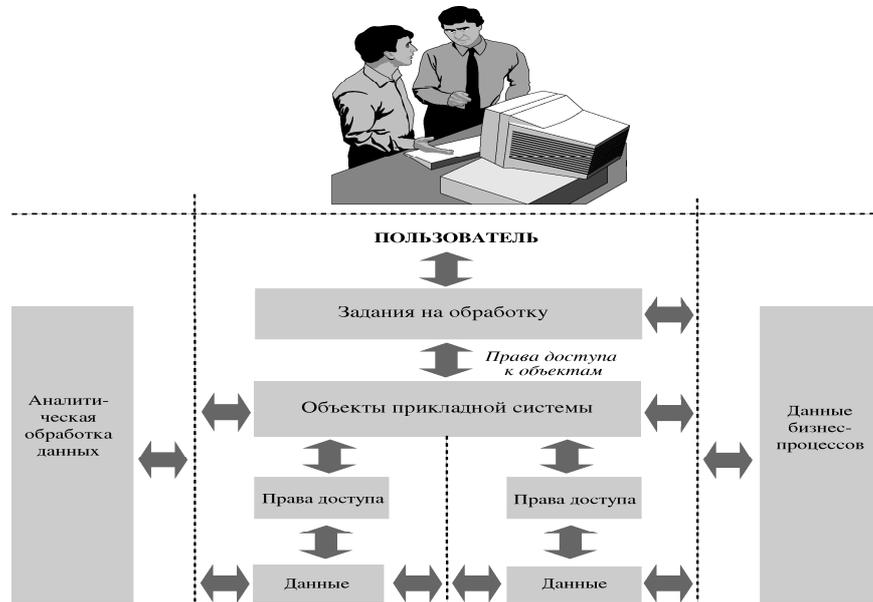


Рис. 6. Построение информационной системы через управление заданиями

Говоря о бизнес логике объектов, следует отметить, что часто она реализуется посредством более или менее сложных механизмов, «автоматов», что, естественно, отражается на предоставляемых возможностях, когда нельзя говорить о сложной системе в целом, а возможно определить только состояние отдельного ее объекта.

Только такая структура прикладной системы (см. рис.7) позволит безболезненно подключать к ней дополнительные функциональные модули без остановки комплекса и без необходимости переучивать пользователей для нового рабочего места, ведь сотрудники все получают в одном месте — ящике «информация для обработки», где находятся задания, связанные с выполнением различных поручений, обработкой документов, приказы, распоряжения, сюда же может попадать информация из личного планировщика и т. д.



Рис.7. Взаимодействие модулей информационной системы

**Общие требования к компонентам комплексной системы.** Практически всегда при создании прикладной системы борются две противоположности (хотя на первый взгляд, это две равноправные и дополняющие друг друга сущности): требуемый функционал и технологичность получаемого комплекса, к которым в реальных условиях подключается еще и фактор времени, в течение которого эта система должна быть готова. Поэтому, наряду с имплантацией в свою систему workflow-ядра, необходимо привести уже имеющиеся у вас механизмы в соответствие с перспективными веяниями архитектуры программирования, основным из которых является *объектно-ориентированный подход* при проектировании и создании системы (при внешней тяжеловесности на первый взгляд данного решения, в дальнейшем, благодаря этому, вы сэкономите массу сил и времени как своим программистам, так и возможным партнерам, которые будут использовать ваши технологии при создании вертикальных решений). Прежде всего делается акцент на выделение объектов прикладной системы (формализуются типы документов), с которыми придется работать сотрудникам, и модулей, которые заключают в себе аналитику обработки данного объекта (на тривиальном уровне — если информация представляет собой документ формата MS Word, тогда аналитический блок должен обеспечить как минимум операции открыть документ, создать новый, закрыть документ, удалить документ). Любую систему необходимо рассматривать с точки зрения перспектив развития и открытости, а для этих целей использование и модификация объектов и методов их обработки подходит как нельзя лучше. Если для получения информации об отвечающем за обработку объекта деловом процессе необходима открытость workflow-системы, то для корректной работы в обратную сторону — отобразить

специализированную форму системы, управлять внутренними атрибутами объектов и т. п. — нужна прозрачность с другой стороны (из чего следует очевидная истина — закрытая система работает только в одну сторону).

**Классификация рынка СЭД:** Аналитические компании IDC и GartnerGroup выделяют шесть категорий технологий, которые составляют рынок средств систем электронного управления документами. Следует заметить, что ни одна классификация, видимо, не является идеальной. В результате некоторые продукты одновременно попадают в несколько категорий и имеют возможности, характерные для продуктов из разных категорий. Ниже перечислены категории технологий СЭД с примерами наиболее известных поставщиков и продуктов в каждом классе:

- ❖ системы СЭД, ориентированные на бизнес-процессы (Business-process EDM): Documentum, FileNet (Panagon и Watermark), Hummingbird (PC DOCS);

- ❖ корпоративные системы СЭД (Enterprise-centric EDM): Lotus (Domino Doc), дополнения к Novell GroupWise, Opent Text (LiveLink), Keyfile Corp., Oracle (Context);

- ❖ системы управления контентом (Content management): Adobe, Excalibur;

- ❖ системы управления информацией (порталы) (Information Management): Excalibur, Oracle Context, PC DOCS/Fulcrum, Verity, Lotus (Domino/Notes, K-station);

- ❖ системы управления образами (Imaging);

- ❖ системы управления потоками работ (Workflow management): Lotus (Domino/Notes и Domino Workflow), Jetform, FileNet, Action Technologies, Staffware;

Дадим более подробные определения и отличительные признаки каждой категории из вышеперечисленных технологий.

**Системы СЭД, ориентированные на бизнес процессы**, как правило, предназначены для специфических вертикальных и горизонтальных приложений, иногда ориентированные на использование в определенной индустрии. Эти решения, как правило, обеспечивают полный жизненный цикл работы с документами, включая технологии работы с образами, управления записями и потоками работ, управление контентом и т.д.

**Корпоративные системы СЭД** обеспечивают корпоративную инфраструктуру для создания, совместной работы над документами и их публикации, доступную, как правило, всем пользователям в организации. Основные возможности этих систем аналогичны системам, ориентированным на бизнес процессы. Однако, их отличительной особенностью является способ использования и распространения. Аналогично таким средствам как текстовые редакторы и электронные таблицы, корпоративные системы СЭД являются стандартным, "приложением по умолчанию" для создания и публикации документов в организации. Как правило, эти средства не ориентированы на использование только в какой-то определенной индустрии или для узко определенной задачи. Они предлагаются и внедряются как общекорпоративные технологии,

доступные практически любой категории пользователей.

**Системы управления контентом** обеспечивают процесс отслеживания создания, доступа, контроля и доставки информации вплоть до уровня разделов документов и объектов для их последующего повторного использования и компиляции. Потенциально доступность информации не в виде документов, а в меньших объектах облегчает процесс обмена информацией между приложениями.

**Системы управления информацией**, называемые также порталами, обеспечивают агрегирование, управление и доставку информации через сети Internet, intranet и extranet. Эти технологии обеспечивают фундамент создания информационных порталов. Системы управления информацией дают возможность организациям накапливать и использовать экспертизу в распределенной корпоративной среде на основе использования бизнес правил, контекста и метаданных. Хотя большинство доступных сегодня технологий обеспечивают, в основном, статические публикации, обеспечение большей интерактивности и средств совместной работы – дело ближайшего будущего.

**Системы управления образами** преобразуют информацию с бумажных носителей в цифровой формат, как правило, это TIFF (Tagged Image File Format), после чего документ может быть использован в работе уже в электронной форме.

**Системы управления потоками работ (workflow)** обеспечивают систематическую маршрутизацию работ любого типа в рамках структурированных и неструктурированных бизнес процессов. Они используются в целях ускорения бизнес процессов, увеличения эффективности и степени контролируемости процессов в организации.

В области работы с документами начинается новый виток спирали - объединение отдельных технологий в интегрированное решение, получившее с легкой руки Gartner Group название Enterprise Content Management (ECM). Пока этот термин не имеет однозначного русского аналога, а дословный перевод «управление информационными ресурсами предприятия» не слишком точно отражает идеи и принципы, на которых базируется ECM. Ясно одно - предприятия все больше нуждаются в единой платформе для автоматизации своей деятельности и бизнес процессов, связанных как с традиционными формами документов, циркулирующих внутри организации, так и с документами Internet и мультимедиа.

**Требования к СЭД, ориентированным на отечественное делопроизводство:**

- ❖ осуществлять ведение пользователей на основе организационно-штатной структуры организации;

- ❖ обеспечивать единую регистрацию всей поступающей корреспонденции, включая письма и обращения граждан, с последующим направлением документов на рассмотрение руководству организации или в ее структурные подразделения;

- ❖ обеспечивать разделение доступа к документам на уровне отдельных

пользователей, по ролевому принципу, и на основе иерархической структуры организации;

- ❖ вести журналы регистрации и контроля исполнения;
- ❖ контролировать сроки исполнения документов, оповещать исполнителя и делопроизводителя о приближении сроков контроля, о невыполненных в срок документах;
- ❖ обеспечивать надежное хранение документов и их описаний;
- ❖ поддерживать шаблоны документов, составные документы, версии и подверсии, перекрестные ссылки между документами;
- ❖ отслеживать документы вне системы, осуществлять выписку документов из системы;
- ❖ поддерживать иерархию категорий для эффективного поиска документа;
- ❖ осуществлять поиск документов: атрибутивный, полнотекстовый, нечеткий поиск;
- ❖ поддерживать разработку документов на стадии проекта, включая коллективную разработку;
- ❖ обеспечивать жизненный цикл документа (его создание, хранение версий, публикация, блокировка доступа к изъятому документу, передача документа для хранения в архиве);
- ❖ поддерживать визирование, согласование, утверждение документов;
- ❖ осуществлять движение документов - документооборот, поддерживать все виды маршрутизации, автоматическую рассылку уведомлением, обмен сообщениями и поручениями внутри системы, формировать реестры отправки во внешние организации;
- ❖ допускать задание пользователем различных типов документов, создания и редактирования карточек для них;
- ❖ вести классификаторы документов (по типу, виду и т.п.), справочники внешних и внутренних организаций, др. справочники;
- ❖ осуществлять жесткое разграничение полномочий в системе, поддерживать роли, осуществлять протоколирование и аудит действий пользователей;
- ❖ алгоритмы шифрования при хранении и передаче данных, цифровую подпись;
- ❖ вести дела документов, поддерживать функцию списания документов в дело, передачу дел на хранение в архив;
- ❖ необходимо наличие развитых средств администрирования;
- ❖ поддерживать удаленный доступ к информации, в том числе территориально распределенные организации;
- ❖ единую регистрацию всей исходящей корреспонденции и внутренних документов организации;
- ❖ формировать требуемые отчеты, в т.ч. статистические отчеты по делопроизводству организации.

**Составляющие эффективности СЭД, при этом, следующие:**

- ❖ на 20-25% возрастает производительность труда персонала при

использовании электронного документооборота;

- ❖ стоимость архивного хранения электронных документов на 80% ниже в сравнении с бумажными;

- ❖ во всех подразделениях и в организации в целом вводится унифицированная, формализованная и строго регламентированная технология делопроизводства;

- ❖ организация становится полностью управляемой. Появляется возможность ответить на любой вопрос по документам и исполнителям, осуществлять анализ и управление документационной деятельностью;

- ❖ поскольку компьютерная сеть может охватывать не только центральный офис организации, но и ее территориально-удаленные подразделения, то управляемость может распространяться на всю территориально-распределенную структуру организации;

- ❖ система автоматизации делопроизводства, по сути, является носителем строго формализованной и строго документированной технологической информации о правилах и порядке работы с документами. В результате уменьшается зависимость организации от персонала как физического носителя технологических знаний и правил работы с документами;

- ❖ создаются условия для резкого ускорения прохождения документов по организации, особенно при организации электронного документооборота. В государственных структурах - это одна из проблем эффективности функционирования госаппарата. В коммерческих структурах это позволяет повысить конкурентоспособность организации, получить материальный выигрыш за счет более быстрой реакции на изменение конъюнктуры;

- ❖ минимизируется трудоемкость делопроизводственных операций. При этом, однако, нужно иметь в виду, что необходимость ввода полной и точной информации о документе, скажем, при его первичной регистрации может потребовать дополнительных усилий на некоторых рабочих местах, тогда как трудоемкость работы на других рабочих местах, использующих эту информацию, может сократиться, как показывает опыт, в несколько раз;

- ❖ качественный выигрыш достигается при организации взаимоувязанного электронного документооборота между организациями, поскольку полностью отпадают проблемы, связанные с изготовлением и пересылкой бумажных документов, а затем - в повторном вводе реквизитов и текстов полученных документов;

- ❖ СЭД позволяет решить ключевой вопрос делопроизводства. Она позволяет реализовать любую степень децентрализации делопроизводства при одновременном обеспечении централизованного учета и контроля. Вплоть до того, что каждый специалист может самостоятельно регистрировать документы и направлять их для дальнейшей работы (в пределах своей компетенции, конечно), находясь при этом под полным контролем своих руководителей. Таким образом, организация может динамично перестраивать свою управленческую структуру без потери управляемости;

❖ создаются предпосылки к реализации более эффективных схем документационного управления. В традиционной практике документы в подразделения организации, будь то локальные или удаленные, направляются по цепочке от руководителя к руководителю до конкретных исполнителей. Это позволяет руководителю каждого уровня эффективно контролировать деятельность своих сотрудников, однако длинные и зачастую формальные цепочки резко снижают эффективность управления. Правильно построенная СЭД позволяет направлять документы непосредственно лицам, которые будут исполнять поручения, при этом руководители сохраняют полный контроль как за прохождением самого поручения, так и за его исполнением;

❖ появляются предпосылки для организации электронного документооборота в том объеме, к которому готова сама организация. Правильно построенная СЭД должна работать как с данными о прохождении и исполнении документов (независимо от того, бумажные они или электронные), так и с самими электронными документами, к которым эти данные относятся. Включение электронных документов в делопроизводственный цикл позволяет перейти на качественно новый уровень эффективности работы с документами, поскольку технологии работы с электронными документами (редактирование, перемещение, тиражирование и др.) принципиально более эффективны;

❖ снимается часть проблем, связанных с территориальной удаленностью подразделений организации или индивидуальных рабочих мест (в том числе мобильных) ее сотрудников. СЭД позволяет организовать нормальную работу с документами всех сотрудников независимо от местонахождения их рабочих мест (конечно, при наличии линий связи);

❖ СЭД позволяет обеспечить внедрение современных систем управления ресурсами предприятия (так называемые ERP-системы). Системы управления ресурсами обычно содержат лишь данные о бизнес процессах и не хранят документов как таковых, в то время как СЭД может взять на себя функции документационного контура управления, в том числе связанного с бизнес процессами;

### Методы проектирования систем ведения электронной документации в информационной среде

Информация должна рассматриваться не только как некая принадлежащая предприятию ценность, но и как исходная точка для построения информационной системы, обслуживающей предприятие. На практике во многих организациях пришли к убеждению, что правильное понимание информационного аспекта служит необходимой предпосылкой для построения высококачественных и целостных информационных систем. По этой причине ведущим направлением в разработке корпоративной информационной системы (КИС) является переход от процедурно-ориентированных методов разработки к информационно-ориентированным.

