

Информационное обеспечение автоматизированных информационных СИСТЕМ

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

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

1. Совокупность форм, методов и технологий отражающих структуру и организацию подготовки информации для автоматизированной обработки
2. Характеристики и структуру информации используемой для обмена между пользователями в процессе управления экономическими объектами
3. Методы и технологии использования данных, полученных в результате автоматизированной обработки, для принятия управленческих решений

Классификаторы

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

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

С точки зрения компьютера это совершенно разные слова. Поэтому для обозначения лучше использовать цифровой код. Такой, например, как номер телефона. Но лучше в качестве универсального номера для обозначения

предприятия использовать уникальный цифровой код не связанный с его реквизитами. Совокупность таких кодов называется классификатором. Классификаторы повсеместно используются для обозначения экономических объектов. Например, почта России традиционно использовала внутреннюю систему кодирования предприятий связи (6 цифр) - известный всем индекс на почтовом конверте. Свой код предприятия введен Министерством по налогам и сборам. Все налогоплательщики пронумерованы в компьютерах МНС РФ специальным кодом - так называемым идентификационным номером налогоплательщика, ИНН.

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

Существуют международные системы кодирования банков. Например, международный банковский идентификационный код BIC (Bank Identification Code) длиной 8-11 знаков, который присваивается всем банкам - участникам международной системы межбанковских переводов S.W.I.F.T.

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

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

Единая система кодирования предприятий должна отвечать следующим требованиям.

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

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

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

Стандартом ассоциации EAN International (General EAN/UCC Specifications) введен в действие единый международный Глобальный идентификационный

номер EAN/UCC Global Location Number (GLN).

GLN представляет собой 13-разрядный цифровой код (номер),

предназначенный для точного и краткого обозначения:

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

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

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

Системы кодирования

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

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

Наиболее распространены следующие системы кодирования:

- Порядковая
- Серийная
- Разрядная
- Коды повторений.

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

Пример. Кодирование списка студентов в группе.

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

Недостатки: отсутствие возможности расширения наименований без нарушения принятой классификации.

Серийная система кодирования предусматривает присвоение серии номеров признакам старшего порядка и самостоятельных кодов признакам младшего порядка. В отведённых сериях каждой группы признаков старшего порядка оставляются резервные номера для кодирования новых наименований.

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

Преимущества:

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

Недостатки:

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

Десятичная или разрядная система кодирования характеризуется закреплением одного или нескольких цифровых разрядов за каждым группировочным признаком внутри многопризначной номенклатуры, т.е. закреплением за признаками номеров, кратных десяти. Каждый разряд или

несколько цифровых разрядов определяет какой-то признак классификации номенклатуры.

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

Преимущества:

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

Недостатки:

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

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

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

Преимущества:

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

Недостатки: неудобны при использовании в автоматизированной обработке.

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

Смешанная или комбинированная система кодирования основана на использовании различных систем, изложенных выше. Смешанные коды применяются для кодирования многозначных номенклатур, содержащих несколько различных признаков. Они широко используются для кодирования производственных затрат, когда балансовые счета, субсчета и статьи

расходов обозначаются кодами повторения, а изделия и заказы кодируются с помощью серийной, десятичной или другой системы кодирования. При смешанной системе наиболее рационально используются преимущества каждой из применяемых систем кодирования.

Манипулирование экономической информацией

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

Процесс решения этих задач включает следующие этапы:

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

Совокупность всех этих действий называется манипулированием экономической информацией.

Организация сбора экономической информации

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

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

Технологии перевода бумажных документов в электронные

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

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

Обычно процесс сканирования - это промежуточная стадия получения электронного документа. Для того чтобы можно было редактировать документ, осуществлять поиск по нему или использовать его фрагменты при подготовке новых документов и т.д., необходимо перевести полученный образ в текстовый документ, понятный офисным программам. Поэтому следующая задача заключается в распознавании отсканированных документов. Для этого необходим специальный инструмент, способный перевести изображение в текстовый редактируемый электронный документ. Такие инструменты существуют, их общее название - программы оптического распознавания символов (optical character recognition, OCR). С помощью OCR-программы компьютер сможет "прочитать" на отсканированной странице текст, отделив его от иллюстраций и прочих элементов оформления, найти таблицы и "разобраться" в их содержимом. А затем скомпоновать все это заново, воссоздав внешний вид страницы.

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

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

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

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

Перевод каждого из перечисленных видов документов имеет свою специфику. Если вводятся фотографии, то достаточно электронного

изображения, если документ содержит текст, его необходимо распознать, если это форматированный текст с рисунками, то нужно не только распознать текст, но и восстановить формат документа, скорее всего, сам документ вообще не нужен, важна только содержащаяся в нем информация.