Информационное проектирование становится самой популярной методологией, пронизывающей все стадии жизненного цикла системы.

## **Информационное проектирование**

Информационное проектирование зачастую характеризуется как подход к созданию систем, сконцентрированный на информации. Это также всеобъемлющая стратегия, основанная на информационном планировании и уяснении целей системы. Главная посылка, на которой строится данная стратегия, состоит в том, что целостность информационных систем определяется степенью охвата информационных элементов объекта соответствующей логической моделью.

Информационное проектирование предполагает прежде всего глубокое исследование всех информационных систем, преследующее своей целью поиск ответа на вопрос: каким образом информация используется и разделяется системами. *Программные структуры выступают по отношению к информационной модели уже "надстройкой", определяющей общую информационную инфраструктуру для создания соответствующих систем на предприятии.* Таким образом, процедуры логически вытекают из данных, а качество информационных систем определяется качеством построения информационной модели.

Информационное обследование организации - трудоемкая и продолжительная работа. Только опытным аналитикам, вооруженным мощным инструментарием и технологией, под силу обеспечить точность и целостность описания информации.

Моделирование взаимосвязей между сущностями (объектами) представляет собой метод информационного проектирования, предназначенный для построения высококачественной информационной модели. Информационное моделирование служит стандартным средством определения данных и взаимосвязей между ними, применяемым во всех информационных системах. Оно существенно повышает качество системы и продуктивность программного обеспечения.

Моделирование взаимосвязей между сущностями (МВМС) является методом описания информационных запросов, циркулирующих в организации. В простейшем виде МВМС ставит своей целью поиск ответов на следующие вопросы: какие в организации имеются объекты, заслуживающие внимания (назовем их сущностями), какими эти объекты обладают качествами и свойствами (назовем их атрибутами) и какие между этими сущностями и атрибутами просматриваются взаимосвязи. Все это должно рассматриваться в контексте поставленной задачи и с учетом всех аспектов ее реализации в данной информационной модели.

За последние годы сформировалось новое направление в создании СЭД - CASE (Computer-Aided Software/System Engineering). В настоящее время не существует общепринятого определения CASE. Содержание этого понятия обычно определяется перечнем задач, решаемых с помощью CASE, а также

совокупностью применяемых методов и средств. Грубо говоря, CASE-технология представляет собой совокупность методологий анализа, проектирования, разработки и сопровождения сложных систем СЭД, поддержанную комплексом взаимосвязанных средств автоматизации. CASE - это инструментарий для системных аналитиков, разработчиков и программистов, позволяющий автоматизировать процесс проектирования и разработки СЭД.

К настоящему моменту направление CASE оформилось в самостоятельное наукоемкое направление в разработке СЭД, повлекшее за собой образование мощной CASE-индустрии, объединившей сотни фирм и компаний различной ориентации. Среди них выделяются компании-разработчики средств анализа и проектирования СЭД с широкой сетью дистрибьютерских и дилерских фирм; фирмы-разработчики специальных средств с ориентацией на узкие предметные области или на отдельные этапы жизненного цикла СЭД; обучающие фирмы, которые организуют семинары и курсы подготовки специалистов; консалтинговые фирмы, оказывающие практическую помощь при использовании CASE-пакетов для разработки конкретных приложений; фирмы, специализирующиеся на выпуске периодических журналов и бюллетеней по CASE. Основными покупателями CASE-пакетов за рубежом являются военные организации, центры обработки данных и коммерческие фирмы по разработке СЭД.

CASE позволяет не только создавать "правильные" продукты, но и обеспечить "правильный" процесс их создания. Основная цель CASE состоит в том, чтобы отделить проектирование СЭД от его кодирования и последующих этапов разработки, а также скрыть от разработчиков все детали среды разработки и функционирования СЭД. Чем больше деятельности будет вынесено в проектирование из кодирования, тем лучше.

При использовании CASE-технологий изменяются все этапы жизненного цикла программной системы, при этом наибольшие изменения касаются этапов анализа и проектирования.

**Жизненный цикл СЭД и его критические этапы.** В основе деятельности по созданию и использованию программного обеспечения (ПО) СЭД лежит понятие его жизненного цикла (ЖЦ). ЖЦ является моделью создания и использования ПО, отражающей его различные состояния, начиная с момента возникновения необходимости в данном программном изделии и заканчивая моментом его полного выхода из употребления у всех пользователей.

Традиционно выделяются следующие основные этапы ЖЦ ПО СЭД:

- ❖ анализ требований;
- ❖ проектирование;
- ❖ кодирование (программирование);
- ❖ тестирование и отладка;
- ❖ эксплуатация и сопровождение.

ЖЦ образуется в соответствии с принципом нисходящего проектирования и, как правило, носит итерационный характер:

реализованные этапы, начиная с самых ранних, циклически повторяются в соответствии с изменениями требований и внешних условий, введением ограничений и т.п. На каждом этапе ЖЦ порождается определенный набор документов и технических решений, при этом *для каждого этапа исходными являются документы и решения, полученные на предыдущем этапе*. Каждый этап завершается верификацией порожденных документов и решений с целью проверки их соответствия исходным.

## Разработка офисных информационных систем

В данном разделе рассматриваются вопросы разработки офисных информационных систем (ИС) с учетом задач бизнес-реинжиниринга в их актуальной трактовке. Рассмотрим правила выполнения BPR - business process reengineering - по Хаммеру/Чампи. Рассмотрение будем проводить в следующих направлениях:

- ❖ общих основополагающих положений бизнес-анализа;
- ❖ основных положений и правил BPR;
- ❖ особенностей применения средств Workflow в условиях BPR.

В основе технологии Workflow лежат следующие понятия:

**объект** - информационный, материальный или финансовый объект, используемый в бизнес-процессе (например транзакция, письмо, оборудование, счет);

**событие** - внешнее (не контролируемое в рамках процесса) действие, произошедшее с объектом (скажем обработка транзакции, получение письма, поломка оборудования, изменение ставки налога);

**операция** - элементарное действие, выполняемое в рамках рассматриваемого бизнес-процесса (допустим подготовка письма, замена оборудования, оплата счета);

**исполнитель** - должностное лицо, ответственное за выполнение одной или нескольких операций бизнес-процесса (к примеру менеджер, сотрудник архива, директор).

Взаимоотношения между базовыми понятиями технологии Workflow отображены на концептуальной информационной модели, представленной на рисунке 8.

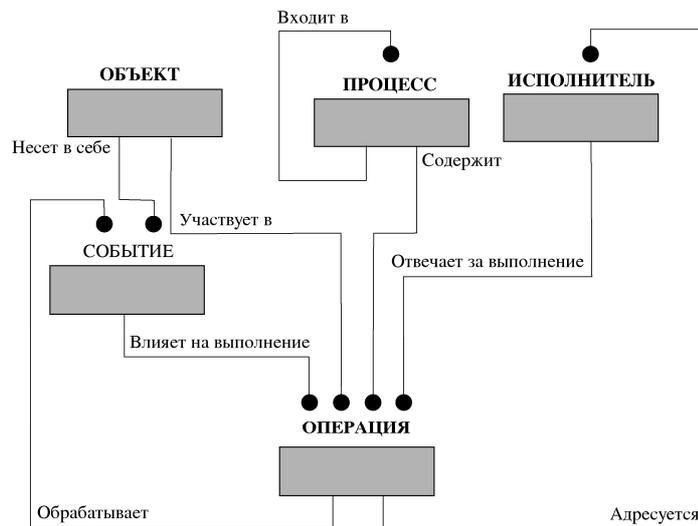


Рис. 8. Концептуальная информационная модель технологии Workflow

Несмотря на то что модель подготовлена в соответствии с требованиями стандарта IDEF1X, ее общая интерпретация и анализ не требуют специального изучения правил используемой методологии. В рамках этой модели ПРОЦЕСС состоит из ОПЕРАЦИЙ и других ПРОЦЕССОВ. ОПЕРАЦИЯ адресуется ИСПОЛНИТЕЛЯМ, которые, в свою очередь, отвечают за выполнение одной или нескольких ОПЕРАЦИЙ. ОБЪЕКТЫ участвуют в выполнении ОПЕРАЦИИ. СОБЫТИЯ могут влиять на выполнение ОПЕРАЦИЙ, например, изменяя результат операций или последовательность их выполнения. ОПЕРАЦИИ обрабатывают СОБЫТИЯ, являясь реакцией системы на происходящие СОБЫТИЯ. Жизненный цикл ОБЪЕКТА связан с внешними СОБЫТИЯМИ и ОПЕРАЦИЯМИ, выполняемыми в составе ПРОЦЕССА.

**Разработка единой структуры управления автоматизированным комплексом.** На основании декомпозиции процессов в объекте деятельности и структуры (DFS) в автоматизированном комплексе должны быть сформированы, как минимум, 3 уровня иерархии управления:

- ❖ первый уровень генерального управления стратегией функционирования предприятия (метауправление). Этот уровень работает на верхней части иерархии целей.

- ❖ второй уровень – управление функциональными подразделениями предприятия. Этот уровень работает в средней части иерархии целей, обычно это финансово-хозяйственная деятельность.

- ❖ третий уровень служит для управления технологическими и бизнес-процессами в функциональных подразделениях. Уровень работает на нижней части иерархии целей, являясь источником информации и местом приложения управляющих воздействий.

В зависимости от конкретного характера, вида и объема деятельности предприятия количество уровней иерархии управления может возрастать. Далее, в соответствии с функциональным назначением алгоритмических модулей в структуре должны быть сформированы подразделения СЭД

(отделения, отделы, группы, цеха, и т.д.), выполняющие одну или несколько законченных функций по контролю и управлению процессами в объекте деятельности и его связи с внешними системами.

## Информационные потоки

Информационный поток — информация, рассматриваемая в процессе ее движения в пространстве и времени в определенном направлении.

Правильный выбор или разработка программных продуктов для автоматизации информационных потоков в рамках информационных систем — первейшая задача современных организаций.

У любых информационных потоков есть общий источник и общий приемник. Поток, состоящий из смысловых структурных элементов, называют *сообщением*.

При управлении количеством и качеством информации происходит ее возникновение, прием, передача и переработка. *Действие информации* заключается в снятии *неопределенности ситуации*. В АСУ одной из задач является *передача управленческому персоналу минимума информации, необходимой для определения состояния производства и принятия решения*.

Массивы информации (БД) в системе должны быть оптимально организованы на основе использования единой системы классификации и кодирования технико-экономической информации, унифицированных систем документации. Информационное обеспечение АСУ должно предоставить всем функциональным подсистемам необходимую информацию в требуемом объеме, в требуемые сроки и в удобной для использования форме. В процессе управления для создания информационного обеспечения (ИО) должны быть осуществлены:

1. сбор информации о состоянии внешней среды и объекте управления, т. е. *создание информации, называемой первичной, текущей, входной*;
2. подготовка и сборка информации в соответствии с некоторой моделью управления, т. е. *создание промежуточной информации*;
3. выработка управляющих воздействий, т. е. *создание оперативной и управляющей информации*.

При сборе информации *исходными документами* являются те, которые служат источниками информации для других документов. *Производные документы* (показатели) формируются на основании других документов. *Первичные документы непосредственно отражают входную информацию. Конечные документы — выходные документы, а также непосредственно влияющие на объект управления.*

*Формированием документа* (показателя) называется процесс перехода от *исходных* документов (показателей) к *производным*. Не всегда выполняют преобразование исходных показателей, иногда часть показателей документа просто переносят в другой документ.

В АСУ на нижних уровнях действует детальная информация о состоянии объекта управления. По мере движения информации от нижних уровней управления к высшим она должна быть избавлена от лишних подробностей, бесполезной детализации. Этот процесс называется *интеграцией* или *сжатием информации*, а сама информация — *осведомительной*.

*Распорядительная* информация (идущая от верхних уровней к нижним) на средних уровнях «размножается», так как средние уровни генерируют дополнительную информацию.

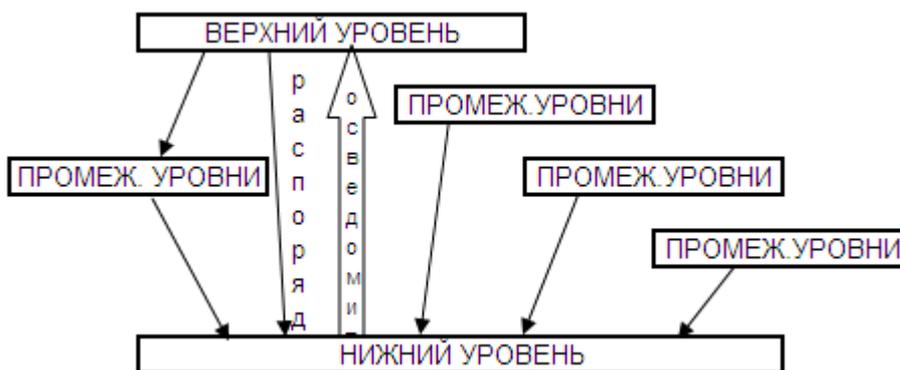


Рис. 9. Схема движения ИП

Степень интеграции или степень размножения  $\beta$  оценивается коэффициентом  $\beta = \frac{J_{вых}}{J_{вх}}$ , где  $J$  — количество информации. Для осведомительной информации  $\beta \leq 1$ , а для распорядительной —  $\beta \geq 1$ .

### ФОРМИРОВАНИЕ ИНФОРМАЦИОННЫХ ПОТОКОВ

Прежде всего собирается информация о состоянии внешней среды и объекте управления. Она называется *первичной*. Информация, получаемая в результате обработки первичной информации (сортировки, группировки, выделения, вычислений и т. д.), называется *промежуточной* или *вторичной* информацией.

Информация, полученная для выработки управляющих воздействий, называется *управляющей* (оперативной), требующей немедленной реакции системы управления.

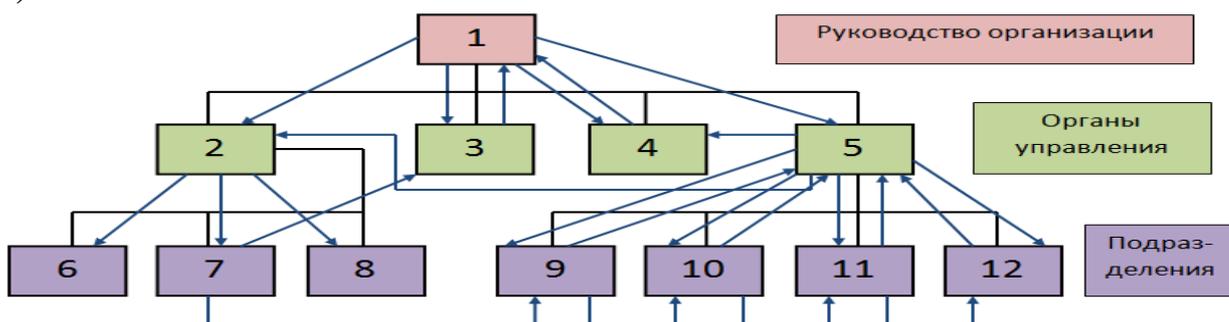
По критерию стабильности выделяют *переменную* и *постоянную* информацию. *Переменная* информация отражает фактическое состояние объекта и, как правило, участвует в одном цикле обработки. *Постоянную* информацию многократно используют для обработки переменной информации.

Создание информационного обеспечения автоматизированной информационной системы (АСУ) начинают с анализа информационных потоков: обследуют реальные потоки и анализируют полученные результаты.

Составляют структурную схему потоков информации. Обследуют и изучают существующие потоки, определяют реальные характеристики документов. Главная цель — выявить возможность автоматизации процессов сбора, передачи и переработки сообщений или их частей.

Эффективным инструментом структурирования информации является математическое моделирование. Создание модели ИП должно основываться на результатах детального обследования объекта управления, выявления закономерностей, функционирования в нем информационных потоков и определения решаемых задач. Должна быть определена связь логических и физических уровней организации (рис.10 ).

а)



б)

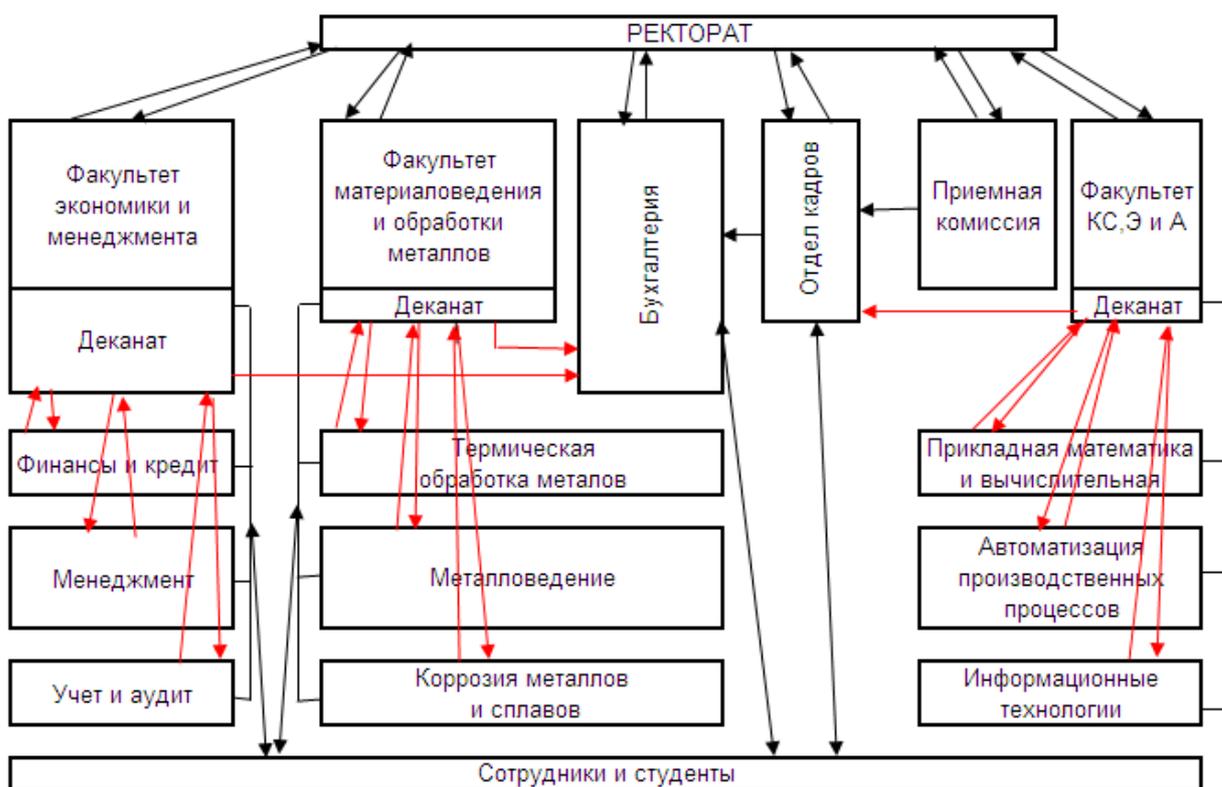


Рис.10. а – пример произвольной схемы действующих ИП, б – движение ИП в ВУЗе.

Далее выполняется анализ структуры документов и разрабатываются их унифицированные формы, пригодные для обработки в ЭВМ. Прежде чем

приступить к автоматизации информационных потоков, должны быть выполнены:

- системный анализ информационных потоков — уровень проблем, среда распространения, тип информации, взаимосвязь потоков;
- формирование целевых установок — политических, хозяйственных, нравственных, технических;
- определение переменных управления — экономических, организационных, правовых, информационных.

*Предметом труда* в АИС является информация, она может классифицироваться по различным признакам, например:

- по форме представления — аналоговая, дискретная;
- по степени детализации — детальная, интегрированная;
- по форме отображения — графическая, текстовая;
- по видам используемых параметров — техническая, социальная, биологическая, экономическая и т. д.

Потоки информации могут формироваться:

- как речевая передача информации;
- как передача информации в виде обычных документов с ручной доставкой;
- с использованием технических средств ручного ввода;
- с автоматическим вводом сообщений.

Поток информации рассматривается как совокупность двух понятий — схемы и элементов потока.

Схема потока информации задается указанием отношения вхождения относительно каждого элемента потока.

Элементами потока могут быть документы, элементы документов (показатели, реквизиты), операторы (люди, устройства, подразделения). Операторы могут быть источниками и потребителями.

В потоке информации (ПИ) определяются два основных параметра — *направление* и *плотность потока*. Направление потока задается местом его входа (наименование или шифр) подразделения. Плотность (значение) потока — объем информации (бит, количество документов, строк, знаков и т. д.), передаваемый в единицу времени (длительность передачи, приема или обработки).

Для отображения информационных потоков строят информационно-логические схемы взаимодействия их отдельных функций и частей.

Один из используемых методов - метод графов (рис.1.2). Элемент потока сопоставляют с вершинами графа, и каждую пару вершин соединяют дугой.

В математической теории графов и информатике граф — это совокупность непустого множества вершин и множества пар вершин. Объекты представляются как вершины, или узлы графа, а связи — как дуги, или рёбра. Вершины — некоторые сущности (объекты, события, процессы, явления), а дуги — отношения между ними.

Например, строение любой Web-страницы или сайта можно смоделировать при помощи ориентированного графа, в котором вершины — это статьи, а дуги (ориентированные рёбра) — гиперссылки (см. рисунок).

### ДАННЫЕ, ПЕРЕДАЮЩИЕСЯ В ИП

Система классификации и кодирования информации обеспечивает формализацию информации в виде, пригодном для последующей обработки ее на ЭВМ.

При обработке информации с помощью технических средств используют различные структуры: скалярные данные, массивы, блоки, записи и т. д.

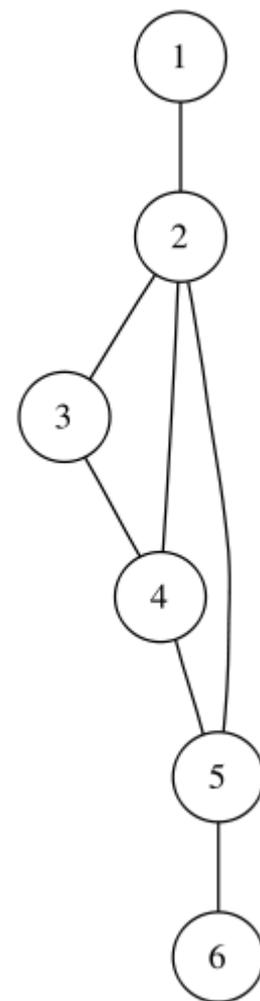
Данные — информация, представленная в формализованном виде, позволяющем передавать или обрабатывать ее с помощью ТС.

Массив — множество данных, содержащих достаточно полное описание информационной совокупности, состоящей из однотипных объектов. Смысловое содержание массива обычно указывается в его названии. Массив состоит из отдельных записей, имеющих одинаковое смысловое содержание. Каждая запись массива состоит из реквизитов. Длина реквизита — количество ячеек памяти, записанных последовательно.

Блок — несколько записей, объединенных в одну физическую запись.

На рис. 11 представлена классификация массивов на основе следующих классификационных признаков:

- семантического содержания, т.е. анализа ключевых понятий, составляющих рассматриваемую предметную область;
- технологии использования;
- носителя информации;
- технических характеристик.



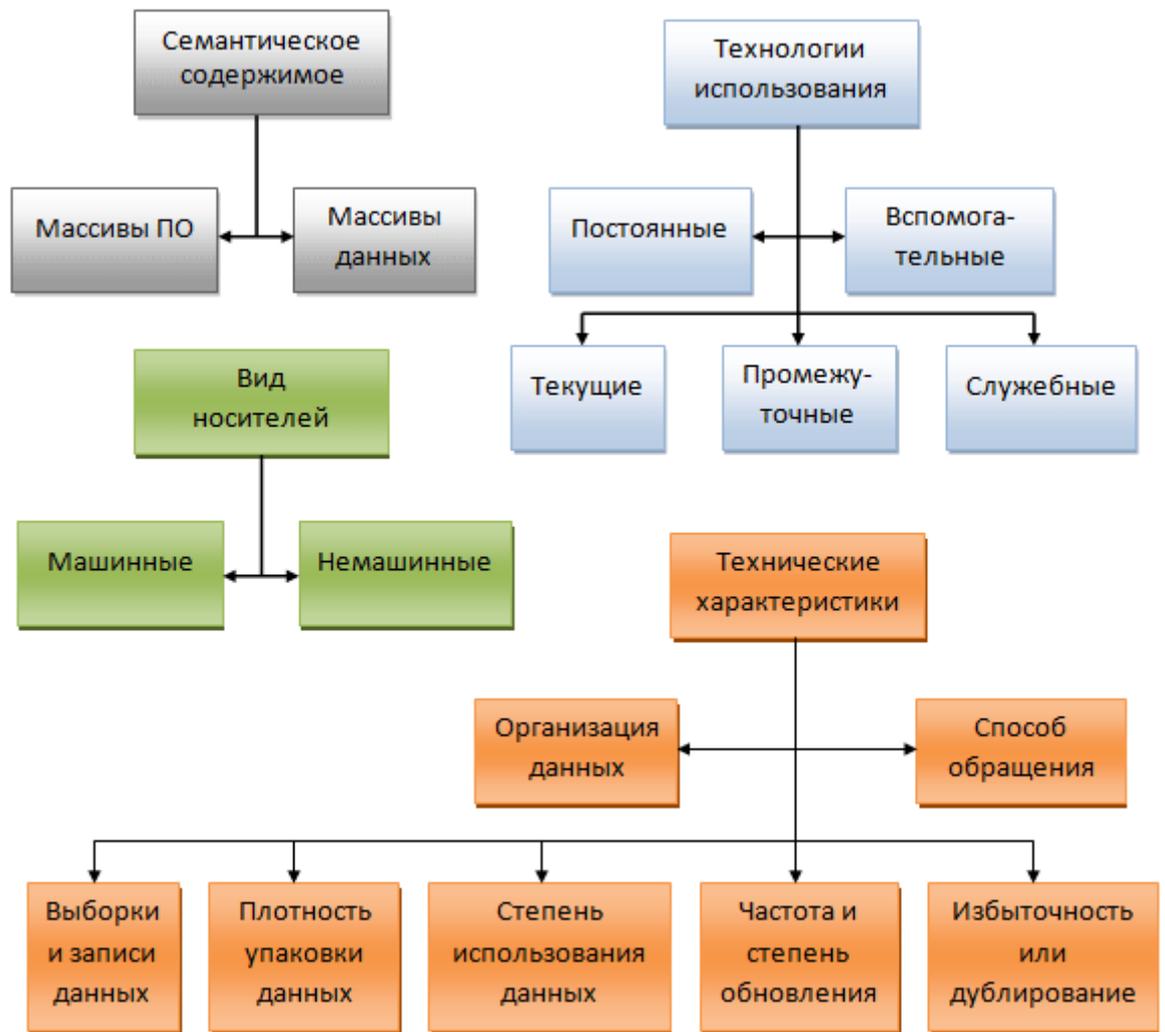


Рис. 11. Классификация массивов

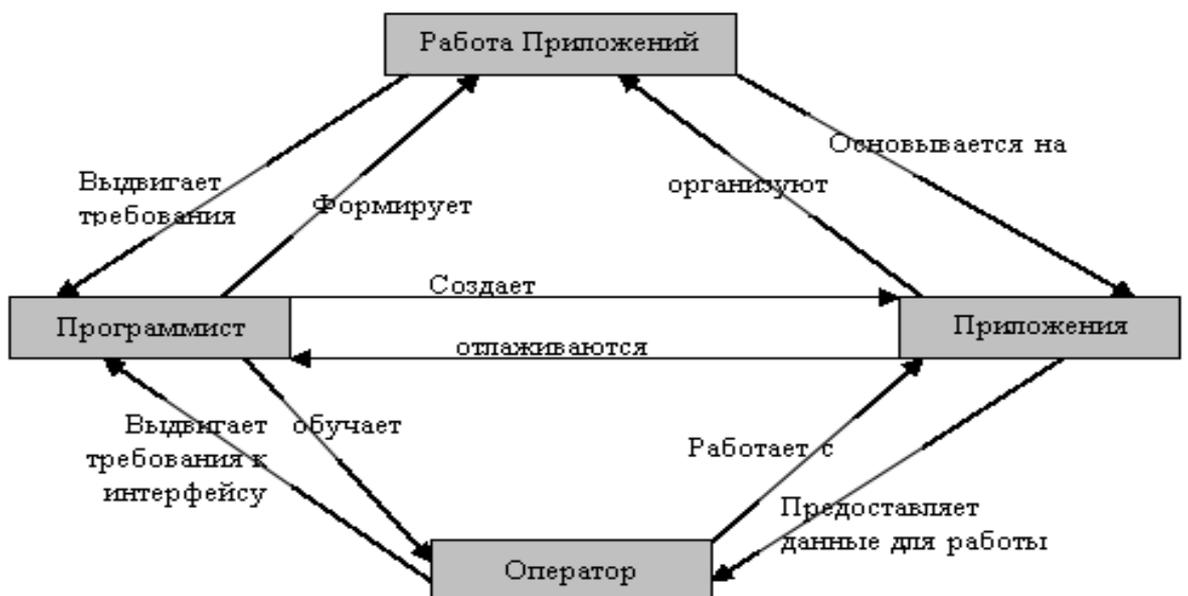


Рис.12 Семантическая сеть. Дуги данной сети обозначают соответствие.

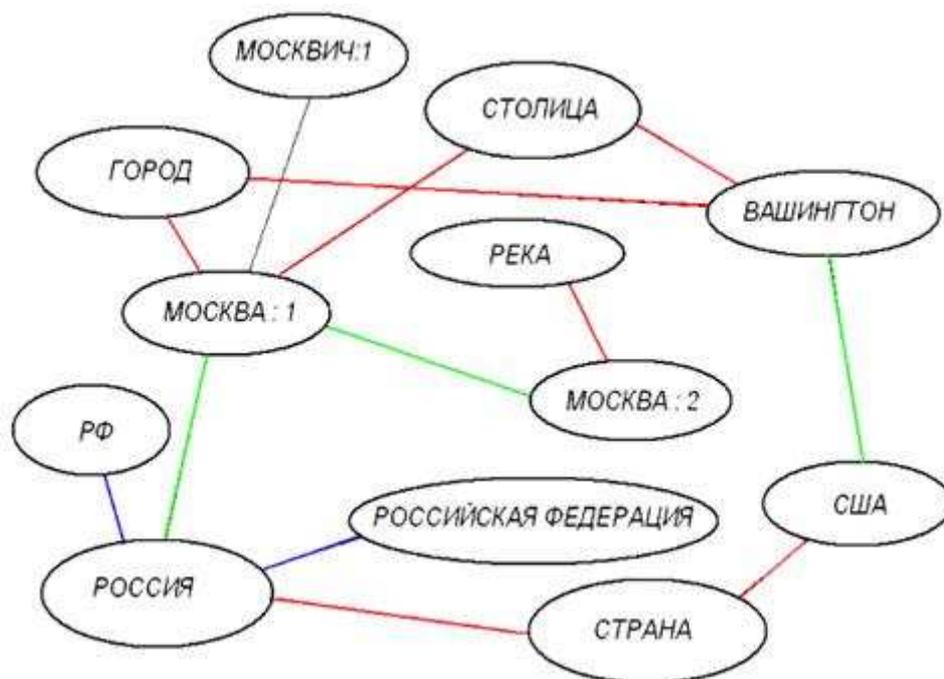


Рис. 13 Семантическая сеть

Семантика (от др.-греч. σηματικός — обозначающий) — раздел языкознания, изучающий значение единиц языка. В качестве инструмента изучения применяют семантический анализ.

Латентно-семантический анализ (ЛСА) — это метод обработки информации на естественном языке, анализирующий взаимосвязь между коллекцией документов и терминами в них встречающимися, сопоставляющий некоторые факторы (тематики) всем документам и термам.

Латентность- скрытость, отсутствие явного проявления.

В основе метода латентно-семантического анализа лежат принципы факторного анализа, в частности, выявление латентных связей изучаемых явлений или объектов. При классификации / кластеризации документов этот метод используется для извлечения контекстно-зависимых значений лексических единиц при помощи статистической обработки больших корпусов текстов[

В области информационного поиска данный подход называют латентно-семантическим индексированием (ЛСИ).

#### **Работа с массивами данных**

Организация массивов — упорядочение его записей и их физическое размещение в памяти системы. Как правило, массивы упорядочивают по ключу — наиболее важному признаку. Организация (структура) бывает:

- последовательно-смежная;
- цепная;
- ветвящаяся;
- списковая.

Под обработкой массива понимают процесс его преобразования. Типичные операции с массивами: сортировка, слияние, разбиение, поиск, изменение носителей. Сортировка — по ключу, методом замещения, отбора, вставки, квадратичной выборки. По времени различают доступ параллельный, ассоциативный, прямой, последовательный, с равным временем, индексно-последовательный, расчлененный.

Поиск — процедура выделения из некоторого множества объектов подмножества, содержащего только те объекты, которые удовлетворяют некоторому условию. Виды поиска: по совпадению, по интервалу близости, сложному арифметическому условию, семантическому условию, логической совокупности нескольких условий и т. д.

Методы поиска: перебор, вычисление адресов, дихотомический поиск.

Перебор — проверка условий поиска для всех объектов, входящих в состав данного множества.

Вычисление адресов — определение адреса в виде некоторой однозначной функции от заданного условия поиска, определение зоны памяти, ограниченной или произвольной по размеру, которая является - линейной функцией от кода, входящего в условие поиска.

Дихотомический поиск — после каждой проверки уменьшается область поиска примерно в 2 раза. Кроме проверки выполнения условий поиска каждый раз надо определить знак отклонения от заданного условия поиска.

Современные средства электронного документооборота дают возможность создать интегральные системы с одновременной обработкой структурных и неструктурных данных, включая письма, приказы, инструкции, расписания, видеoinформацию и т. д. Управление обычной электронной почтой предусматривает автоматизацию рассылки информации по адресам, т. е. с помощью ИС происходит управление всем документооборотом.

Цель автоматизации разнообразных потоков информации — совершенствование существующего документооборота, форм документов, сокращение их числа и копий, оптимизация маршрутов движения документов и алгоритма их формирования. Одна из основных задач состоит в разработке и внедрении средств и методов использования вычислительной техники для перевода документооборота из бумажной формы в электронную. Современные сетевые информационные технологии позволяют решить эту задачу. Все банки мира уже связаны сетевыми электронными сетями, и финансовые документы циркулируют в основном в электронном виде. Постепенно выходят из обращения бумажные акции предприятий и другие ценные бумаги. Их заменяют электронные депозитарии, т. е. базы данных

(БД), в которых сведения об акционерах хранятся в виде записей. Сравнительно недавно появились электронные деньги — это тоже записи БД. Движение электронных денег происходит по безбумажной технологии путем переноса данных из одних записей в другие. В качестве электронных денег служат пластиковые карты, содержащие сведения о владельце электронного счета на магнитной полосе, или смарт-карты, которые записывают на микросхему, встроенную в эту карту. По безбумажной технологии сегодня работает большинство средств массовой информации. Все этапы подготовки газет, рекламы, книг проводятся на компьютере с помощью автоматизированных издательских систем.

Многозадачные операционные системы (ОС) типа Windows позволяют одновременно создавать и редактировать тексты, а компьютерные сети объединяют в автономные рабочие группы. Такая рабочая группа может обходиться без бумажных документов до полного завершения работы над системой (проектом). Только когда работа закончена, можно составить итоговый бумажный документ. Основным препятствием на пути создания безбумажной технологии стоит проблема ввода исходных данных в электронном виде. Эта проблема решается путем создания и внедрения специальных аппаратов и специальных средств перевода информации разного вида в электронную форму. Автоматизация ввода информации в компьютер, т. е. перевод бумажных документов в электронный вид, осуществляется путем сканирования.

### **Формирование потоков данных в системе управления**

Информационные потоки в функционально-алгоритмической структуре представляют собой сеть обмена данными — «каркас» системы, внутри которого размещены алгоритмические модули. Данные, содержащиеся в информационных потоках, классифицируются по двум категориям:

- ❖ по степени детализации, определяемой уровнем иерархии управления;

- ❖ по семантике, определяемой результатами работы конкретных алгоритмических модулей, реализующих стандартный инвариантный к предметной области типовой набор функций управления.

Конкретное содержание информационных потоков определяется алгоритмами взаимодействия пользователей в рамках структуры управления. Эти алгоритмы должны быть разработаны на основании типовых технологических операций, типовых функций управления и их типового документального оформления. На этой основе разрабатываются индивидуальные «маршруты» деятельности, соответствующие конкретной специфике предприятия и предметной области и соответствующий им документооборот.

**Разработка технического задания на аппаратно-программный комплекс системы.** В состав этого задания должно быть включено следующее:

- ❖ перечень функций аппаратно-программного комплекса по обеспечению приема, хранения, обработки, обмена и выдачи данных и требования к выполнению этих функций;
- ❖ требования к архитектуре аппаратно-программного комплекса;
- ❖ требования к программной, аппаратной и сетевой платформе;
- ❖ требования к информационному обеспечению разработки (составу проектной и эксплуатационной документации);
- ❖ требования к администрированию и поддержке системы в процессе эксплуатации;
- ❖ требования к организационному обеспечению разработки и внедрения (стадии, этапы, сетевой график);
- ❖ бизнес-план (технико-экономическое обоснование разработки применительно ко всем стадиям жизненного цикла системы).

Следует отметить, что при разработке и внедрении аппаратно-программного комплекса необходимо учитывать, что у крупных заказчиков, как правило, уже есть информационная система, нуждающаяся в модернизации и реконструкции. Поэтому при внедрении новой системы и сопутствующему ей частичному изменению организационной структуры предприятия необходимо разработать и поэтапно осуществить плавный алгоритм перехода от старой системы к новой без ущерба для функционирования предприятия.

Основные работы, проводимые в рамках разработки проекта реорганизации, выделены в следующие этапы:

1. Функциональное и информационное обследование предприятия. Как указывалось выше, цель этого этапа состоит в сборе подробной информации о направлениях работы, решаемых задачах, приоритетах, показателях развития, показателях эффективности, организационной структуре, выполняемых функциях, сопутствующих им материальных, финансовых и информационных потоках, взаимодействии с внешними абонентами, используемых средствах автоматизации и др.

2. Определение миссии предприятия, иерархии целей, разработка функциональных моделей существующей (“как есть”) организации бизнес-процессов. Цель этого этапа состоит в формальном представлении собранной информации в соответствии с согласованной целью проекта.

3. Анализ организации бизнеса. Оценка эффективности существующей организации бизнес-процессов. На этом этапе выделяются показатели эффективности выполнения бизнес-процессов, выполняется анализ существующего уровня и методов их информационной поддержки, функционально-стоимостной анализ, выделение центров затрат, динамический анализ загрузки и распределения ресурсов и др.

4. Разработка функциональных и информационных моделей рациональной (“как должно быть”) организации бизнес-процессов, в соответствии с установленной иерархией целей предприятия, на основе согласованных принципов и возможностей информационных технологий. На этом этапе проводится также формирование предложений по

совершенствованию системы управления предприятия и системы информационной поддержки бизнес-процессов.

5. Оценка эффективности предлагаемых решений. На этом этапе проводится расчет значений, выделенных ранее показателей эффективности бизнес-процессов, соответствующих моделям “как должно быть”, их сопоставление с значениями, полученными для моделей “как есть”, анализ различий.

6. Разработка поэтапного плана реорганизации и проекта корпоративной информационной системы. На этом выполняется также анализ окупаемости инвестиций.

## **ОБМЕН ИНФОРМАЦИОННЫМИ ПОТОКАМИ В ВУЗе**

Рассмотрим разработку программного обеспечения документооборота структурного подразделения высшего учебного заведения, которое включает систему формирования текущих запросов и отчетности

Одним из важнейших условий успешного функционирования любого вуза является эффективный обмен информационными потоками между его подразделениями, т.е. наличие системы качественного документооборота и автоматизированных систем сбора и обработки данных.

Отчеты о проведенной работе конкретных преподавателей и структурных подразделений, как правило, выполняются в виде бумажных документов - журнала "Индивидуальный план работы преподавателя и ее учет" и вкладыша к нему. Также к отчетной документации относятся рабочие журналы, отражающие ход выполнения учебного процесса, зачетные и экзаменационные ведомости. Надо отметить тот факт, что вкладыш в журнал представляет собой несколько табличных форм, требующих ежедневного заполнения и ежемесячного расчета сводных ведомостей данных, а данные в журнал "Индивидуальный план работы преподавателя и ее учет" вносятся конце полугодия и учебного года. Конечно, что эти документы находятся в течение всего периода учебного процесса на руках преподавателей, что значительно препятствует оперативному анализу и контролю объема и содержания работы, проводимой как для конкретного преподавателя, так и для структурного подразделения (кафедры) в целом. Практически оперативный анализ и контроль для любой текущей даты при бумажном ведении отчетной документации в структурному подразделению невозможен. Необходимо отметить тот факт, что индивидуальные журналы, отражающие ход выполнения учебного процесса, ведутся каждым преподавателем самостоятельно для части структурной единицы (подгруппы студентов), что не позволяет проводить в целом оперативный анализ качества усвоения дисциплины.

Одной из основных проблем, которую необходимо решить при реализации данной системы, является проблема множественных отношений. Она заключается в том, что несколько преподавателей ведут дисциплины, как на одном потоке, так и в одной группе, например, лабораторные работы.

Следовательно, каждый преподаватель должен иметь общий доступ к журналам групп. Решить данную задачу возможно путем структуризации потоков и студенческих групп относительно дисциплины, которая читается. На первый взгляд, образуется некоторый избыток данных. Однако, если глубже рассмотреть этот вопрос, то становится очевидным тот факт, что в реальных условиях рабочие планы и программы дисциплин, как правило, для каждой специальности или потока разные. Таким образом, исполнительная структура относительно дисциплины может быть представлена в виде следующего соотношения категорий :

**ДИСЦИПЛИНА: ПОТЕК (ЛЕКТОР) : ГРУППА (АССИСТЕНТ).**

Здесь имеется в виду следующее: категория **дисциплина** являет собой список дисциплин, которые читаются во всех группах и на всех потоках. Категория **поток** являет собой список потоков (специальностей), которые обеспечивает подкатегория **лектор**, которая в свою очередь относится к категории **преподавателей**.

Категория **группа**, в свою очередь, являет собой список групп, которые обеспечивает подкатегория **ассистент**, которая в свою очередь входит в категорию **преподаватель**.

Относительно вложенных документов структура будет иметь такой вид:

**ДИСЦИПЛИНА: ПОТОК (учебно-методические материалы) : ГРУППА (журналы, зачетно-экзаменационные сведения).**

Каждый преподаватель, который входит в структуру обеспечения потока должен иметь доступ как к учебно-методическим материалам, так и к журналам и сведениям конкретных групп. При этом, соответствующие записи в один журнал могут осуществляться двумя или больше преподавателями.

Возникает проблема, связанная с некоторыми противоречиями и осложнениями в реализации прав доступа только к той информации, которая касается конкретного преподавателя. Однако здесь следует иметь в виду тот факт, что система должна обеспечить функционирование в реальных условиях, то есть доступ к рабочим материалам студенческих групп могут иметь и другие преподаватели, что предопределенно жизненной необходимостью, например, болезнью преподавателя. Таким образом, разграничение прав доступа со стороны преподавателей не должно быть очень жестким.

Известно, что отношения многие-ко-многим не могут быть реализованы в реляционной модели с жесткими связями, потому что любой набор данных DataSet имеет лишь один уровень соединения DataSource. Таким образом, реализация системы в традиционных отношениях Master : Detail является затрудненным. Также следует заметить тот факт, что сортировка полей в Detail -представлении невозможна, что приводит к необходимости использования дополнительных представлений вида LookUpTable или ClientDataSet. Решить эту проблему возможно путем структуризации данных SQL -запросом или применением переопределяемых

динамических фильтров. Предпочтение желательно оказать второму способу, потому что он обеспечивает более высокое быстродействие в рамках малораспределенных локальных сетей.

Общую структуру системы, которая решает поставленные задачи, можно представить в виде, который приведен на рисунке 14 :

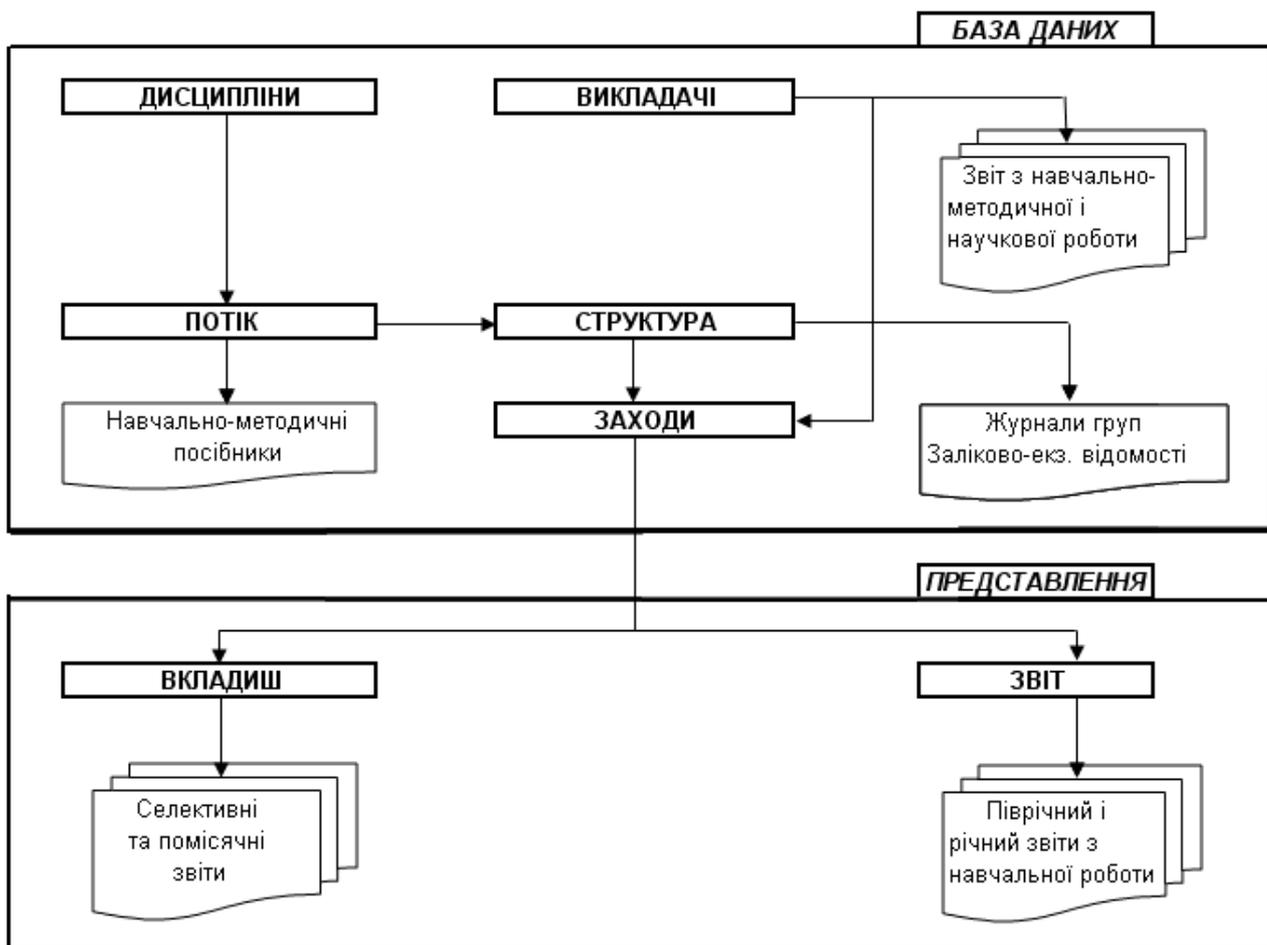


Рис.14. Структурная схема системы

В приведенной схеме четко распределяются принадлежность объектов описания к базе данных и представлений. Как в базе данных, так и в представлениях связь между объектами имеет условный характер. Все связи определяются динамическими фильтрами и реализуются непосредственно в клиентских приложениях. В обычных полях базы данных описываются характеристики объектов, а в бинарных полях базы данных хранятся документы. Объекты - категории ДИСЦИПЛИНЫ, ПРЕПОДАВАТЕЛИ и ПОТОК - представляют собой условно справочные таблицы, которые заполняются преподавателями в процессе регистрации и в процессе работы по мере необходимости. Объект **структура** описывается преподавателем - лектором. Данный объект описывает структуру как *отдельные группы студентов* относительно дисциплин. Объект **мероприятия** представляет собой результирующую таблицу, в которую заносятся проведенные каждым

преподавателем мероприятия. На основании объекта МЕРОПРИЯТИЯ по требованию пользователя создаются *представления* - отчеты о проведенных на текущий период мероприятиях для каждого преподавателя. Представления не хранятся в базе данных. Методика построения проекта описана в методическом пособии к лабораторным работам по данной дисциплине.

### **Обоснование выбора технологии и среды разработки**

Одной из наилучших сред разработки СУБД есть среда Delphi, которая позволяет при минимальных расходах создавать мощные дополнения за счет большего количества предопределенных классов и легкой интеграции новых классов. В рассматриваемом проекте используется компонент хранения ActiveX документов в бинарных полях баз данных, совместимый с форматом mdb Access, а также тот факт, что в среде Delphi представлен широкий набор компонентов, которые поддерживают технологию ADO, разработка системы выполнена в указанной среде. Также, что технология ADO позволяет создавать клиентские приложения баз данных без установления дополнительного серверного программного обеспечения за счет интеграции в операционной системе интерфейсов провайдеров доступа к данным. Таким образом, разработанные в среде Delphi приложения СУБД представляют собой по существу один исполняемый файл, который не требует установления и регистрации дополнительных компонентов приложения.

### **Хранилища данных**

*Электронный архив (ЭА), или Хранилище данных (ХД), или Data Warehouse (DW) - это база данных, хранящая данные, агрегированные по многим измерениям. Пополнение ХД происходит на периодической основе. При этом автоматически формируются новые агрегаты данных, зависящие от старых. Доступ к ХД организован особым образом на основе модели многомерного куба.*

*Хранилище данных - это не автоматизированная система принятия решений, не экспертная система, не система логического вывода, а "всего лишь" оптимально организованная база данных, обеспечивающая максимально быстрый и комфортный доступ к информации, необходимой при принятии решений.*

DW есть предметно ориентированный, интегрированный, неизменный, поддерживающий хронологию набор данных, предназначенный для поддержки принятия решений. В этом определении соединены две различные функции:

а) сбор, организация и подготовка данных для анализа в виде постоянно наращиваемой базы данных;

б) собственно анализ как элемент принятия решений. Принятие решений в качестве сферы применения DW существенно сужает определение.

Если в определении оставить лишь анализ (как элемент научных,

технологических и экологических систем), круг использования данной концепции может быть значительно расширен.

Очень важен основной принцип действия DW: единожды занесенные в DW данные затем многократно извлекаются из него и используются для анализа. Отсюда вытекает одно из основных преимуществ использования DW в работе предприятия -- контроль за критически важной информацией, полученной из различных источников, как за производственным ресурсом.

Наиболее уязвимым местом использования DW на предприятии, с точки зрения бизнеса, является корректность его данных, полученных из разных источников. Данные перед загрузкой в DW должны быть либо "очищены от шума", либо обработаны методами нечеткой логики, допускающей наличие противоречивых фактов. Например, данные о предприятии-партнере могут быть получены от разных экспертов, чьи оценки порой бывают диаметрально противоположными.

### **Концепция хранилища данных.**

Каждый, кто хоть раз в жизни проектировал базу данных, знает, насколько это сложный, чреватый ошибками и дорогостоящий процесс. Действительно, построить нормализованную БД реального предприятия, содержащую сотни или даже тысячи таблиц, - чрезвычайно сложная задача даже при наличии хороших CASE-средств.

К проектированию хранилищ данных обычно предъявляются следующие требования:

- ❖ структура данных хранилища должна быть понятна пользователям;
- ❖ полнота и достоверность хранимых данных;
- ❖ должны быть выделены статические данные, которые регулярно модифицируются: ежедневно, еженедельно, ежеквартально;
- ❖ поддержка внутренней непротиворечивости данных;
- ❖ должны быть упрощены требования к запросам с целью исключения запросов, которые могли бы требовать множественных утверждений SQL в традиционных реляционных СУБД;
- ❖ поддержка высокой скорости получения данных из хранилища;
- ❖ должна быть обеспечена поддержка сложных запросов SQL, которые требуют последовательной обработки тысяч или миллионов записей;
- ❖ возможность получения и сравнения так называемых срезов данных (slice and dice);
- ❖ наличие удобных утилит просмотра данных в хранилище;
- ❖ поддержка качественного процесса пополнения данных.

Эти требования существенно отличают структуру реляционных СУБД и хранилищ данных. Нормализация данных в реляционных СУБД приводит к созданию множества связанных между собой таблиц. В результате, выполнение сложных запросов неизбежно приводит к объединению многих таблиц, что существенно увеличивает время отклика. Проектирование хранилища данных подразумевает создание денормализованной структуры данных (допускается избыточность данных и возможность возникновения аномалий при манипулировании данными), ориентированной в первую

очередь на высокую производительность при выполнении аналитических запросов. Нормализация делает модель хранилища слишком сложной, затрудняет ее понимание и ухудшает эффективность выполнения запроса.

В размерном моделировании принят стандарт модели, называемый **схемой звезда (star schema)**, которая обеспечивает высокую скорость выполнения запроса посредством денормализации и разделения данных. Невозможно создать универсальную денормализованную структуру данных, обеспечивающую высокую производительность при выполнении любого аналитического запроса. Поэтому схема звезда строится так, чтобы обеспечить наивысшую производительность при выполнении одного самого важного запроса, либо для группы похожих запросов.

Схема звезда обычно содержит одну большую таблицу, называемую таблицей фактов (*fact table*), помещенную в центр, и окружающие ее меньшие таблицы, называемые таблицами размерности (*dimensional table*), соединенные с таблицей фактов в виде звезды радиальными связями. В этих связях таблицы размерности являются родительскими, таблица фактов - дочерней. Схема звезда может иметь также консольные таблицы (*outrigger table*), присоединенные к таблице размерности. Консольные таблицы являются родительскими, таблицы размерности - дочерними.

Прежде чем создать базу данных со схемой типа звезда, необходимо проанализировать бизнес-правила предметной области с целью выяснения центрального вопроса, ответ на который наиболее важен. Все прочие вопросы должны быть объединены вокруг этого основного вопроса и моделирование должно начинаться с него. Данные, необходимые для ответа на этот вопрос, должны быть помещены в центральную таблицу модели - таблицу фактов. На рис. 4.1 показан пример схемы «звезда».



Рис. 15. Пример схемы «звезда»

Описанию (star schema) и рекомендациям по ее применению полностью посвящены работы. Идея схемы звезды заключается в том, что имеются таблицы для каждого измерения, а все факты помещаются в одну таблицу, индексируемую множественным ключом, составленным из ключей

отдельных измерений (рис.15). Каждый луч схемы звезды задает, в терминологии Кодда, направление консолидации данных по соответствующему измерению.

**Размерности.** В технологии многомерного моделирования размерность — это аспект, в разрезе которого можно получать, фильтровать, группировать и отображать информацию о фактах. Типичные размерности, встречающиеся практически в любой модели:

- ❖ клиент;
- ❖ продукт;
- ❖ время;
- ❖ география;
- ❖ сотрудник.

Размерности, как правило, имеют многоуровневую иерархическую структуру. Например, размерность ВРЕМЯ может иметь следующую структуру: ГОД, КВАРТАЛ, МЕСЯЦ, ДЕНЬ.

**Таблица фактов.** Таблица фактов является центральной таблицей в схеме звезда. Она может состоять из миллионов строк и содержать суммирующие или фактические данные, которые могут помочь ответить на требуемые вопросы. Она соединяет данные, которые хранились бы во многих таблицах традиционных реляционных базах данных. Таблица фактов и таблицы размерности связаны идентифицирующими связями, при этом первичные ключи таблицы размерности мигрируют в таблицу фактов в качестве внешних ключей. В размерной модели направления связей явно не показываются – они определяются типом таблиц. Первичный ключ таблицы фактов целиком состоит из первичных ключей всех таблиц размерности. В примере рис. 4.1 первичный ключ составлен из четырех внешних ключей: «Поставщик», «Покупатель», «Продукт» и «Период».

Таблицы размерности имеют меньшее количество строк, чем таблицы фактов и содержат описательную информацию. Эти таблицы позволяют пользователю быстро переходить от таблицы фактов к дополнительной информации.

В примере на рис.15 “Таблица\_покупателей”, “Таблица\_периодов”, “Таблица\_продуктов» и “Таблица\_поставщиков” - таблицы размерности, которые позволяют быстро извлекать информацию о том, кто и когда сделал покупку, какой продавец и на какую сумму продал и какие именно товары были проданы. 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).

Факты имеют ряд свойств, на которых мы вкратце остановимся. Более подробное описание фактов и их свойств можно найти в литературе [337-339].

**Аддитивные факты.** Аддитивность определяет возможность суммирования факта вдоль определенной размерности. Аддитивные факты можно суммировать и группировать вдоль всех размерностей на любых уровнях иерархии.

**Полуаддитивные факты.** Полуаддитивный факт — это факт, который можно суммировать вдоль определенных размерностей, и нельзя — вдоль других. Примером может служить остаток на счете (или остаток товара на складе). Данную величину нельзя суммировать вдоль размерности ВРЕМЯ. Однако сумма остатков по счетам вдоль размерности смысл для анализа.

**Неаддитивные факты.** Неаддитивные факты вообще нельзя суммировать. Пример неаддитивного факта — отношение (например, выраженное в процентах). Специалисты рекомендуют моделировать полуаддитивные факты таким образом, чтобы сделать их более аддитивными. Например, представить процент составляющими его величинами.

**Таблицы покрытия.** Таблицы покрытия используются с целью моделирования сочетания размерностей, для которых отсутствуют факты. Например, нужно найти количество категорий продуктов, которые сегодня ни разу не продавались. Таблица фактов продаж не может ответить на данный вопрос, поскольку она регистрирует только факты продаж. Для того чтобы модель позволяла отвечать на подобные вопросы, нужна дополнительная таблица фактов (которая, по сути дела, не содержит фактов), которая и называется таблицей покрытия.

Итак, формально корпоративное хранилище данных можно определить как комплекс аппаратно-программных средств и технологий создания архива (масштаба отрасли, или корпорации, или предприятия) документов в электронном виде. Цель создания ХД состоит в обеспечении оперативного и полноценного доступа ко всем хранящимся и поступающим документам. Для этого требуется решить две основные задачи: ввести массив имеющихся в архиве документов и обеспечить возможность оперативного полнотекстового доступа к электронным документам.

Очень важен основной принцип действия ХД: единожды занесенные в

ХД данные затем многократно извлекаются из него и используются для анализа. Отсюда вытекает одно из основных преимуществ использования ХД в работе предприятия -- контроль за критически важной информацией, полученной из различных источников, как за производственным ресурсом.

Наиболее уязвимым местом использования ХД на предприятии является корректность его данных, полученных из разных источников. Данные перед загрузкой в ХД должны быть либо "очищены от шума", либо обработаны методами нечеткой логики, допускающей наличие противоречивых фактов. Например, данные о предприятии-партнере могут быть получены от разных экспертов, чьи оценки порой бывают диаметрально противоположными.

### **Терминология.**

В области ХД приняты основные термины:

*Хранилище данных.* Это понятие является, наверное, наиболее широко трактуемым. Приведем ставшее уже классическим определение Б. Инмона: Хранилище данных (Data Warehouse) это - "*предметно-ориентированный, интегрированный, неизменяемый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений*". Достаточно часто под ХД подразумевается не только набор данных, но также и вся технология использования ХД. В данном учебнике под ХД мы будем понимать только набор данных, причем не единственный, которой используется в рамках этой технологии.

*Витрина данных.* Это понятие возникло несколько позже термина ХД, поэтому в некоторых источниках оно слито с понятием ХД. В данном учебнике под витриной (или Киоском) данных (Data Mart) мы будем понимать сравнительно небольшое ХД, сконструированное для использования неким подразделением с одним существенным отличием от ХД - в витрине данных конечный пользователь может создавать свои собственные структуры данных. Есть еще одна особенность у витрин данных (ВД) - источником для большинства хранящихся там данных является ХД. Это приводит к тому, что при создании ВД редко используются инструменты по очистке, денормализации и унификации данных.

*Технология ХД.* Под этим термином будем понимать технологию использования всех объектов связанных с ХД, как то:

- ❖ хранилища данных;
- ❖ витрины данных;
- ❖ программное обеспечение.

*Система Поддержки Принятия Решений.* Термин "Система Поддержки Принятия Решений (DSS, СППР)" начал использоваться раньше возникновения концепции ХД, но до сих пор имеет множество трактовок. Ряд авторов используют его для названия всей системы в целом, включая источники данных, ХД, и средства представления и анализа данных.

В данном учебнике этот термин используется для обобщающего названия систем, специализированных на представлении и анализе специализированных данных, например, финансовых. При использовании технологии ХД эти системы не решают вопросы доставки, очистки

интеграции данных. Кроме этого, так как СППР обычно решают вопросы некой конкретной области применения, они не всегда могут быть использованы в применении к специфической организации. В нашей же стране использование стандартных СППР очень часто бывает проблематичной.

*Информационная Система Руководителя.* Информационная Система (ИСП), как нам кажется, является не очень удачным переводом термина Executive Information System (EIS). дело в том, что такие системы обычно являются средством создания приложений без программирования, и поэтому используются не столько руководителями, сколько аналитиками, которые обычно, используя это средство, создают приложения, которыми уже пользуются руководители.

*Средства OLAP.* В большинстве случаев под этим термином понимают удобную и красивую оболочку для навигации по многомерным данным.

*Операционные БД (ОБД).* Этот термин обозначает наши старые, добрые БД и введен для того, что бы подчеркнуть их существенное отличие от БД, используемых для реализации ХД.

*Средства Анализа.* В данном учебнике этот термин обозначает весь спектр приложений для конечного пользователя, включая:

- ❖ ИС
- ❖ СППР
- ❖ Средства OLAP
- ❖ другие специализированные средства анализа, прогноза и представления данных.

*Информационная Система Нового Поколения.* В данном учебнике этот термин (ИСНП) вводится для обозначения всей системы, построенной по Технологии ХД, включая Источники данных, Хранилище данных и Средства Анализа.

### **Основные компоненты информационного хранилища**

*ПО промежуточного слоя.* Обеспечивает сетевой доступ и доступ к базам данных. Сюда относятся сетевые и коммуникационные протоколы, драйверы, системы обмена сообщениями.

*транзакционные БД и внешние источники информации.* Базы данных OLTP-систем исторически предназначались для эффективной обработки структур данных в относительно небольшом числе четко определенных транзакций. Из-за ограниченной целевой направленности "учетных" систем применяемые в них структуры данных плохо подходят для систем поддержки принятия решений. Кроме того, возраст многих установленных OLTP-систем достигает 10 - 15 лет. В настоящее время источниками данных хранилища служат оперативные транзакционные системы, которые обслуживают повседневную учетную деятельность компании. Необходимость включения той или иной транзакционной системы в качестве источника определяется бизнес-требованиями к СППР. Исходя из этих же требований, в качестве источников данных, могут быть рассмотрены внешние системы, в том числе

и Интернет. Детальные данные из источников могут либо напрямую поступать в хранилище, либо предварительно агрегироваться до требуемого уровня обобщения;

*уровень доступа к данным.* Относящееся сюда ПО обеспечивает общение конечных пользователей с информационным хранилищем и загрузку требуемых данных из транзакционных систем. В настоящее время универсальным языком общения служит язык структурированных запросов (SQL);

*загрузка и предварительная обработка.* Этот уровень включает в себя набор средств для загрузки данных из OLTP-систем и внешних источников. Выполняется, как правило, в сочетании с дополнительной обработкой: проверкой данных на чистоту, консолидацией, форматированием, фильтрацией и пр. Данная компонента представляет собой ПО, которое в соответствии с определенным регламентом извлекает данные из источников и приводит их к единому формату, определенному для хранилища и отвечает за формализованную логическую согласованность, качество и интеграцию данных, которые загружаются из источников в оперативный склад данных. Каждый источник данных требует разработки собственного загрузочного модуля. Каждый модуль должен решать два класса задач:

- ❖ начальной загрузки ретроспективных данных,
- ❖ регламентного пополнения хранилища данными из источников.

Данная компонента также извлекает детальные данные по регламенту из оперативного склада, производит их агрегирование, консолидацию, трансформацию и помещает данные в хранилище и витрины данных. Именно в данной подсистеме должны быть определены все бизнес-модели консолидации данных по иерархическим измерениям и вычисления зависимых бизнес-показателей по независимым исходным данным;

*информационное хранилище.* Представляет собой предметно-ориентированную базу или совокупность БД, извлекаемых из источников, которые организованы по сегментам, отражающим конкретную предметную область бизнеса: производство, правило, детальные слабо агрегированные данные;

*метаданные.* Метаданные (репозиторий, "данные о данных"). Играют роль справочника, содержащего сведения об источниках первичных данных, алгоритмах обработки, которым исходные данные были подвергнуты, и т. д. Метаданные - это любые данные о данных. Метаданные играют важную роль в построении Систем Поддержки Принятия Решений (СППР). Одновременно это один из наиболее сложных и недостаточно практически проработанных объектов. В общем случае можно выделить по крайней мере три аспекта метаданных, которые должны присутствовать в системе.

1. С точки зрения пользователей:
  - ❖ метаданные для бизнес-аналитиков,
  - ❖ метаданные для администраторов,
  - ❖ метаданные для разработчиков.
2. С точки зрения предметных областей:

- ❖ структуры данных хранилища,
  - ❖ модели бизнес-процессов,
  - ❖ описания пользователей,
  - ❖ технологические и пр.
3. С точки зрения функциональности системы:
- ❖ метаданные о процессах трансформации,
  - ❖ метаданные по администрированию системы,
  - ❖ метаданные о приложениях,
  - ❖ метаданные о представлении данных пользователям.

Присутствие трех перечисленных аспектов метаданных подразумевает, что, например, прикладные пользователи и разработчики системы будут иметь различное видение технологических аспектов трансформации данных из источников: прикладные пользователи - семантику, состав и периодичность пополнения хранилища данными из источника, разработчики - ER-диаграммы, правила трансформации и интерфейс доступа к данным источника;

*уровень информационного доступа.* Обеспечивает непосредственное общение пользователя с данным DW посредством стандартных систем манипулирования, анализа и предоставления данных типа MS Excel, MS Access, Lotus 1-2-3 и др.

*уровень обработки запросов и представления данных.* Оперативный склад, хранилище витрины данных являются инфраструктурой, которая обеспечивает хранение и администрирование данных. Для извлечения данных, их аналитической обработки и представления конечным пользователям служит специальное ПО. Как правило, можно выделить три типа данного ПО.

Программное обеспечение регламентированной отчетности, которое характеризуется заранее определенными запросами данных и их представлениями бизнес-пользователям. От данного ПО не требуется быстрого времени реакции. Из соображений стоимости эффективности для его реализации в наибольшей степени подходит технология ROLAP (см. далее).

Программное обеспечение нерегламентированных запросов пользователей. Это ПО – основной способ общения бизнес-аналитиков с хранилищем, при котором каждый последующий запрос к данным и вид их представления определяются, как правило, результатами предыдущего запроса. Для приложений данного типа требуется высокая скорость обработки запросов (единицы секунд). Данное ПО реализуется технологией MOLAP (см. далее) и специальными инструментами построения сложных нерегламентированных запросов с интуитивно понятным для бизнес-аналитиков графическим интерфейсом.

Программное обеспечение добычи знаний, которое реализует сложные статистические алгоритмы и алгоритмы искусственного интеллекта, предназначенные для поиска скрытых в данных закономерностей, представления этих закономерностей, представления этих закономерностей в

виде моделей и многовариантного прогнозирования по ним развития ситуаций по схеме «Что если ...?».

Конечно, как правило, такое деление носит весьма условный характер, а границы между соответствующими приложениями могут быть размыты.[342].

*уровень управления (администрирования).* Отслеживает выполнение процедур, необходимых для обновления информационного хранилища или поддержания его состояния. Здесь программируются процедуры подкачки данных, перестройки индексов, выполнения итоговых (суммирующих) расчетов, репликации данных, построения отчетов, формирования сообщений пользователям, контроля целостности и др. К ведению данной подсистемы относятся все задачи, связанные с поддержанием системы и обеспечением ее устойчивой работы и расширения. Можно выделить, по крайней мере, четыре класса задач, расширение которых должна обеспечивать данная подсистема:

1) Администрирование данных, которое включает в себя регулярное пополнение данных из источников, если необходимо, ручной ввод, сверка и корректировка данных в оперативном складе. Администрирование данных ведется, как правило, бизнес-пользователями, а ответственность распределяется по предметно-ориентированным сегментам.

2) Администрирование хранилища данных. В задачу администрирования хранилища входят все вопросы, связанные с поддержанием архитектуры хранилища, обеспечением его эффективной и бесперебойной работы, защитой и восстановлением данных после сбоев.

3) Администрирование доступа к данным обеспечивает сопровождение профилей пользователей, разграничение доступа к конфиденциальным данным, защиту информации от несанкционированного доступа.

**Ориентированность на предметную область.** Первая особенность хранилища данных заключается в его ориентированности на предметный аспект. Предметная направленность контрастирует с классической ориентированностью прикладных приложений на функциональность и процессы. Приложения всегда оперируют функциями, такими, например, как открытие сделки, кредитование, выписка накладной, зачисление на счет и т.д. Хранилище данных организовано вокруг фактов и предметов, таких, как сделка, сумма кредита, покупатель, поставщик, продукт и т.д.

**Независимость:** выделенность информационного хранилища существенно снижает нагрузку на OLTP-системы со стороны аналитических приложений, тем самым производительность существующих систем не ухудшается, а на практике происходит уменьшение времени отклика и улучшение доступности систем.

**Интегрированность.** Наиболее важный аспект хранилища данных состоит в том, что данные, находящиеся в хранилище, интегрированы. Интегрированность проявляется во многих аспектах:

- ❖ в согласованности имен,
- ❖ в согласованности единиц измерения переменных,

- ❖ в согласованности структур данных,
- ❖ в согласованности физических атрибутов данных и др.

Контраст между интеграцией данных в хранилище данных и в прикладном окружении иллюстрируется следующим образом.

Первая причина возможного рассогласования приложений заключается в наличии множества средств разработки. Каждое средство разработки диктует определенные правила, часть из которых индивидуальна для данного средства. Не секрет, что каждый разработчик предпочитает одни средства разработки другим. Если два разработчика используют различные средства разработки, они, как правило, применяют индивидуальные особенности средств, а значит, возникает вероятность несогласованности между создаваемыми системами.

Вторая причина возможного рассогласования приложений заключается в существовании множества способов построения приложения. Способ построения конкретного приложения зависит от стиля разработчика, от времени, когда это приложение разрабатывалось, а также от ряда факторов, характеризующих конкретные условия разработки приложения. Все это отражается на используемых способах задания ключевых структур, способах кодирования, обозначения данных, физических характеристиках данных и т.д. Таким образом, если два разработчика создают различные способы построения приложений, имеется высокая вероятность того, что полной согласованности между системами не будет.

Интеграция данных по единицам измерения атрибутов состоит в следующем. Разработчики приложений к вопросу о способе задания размеров продукции могут подходить несколькими путями. Размеры могут задаваться в сантиметрах, дюймах, ядрах и т.д. Каков бы ни был источник данных, если информация поступит в хранилище, она должна быть приведена к одним и тем же единицам измерения, принятым в качестве стандарта в хранилище.

**Историчность и стабильность:** OLTP-системы оперируют с актуальными данными, срок применения и хранения которых обычно не превышает величины текущего бизнес-периода (полугода-год), в то время как информационное хранилище данных нацелено на долговременное хранение информации в течении 10-15 лет. Стабильность означает, что фактическая информация в хранилище данных не обновляется и не удаляется, а только специальным образом адаптируется к изменениям бизнес-атрибутов. Таким образом, появляется возможность осуществлять исторический анализ информации.

**Зависимость от времени.** Все данные в хранилище в определенный момент времени совместны (непротиворечивы). Для оперативных систем эта базовая характеристика данных соответствует совместности данных в момент доступа. Когда в оперативной среде осуществляется доступ к данным, ожидается, что данные имеют совместные значения только в момент доступа к ним.

Зависимость от времени хранилища данных проявляется в следующем.

Данные в хранилище представлены за временной промежуток от года до 10 лет. В оперативной среде представление данных осуществляется в промежутке от текущего значения до нескольких десятков дней. Приложения с высокой производительностью для обеспечения эффективного процесса транзакций должны работать с минимальным количеством данных. Следовательно, оперативные приложения ориентированы на короткий временной промежуток.

Другое проявление зависимости хранилища данных от времени заключается в его структуре. Каждая структура хранилища включает – явно или неявно – элемент времени.

Третье проявление зависимости хранилища данных от времени состоит в неукоснительном выполнении правила, что данные, однажды корректно в хранилище записанные, не могут быть обновлены. Хранилище данных с точки зрения практического использования представляет собой большую серию моментальных снимков. Естественно, если моментальный снимок данных был сделан некорректно, он может быть изменен. Но если был получен корректный моментальный снимок, то, однажды сделанный, он в последующем изменению не подлежит. Оперативные данные, будучи корректны в момент доступа к ним, могут обновляться по мере необходимости.

**Постоянство.** Четвертая определяющая характеристика хранилища данных – это постоянство. В оперативной среде операции обновления, добавления, удаления и изменения производятся над записями регулярно. Базовые манипуляции с данными хранилища ограничены начальной загрузкой данных и доступом к ним. В хранилище данных обновление данных не производится. Исходные (исторические) данные, после того как они были согласованы, верифицированы и внесены в хранилище данных, остаются неизменными и используются исключительно в режиме чтения.

Существуют важные последствия различия обработки данных в оперативной среде и обработки в хранилище данных. На уровне проектирования хранилища данных необходимость в поддержке механизмов, обеспечивающих корректность обновлений, отпадает – обновления в хранилище данных не производятся. Это означает, что на физическом уровне проектирования при решении проблемы нормализации и физической денормализации доступ к данным может оптимизироваться без каких-либо ограничений. Другое последствие простоты работы с данными хранилища касается технологии работы с данными. Технология работы с данными в оперативной среде отличается большей сложностью. Она поддерживает функции оперативного резервного копирования и восстановления, обеспечивает целостность данных, включает механизмы разрешения конфликтов и тупиковых ситуаций. Для обработки информации в хранилище данных указанные функции не столь критичны.

Характеристики хранилища данных – ориентированность на предметную область при проектировании, интегрированность данных, зависимость от времени и простота управления данными – определяют среду,

которая существенно отличается от классической транзакционной среды.

Источником почти всех данных среды хранилища данных являются оперативные среды. Может возникнуть ощущение, что существует огромная избыточность данных в обеих средах. Однако на практике избыточность данных в средах минимальна, поскольку:

1) При передаче данных из оперативной среды в хранилище данных эти данные фильтруются. Многие данные вообще никогда не выгружаются из оперативной среды. В хранилище данных передается только информация, используемая для обработки в системе поддержки принятия решений.

2) Временной горизонт в средах существенно различается. Данные в оперативной среде всегда являются текущими. Данные в хранилище имеют хронологию. С точки зрения временного горизонта пересечение между оперативной средой и средой хранилища данных минимально.

3) Хранилище данных содержит агрегированные (итоговые) данные, которые никогда не включаются в оперативную среду.

4) Передача данных из оперативной среды в хранилище данных сопровождается фундаментальными преобразованиями. Большинство данных при поступлении в хранилище видоизменяется.

**Сложные многомерные данные.** Традиционные многомерные модели данных и методы их реализации предполагают, что:

- ❖ все факты напрямую отображаются на значения измерений более низкого уровня, причем ровно на одно значение в каждом измерении;

- ❖ иерархии измерений представляют собой сбалансированные деревья.

Если эти предположения не выполняются, то стандартные модели и системы оказываются неадекватными. Особенно серьезные проблемы вызывают комплексные многомерные данные, поскольку они не являются суммируемыми (summarizable) — агрегированные результаты более высокого уровня нельзя получить из агрегированных результатов более низкого уровня. Запросы по результатам более низкого уровня будут приносить неверные данные, или предварительные вычисления, сохранение и последующее использование их результатов в данном случае невозможны. Вместо этого агрегированные результаты должны вычисляться непосредственно из базовых данных, что значительно увеличивает затраты на вычисления.

Суммирование требует применения распределенных агрегированных функций и значений иерархии измерений. Неформально иерархия измерений является «строгой», если ни одно из значений измерений не имеет более одного прямого родителя, «сюрьективной» (onto), если иерархия сбалансирована, и «покрывающей» (covering), если ни один локальный путь не «перескакивает» через уровень. Интуитивно это значит, что иерархии измерений должны быть сбалансированными деревьями.

Нерегулярные иерархии возникают в разных приложениях, в том числе в иерархии административных структур, иерархии медицинских диагнозов и иерархии концепций для Web-порталов, подобных Yahoo. Одно из решений — нормализовать нерегулярные иерархии, процесс, который

предусматривает пополнение несюрьективных и непокрывающих иерархий фиктивными значениями измерений, и перестраивает наборы родителей, для того чтобы решить проблемы нестрогих иерархий. Это преобразование может выполняться прозрачным для пользователя образом.

### **Архитектура и компоненты корпоративного Хранилища данных**

**Среда Microsoft Data Warehousing Framework.** Процессы создания, поддержки и использования хранилищ данных традиционно требовали значительных затрат, что в первую очередь было вызвано высокой стоимостью доступных на рынке специализированных инструментов. Эти инструменты практически не интегрировались между собой, так как были основаны не на открытых и стандартных, а на частных и закрытых протоколах, интерфейсах и т.д. Сложность и дороговизна делали практически невозможным построение хранилищ данных в небольших и средних фирмах, в то время как потребность в анализе данных испытывает любая фирма, независимо от масштаба.

Корпорация Microsoft приняла определенные меры, связанные с хранилищами данных и необходимостью по созданию инструментальной и технологической среды, которая позволила бы минимизировать затраты на создание хранилищ данных и сделала бы этот процесс доступным для массового пользователя. Это привело к созданию **Microsoft Data Warehousing Framework** ([http://mf.grsu.by/other/lib/olap/bd\\_wh/doc18.htm](http://mf.grsu.by/other/lib/olap/bd_wh/doc18.htm)) (рис. 16) — спецификации среды создания и использования хранилищ данных. Данная спецификация определяет развитие не только новой линии продуктов Microsoft (например, Microsoft SQL Server), но и технологий, обеспечивающих интеграцию продуктов различных производителей. Открытость среды Microsoft Data Warehousing Framework обеспечила ее поддержку многими производителями ПО, что, в свою очередь, дает возможность конечным пользователям выбирать наиболее понравившиеся им инструменты для построения своих решений.

Цель Microsoft Data Warehousing Framework — упростить разработку, внедрение и администрирование решений на основе хранилищ данных. Эта спецификация призвана обеспечить:

- ❖ открытую архитектуру, которая легко интегрируется и расширяется третьими фирмами;
- ❖ экспорт и импорт гетерогенных данных наряду с их проверкой, очисткой и возможным ведением истории накопления;
- ❖ доступ к разделяемым метаданным со стороны процессов разработки хранилища, извлечения и трансформации данных, управления сервером и анализа данных конечными пользователями;
- ❖ встроенные службы планирования задач, управления дисковой памятью, мониторинга производительности, оповещения и реакции на события.



Рис. 16. Microsoft Data Warehousing Framework

### Основные компоненты Microsoft Data Warehousing Framework.

Основными компонентами хранилища данных являются следующие:

- ❖ OLE DB — стандарт обмена данными;
- ❖ метаданные;
- ❖ средства хранения данных;
- ❖ средства OLAP-анализа;
- ❖ средства переноса и трансформации данных;
- ❖ средства представления и анализа данных;
- ❖ средства администрирования.

**OLE DB — стандарт обмена данными.** Построение хранилищ данных требует, с одной стороны, взаимодействия с различными оперативными БД для извлечения данных, а с другой — обмена данными и метаданными между различными компонентами. И та и другая задача решается крайне сложно — при отсутствии единого интерфейса для доступа к разнородным данным. Но такой интерфейс существует — это **OLE DB**. OLE DB целиком основан на открытой модели **COM** (Component Object Model) и представляет собой набор интерфейсов, которые могут быть использованы, например, в приложениях на Visual C++. Для упрощения использования OLE DB создан набор ActiveX-компонентов — **ActiveX Data Objects (ADO)**. Эти компоненты могут вызываться из приложений на Visual Basic, Access, Excel, встраиваться в активные Web-страницы и т.п. Практически все компоненты, которые мы будем обсуждать ниже, используют OLE DB для доступа к данным. OLE DB обеспечивает доступ не только к реляционным данным, но и к таким ресурсам, как почтовые сообщения, файловые каталоги, полнотекстовые индексы и т.п.

**Метаданные.** Одна из наиболее важных задач при построении хранилища данных — интеграция различных компонентов и инструментов, используемых для проектирования, хранения данных, переноса и

трансформации, а также анализа данных. Ключевым моментом при такой интеграции является возможность использования разделяемых метаданных (то есть данных о данных).

Центральным компонентом Data Warehousing Framework является хранилище метаданных (репозитарий) — Microsoft Repository, поставляемое как один из компонентов Microsoft SQL Server. Microsoft Repository — это база данных, которая хранит описательную информацию о компонентах программного обеспечения и об их отношениях. Microsoft Repository состоит из набора **Открытых информационных моделей (Open Information Model, OIM)**, а также набора опубликованных COM-интерфейсов. Открытые информационные модели — это объектные модели определенного типа информации, при этом они достаточно гибки, чтобы обеспечить поддержку новых типов информации. Корпорация Microsoft, опираясь на сотрудничество с представителями отрасли, уже разработала модели (OIM) для схемы баз данных (Database Schema), преобразования данных (Data Transformations) и OLAP. Последующие модели будут поддерживать репликацию, планирование задач, семантические модели и информационный справочник, который предназначен для обеспечения метаданными конечного пользователя.

Коалиция метаданных (The Metadata Coalition), отраслевой консорциум 53 производителей, заявила о поддержке Microsoft Repository. Открытые информационные модели получили широкую поддержку у независимых разработчиков ПО.

**Средства хранения данных.** Сердцем хранилища данных является, безусловно, СУБД, обеспечивающая надежное и производительное хранение и обработку данных. Как правило, данные из оперативных БД перемещаются в реляционное хранилище, где они становятся доступными для анализа. В дальнейшем, при использовании OLAP-средств, они могут быть перемещены в многомерную СУБД либо будут выбираться процессором многомерных запросов прямо из реляционных таблиц. Microsoft SQL Server обеспечивает как реляционный, так и многомерный вид хранения. Подробную информацию о Microsoft SQL Server можно найти в разделе «Microsoft SQL Server». Ниже кратко перечислены его основные характеристики: сначала возможности реляционной СУБД, а затем — многомерной.

Microsoft SQL Server обладает целым рядом свойств, делающих его превосходной платформой для построения хранилищ данных:

- ❖ поддержка баз данных, размер которых исчисляется терабайтами;
- ❖ масштабируемость как «вверх» — в сторону мощнейших современных аппаратных платформ для поддержки очень больших баз данных, так и «вниз» — в сторону серверов небольших рабочих групп и даже настольных и мобильных компьютеров (при этом обеспечивается полная совместимость);
- ❖ улучшенная обработка запросов, обеспечивающая оптимизацию и эффективное выполнение сложных запросов, типичных для хранилищ данных, в частности, запросов по схеме типа «звезда»;

- ❖ средства параллельного выполнения сложных запросов;
- ❖ эффективные средства настройки производительности, загрузки данных и построения индексов;
- ❖ распределенные запросы, позволяющие выбирать связанные данные из различных OLE DB-источников;
- ❖ надежные и эффективные средства тиражирования данных, незаменимые при поддержке нескольких связанных хранилищ или киосков данных.

Кроме того, средства тиражирования по-прежнему остаются одним из механизмов перемещения данных из оперативной БД в хранилище. Ниже рассматривается ряд механизмов, входящих в состав SQL Server.

**Средства OLAP-анализа.** OLAP все более популярная технология, которая может коренным образом усовершенствовать анализ данных. Microsoft SQL Server OLAP Services — это новый, полнофункциональный OLAP-сервер, поставляемый в составе SQL Server. OLAP Services включает в себя собственно сервер, доступный по протоколу OLE DB for OLAP, а также клиентский компонент, являющийся поставщиком протокола OLE DB for OLAP и обеспечивающий эффективное кэширование и возможность локального сохранения многомерных выборок для их дальнейшего анализа без подключения к OLAP-серверу.

Традиционно OLAP характеризовался дорогим инструментарием и сложным процессом реализации. Включение OLAP-функциональности в Microsoft SQL Server сделает многомерный анализ значительно более приемлемым с точки зрения затрат для небольших и средних организаций. Кроме того, небольшие группы или отделы в крупных организациях также смогут в полной мере воспользоваться новыми возможностями анализа — от сложной отчетности до продвинутых систем принятия решений.

**Средства переноса и трансформации данных.** Организация извлечения данных из оперативных БД, их очистки, интеграции и помещения в хранилище может потребовать значительных усилий и затрат, если не пользоваться встроенной в Microsoft SQL Server службой — Data Transformation Services (DTS). DTS обладает следующими свойствами:

- ❖ он на 100% использует OLE DB для доступа как к источнику, так и к приемнику данных. Благодаря этому DTS может извлекать и преобразовывать данные практически из любых источников (и, соответственно, помещать их в любые приемники данных);
- ❖ для переноса и трансформации данных используется расширяемый набор ActiveX-объектов, которыми легко управлять при помощи языка сценариев, например VBScript или JavaScript. Таким образом, имеются практически неограниченные возможности управления переносом и преобразованием данных;
- ❖ DTS способен интегрироваться с Microsoft Repository для использования метаданных об источнике, приемнике и схеме преобразования данных;
- ❖ задания по переносу и преобразованию данных, которые могут

включать в себя множество последовательных шагов, оформляются в виде пакетов (DTS Package), которые могут быть сохранены в хранилище метаданных (Repository), в базе SQL Server или в файле. Пакеты могут затем автоматически выполняться по расписанию при помощи сервиса SQL Server Agent.

**Средства представления и анализа данных.** Именно в сфере средства представления и анализа данных следует ожидать (и уже можно видеть) наибольшее количество продуктов, предлагаемых третьими фирмами, хотя и Microsoft предлагает здесь не только базовые технологии, но и средства для конечного пользователя. К ним относятся компоненты нового поколения Microsoft Office — Office 2000, прежде всего Microsoft Excel. Его популярное средство анализа данных PivotTable теперь сможет задействовать всю мощь OLAP-сервера, подключаясь к нему через упоминавшийся выше клиентский компонент PivotTable Services.

Еще одно средство от Microsoft — English Query — позволяет строить запросы к SQL Server на естественном языке (по-английски).

**Средства администрирования.** Существенной составляющей расходов на внедрение хранилища данных являются расходы на текущее сопровождение и администрирование хранилища. Средства администрирования, в том числе средства автоматизации выполнения административных задач, предусмотренные Data Warehousing Framework и включенные в состав Microsoft SQL Server, позволяют значительно сократить эти расходы.

Единой средой администрирования различных компонентов является **Microsoft Management Console**. Средства управления каждым конкретным компонентом (например, SQL Server или OLAP Services) представляют собой так называемый snap-in, то есть модуль администрирования, использующий единые средства пользовательского интерфейса. Средство управления SQL Server — **SQL Enterprise Manager** включает в себя более 25 программ-мастеров (Wizards), помогающих не слишком искушенному администратору решать самые важные задачи, в том числе создавать и копировать базы данных, производить настройку тиражирования, импорт/экспорт данных, управлять правами пользователей и т.п.

Кроме того, в SQL Enterprise Manager входят средства создания и редактирования графических диаграмм баз данных, значительно облегчающих создание и модификацию структуры хранилища.

Средства автоматизации администрирования позволяют создавать многоступенчатые задания, состоящие как из команд языка Transact-SQL, так и из сценариев на языках VBScript или JavaScript. При этом выполнение последующих шагов может быть поставлено в зависимость от результатов выполнения предыдущих. Эти задания могут охватывать множество серверов и выполняются по заданному расписанию.

**Основные поставщики ПО хранилищ данных:** корпорация «Электронный Архив», Arbor, Business Objects, Carleton, Cognos, Hewlett-Packard, IBM, Information Builders, Informix, Intellidex, Microsoft, MSP, NCR,

Oracle, Platinum Technology, Praxis, Prism, Pyramid, Red Brick, SAS Institute, Sequent, Software AG, Sybase, Tandem и др.

Все эти фирмы имеют страницы в Internet, где приводятся подробные сведения об их продуктах и услугах. Стоит отдельно отметить альянс Arbor и Seagate при встраивании OLAP в Crystal Info для СУБД Essbase.

### **Оперативная аналитическая обработка данных (OLAP)**

OLAP (On-Line Analytical Processing) - 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. Вся работа с OLAP-системой происходит в терминах предметной области.

В последнее время много написано про OLAP. Можно сказать, что наблюдается некоторый бум вокруг этих технологий. Правда, для нас этот бум несколько запоздал, но связано это, конечно, с общей ситуацией в стране. OLAP - аббревиатура от английского On-Line Analytical Processing - это название не конкретного продукта, а целой технологии. По-русски удобнее всего называть OLAP оперативной аналитической обработкой. Хотя в некоторых изданиях аналитическую обработку называют и онлайн-овой, и интерактивной, однако прилагательное “оперативная” как нельзя более точно отражает смысл технологии OLAP.

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. Вся работа с OLAP-системой происходит в терминах предметной области.

OLAP-системы являются частью более общего понятия Business Intelligence, которое включает в себя помимо традиционного OLAP-сервиса средства организации совместного использования документов, возникающих в процессе работы пользователей хранилища. Технология Business Intelligence обеспечивает электронный обмен отчетными документами, разграничение прав пользователей, доступ к аналитической информации из Интернет и Интранет.

В основе концепции OLAP лежит принцип многомерного представления данных.

**ROLAP.** Реляционная (relational) OLAP. Как и подразумевается названием, многомерная структура в таких инструментах реализуется реляционными таблицами. А данные в процессе анализа, соответственно, выбираются из реляционной базы данных аналитическим инструментом. Главный недостаток ROLAP по сравнению с многомерными СУБД - меньшая производительность. Для обеспечения производительности, сравнимой с

MOLAP, реляционные системы требуют тщательной проработки схемы базы данных и настройки индексов, то есть больших усилий со стороны администраторов БД. Только при использовании звездообразных схем производительность хорошо настроенных реляционных систем может быть приближена к производительности систем на основе многомерных баз данных.

В ROLAP, как правило, используются схемы «звезда» и «снежинка», при которых данные хранятся в таблицах фактов и таблицах измерений. Таблица фактов содержит одну строку для каждого факта в кубе. Для каждого измерения отводится отдельный столбец, содержащий значение параметра для конкретного факта, а также столбец для каждого измерения, которое содержит внешний ключ, ссылающийся на таблицу измерений для конкретного измерения.

**MOLAP.** Собственно многомерная (multidimensional) OLAP. В основе продукта лежит нереляционная структура данных, обеспечивающая многомерное хранение, обработку и представление данных. Соответственно и базы данных называют многомерными. Продукты, относящиеся к этому классу, обычно имеют сервер многомерных баз данных. Данные в процессе анализа выбираются исключительно из многомерной структуры. Подобная структура является высокопроизводительной. Хранение данных в многомерных структурах позволяет манипулировать данными как многомерным массивом, благодаря чему скорость вычисления агрегатных значений одинакова для любого из измерений. Однако в этом случае многомерная база данных оказывается избыточной, так как многомерные данные полностью содержат исходные реляционные данные.

Недостатки и преимущества каждого подхода, в общем-то, очевидны. Многомерная OLAP обеспечивает лучшую производительность, но структуры нельзя использовать для обработки больших объемов данных, поскольку большая размерность потребует больших аппаратных ресурсов, а вместе с тем разреженность гиперкубов может быть очень высокой и, следовательно, использование аппаратных мощностей не будет оправданным. Наоборот, реляционная OLAP обеспечивает обработку на больших массивах хранимых данных, так как возможно обеспечение более экономичного хранения, но, вместе с тем, значительно проигрывает в скорости работы многомерной.

В отличие от MOLAP, которая работает лучше, когда данные более менее плотные, серверы ROLAP лучше в тех случаях, когда данные довольно разрежены. Серверы HOLAP применяют подход ROLAP для разреженных областей многомерного пространства и подход MOLAP — для плотных областей. Серверы HOLAP разделяют запрос на несколько подзапросов, направляют их к соответствующим фрагментам данных, комбинируют результаты, а затем предоставляют результат пользователю. Материализация выборочных представлений в HOLAP, выборочное построение индексов, а также планирование запросов и ресурсов аналогично тому, как это реализовано в серверах MOLAP и ROLAP.

**DOLAP** - “настольный” (desktop) OLAP. Речь идет о такой аналитической обработке, где гиперкубы малы, размерность их небольшая, потребности скромны, и для такой аналитической обработки достаточно персональной машины на рабочем столе .

### **Витрины данных - необходимое звено в концепциях построения хранилищ данных**

Функционально ориентированные витрины данных (Data Mart) представляют собой структуры данных, обеспечивающие решение аналитических задач в конкретной функциональной области или подразделении компании, например управление прибыльностью, анализ рынков, анализ ресурсов и проч. Иногда эти структуры хранения данных называют также киосками данных. Витрины данных можно рассматривать как маленькие хранилища, которые создаются с целью информационного обеспечения аналитических задач конкретных управленческих подразделений компании.

Витрина содержит значительно меньше данных, охватывает всего несколько предметных областей и имеет более короткую историю. Витрины данных можно представить в виде логически или физически разделенных подмножеств хранилищ данных. Обычно они строятся для обслуживания нужд определенной группы пользователей.

Источником данных для витрин служат данные хранилища, которые, как правило, агрегируются и консолидируются по различным уровням иерархии. Детальные данные могут также помещаться в витрину или присутствовать в ней в виде ссылок на данные хранилища.

Различные витрины данных содержат разные комбинации и выборки одних и тех же детализированных данных хранилища. В хранилищах данных имеются большие области для сбора/хранения/перемещения существующих данных, откуда данные могут быть перераспределены по или витринам данных, которые как раз и предназначены для доступа пользователей, принимающих решения. В то время как хранилище данных предназначено для управления данными поступающими крупными партиями от их поставщиков (например, операционных систем), а так же для организации и хранения этих данных, витрины данных могут сосредоточиться на упаковке и предложении наборов данных конечным пользователям, зачастую удовлетворяя конкретные потребности. витрина данных по своему исходному определению - это набор тематически связанных баз данных, которые содержат информацию, относящуюся к отдельным аспектам деятельности корпорации. По сути дела, витрина данных - это облегченный вариант хранилища данных, содержащий только тематически объединенные данные. Целевая база данных максимально приближена к конечному пользователю и может содержать тематически ориентированные агрегатные данные. витрина данных, естественно, существенно меньше по объему, чем корпоративное хранилище данных, и для его реализации не требуется особо мощная вычислительная техника.

Фактическим стандартом структуры данных при разработке витрины является схема "звезда" или схема «снежинка».

## ЛИТЕРАТУРА

1. Гладкий А.В. Формальные грамматики и языки. - М.: Наука, 1973. - 368с.
2. Баласанян В.Э. Электронный документооборот – основа эффективного управления современным предприятием / В.Э. Баласанян // Секретарское дело. – 2010.— №2. – С. 46—48.
3. Гайдукова Л.М. Проблемы традиционных технологий документационного обеспечения / Л.М. Гайдукова// – Секретарское дело. – 2011. – №10 – С. 17–22.
4. Глик Д.И. Национальные стандарты в области электронного документооборота / Д.И. Глик// - Секретарское дело. – 2011. – № 9 – С. 45–75.
5. Домарев В.В. Защита информации и безопасность компьютерных систем/ В.В. Домарев. – К.: «Диа-Софт», 2009. – 38 с.
6. Доронина Л.А. Новые информационные технологии хранения документов организации (к постановке проблемы) / Л.А. Доронина, Б.В. Черников // Управление персоналом. – 2011. – №2. – С. 39–43.
7. Живаева О.В. Управление информационными потоками промышленных предприятий на основе оптимизации документооборота. – М., 2009. – 148 с.
8. Кузнецов С.Л. Проблемы выбора программного обеспечения для автоматизации работы с документами в офисе / С.Л. Кузнецов // Секретарское дело. – 2010. – №10. – С. 32–35.
9. Максимович Г.Ю. Современные информационные технологии хранения информации и организация доступа к ней / Г.Ю. Максимович, В.И. Берестова // Секретарское дело. – 2011. – №1 (53) – С. 34 (2011 в).
10. Митяев К.Г. Классификационные схемы (классификаторы) документов в делопроизводстве / К.Г. Митяев // Вестник архивоведения. – 2003. – № 4. – С. 81–83.
11. Плешкевич Е.А. Формирование документной парадигмы / Е.А. Плешкевич // НТИ. – Сер. 1. – 2005. – № 10. – С. 19–22 (2005 в).
12. Сысоева Л.А. Системы электронного управления документами / Л.А. Сысоева // Секретарское дело. – 2012. – №3. – С.29–33.
13. Тихонов В.И. Электронные архивы и электронный документооборот / В.И. Тихонов, И.Ф. Юшин// Отеч. Архивы. – 2006. – №2. – С.17–26.
14. Ткачев А.В. Правовой статус компьютерных документов: основные характеристики / А.В. Ткачев // – М.: ООО «Городец-издат», 2012. – С.62.
15. Developer's Guide for Delphi 7, Borland Inprise Corporation, 100 Enterprise Way, Scotts Valley, CA 95066-3249
16. Object Pascal Language Guide, Borland Inprise Corporation, 100 Enterprise Way, Scotts Valley, CA 95066-3249
17. <http://www.abbyy.ru>
18. <http://www.boss-referent.ru>
19. <http://www.businessdoc.ru>
20. <http://www.documentum.ru>