Ввод формализованных документов

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

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

При заполнении формализованного документа требуется внести информацию в определенные служебные поля, а задача программы при вводе форм - определять эти специальные поля, распознавать информацию в них и ввести ее в базу данных. При этом особенностью ввода форм в компьютер является необходимость распознавания текстов, заполненных от руки. Обычно в том случае, если форма должна распознаваться компьютером, заполняющего просят ввести текст отдельными буквами, и такой текст называют рукопечатным. Технологии распознавания рукопечатных символов обозначаются термином ICR (Intelligent Character Recognition). Распознавание рукопечатных символов представляет собой более сложную задачу по сравнению с распознаванием печатных, поскольку требуется распознать символ, вписанный в форму от руки с учетом возможных его отклонений, обусловленных индивидуальными особенностями почерка.

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

Виды систем ввода документов

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

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

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

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

- **Продукты для персонального ввода неформализованных документов**

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

- **Продукты для промышленного ввода документов**

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

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

Среди российских разработок промышленного ввода документов

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

- **Продукты для персонального ввода форм**

Наиболее известными решениями являются ABBYY FormReader и система Cognitive Forms.

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

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

Процедура создания шаблона формы в большинстве случаев достаточно проста. FormReader позволяет различить и идентифицировать в едином потоке около 100 различных шаблонов и форм.

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

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

На следующем этапе определяются языки, используемые в распознаваемом тексте, типы данных для полей и правила контроля. ABBYY FormReader предлагает большое число готовых правил контроля, которые позволяют гарантировать правильность ввода информации

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

Эта система позволяет вводить от 2000 до 3000 страниц ежедневно. В программный комплект Cognitive Forms входит отдельный модуль "Дизайнер форм", с помощью которого пользователь может самостоятельно создавать формы документов, задавать описания полей и варианты контекстных проверок.

- **Продукты для промышленного ввода форм**

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

Хранение информации. Базы и хранилища данных

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

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

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

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

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

На уровне физической модели электронная БД представляет собой файл или группу файлов в специализированном формате конкретной СУБД.

В настоящее время наибольшее распространение получили реляционные базы данных.

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

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

- данные хранятся в таблицах, состоящих из столбцов ("атрибутов") и строк ("записей");
- на пересечении каждого столбца и строчки стоит в точности одно значение;
- у каждого столбца есть свое имя, которое служит его названием, и все значения в одном столбце имеют один тип;
- между таблицами установлены соответствия, называемые связями;
- запросы к базе данных возвращают результат в виде таблиц, которые тоже могут выступать как объект запросов;
- строки в реляционной базе данных неупорядочены, упорядочивание производится в момент формирования ответа на запрос.

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

Классификация баз данных.

По организации хранения данных:

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

По способу доступа к БД:

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

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

Клиент-серверные СУБД. Такие СУБД состоят из клиентской части (которая входит в состав прикладной программы) и сервера. Клиент-серверные СУБД, в отличие от файл-серверных, обеспечивают разграничение доступа между пользователями и меньше загружают сеть и клиентские машины. Сервер является внешней по отношению к клиенту программой, и по мере надобности его можно заменить другим. Недостаток клиент-серверных СУБД - в самом факте существования сервера (что плохо для локальных программ - в них удобнее встраиваемые СУБД) и больших вычислительных ресурсах, потребляемых сервером. Примеры: Interbase, MS SQL Server, Oracle, DB2, PostgreSQL, MySQL (старше версии 5.0).

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

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

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

Примеры: SQLite, BerkeleyDB, один из вариантов Firebird, один из вариантов MySQL.

В основе концепции хранилищ данных лежат следующие основополагающие идеи:

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

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

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

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

Достоинствами архитектуры классического хранилища данных являются:

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

- многообразие механизмов обработки и представления данных.

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

Организация обмена данными для управления экономическими объектами

Использование Интернет для обмена данными в среде управления

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

- глобальное сообщество произвольно объединяемых мировых сетей, которые используются для свободного обмена данными, информацией и знаниями;
- совокупность технологий, которые реализуют обмен данными на основе использования семейства протоколов TCP/IP (Transmission Control Protocol / Internet Protocol), называемых Internet-технологиями или технологиями Internet.

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

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

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

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

Эти решения оказались успешными, стандартизация протоколов позволила подключать к сети компьютеры с различным базовым программным обеспечением. Появилось понятие "трафик", трактуемое в единицах обмена информацией, которым стали измерять реальную загрузку сети. Технология передачи данных IP-пакетами оказалась чрезвычайно перспективной в техническом отношении, однако в чисто пользовательском плане ее необходимо было дорабатывать, так как скорость передачи данных не могла компенсировать значительные затраты времени на поиск нужной информации в огромных массивах данных.

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

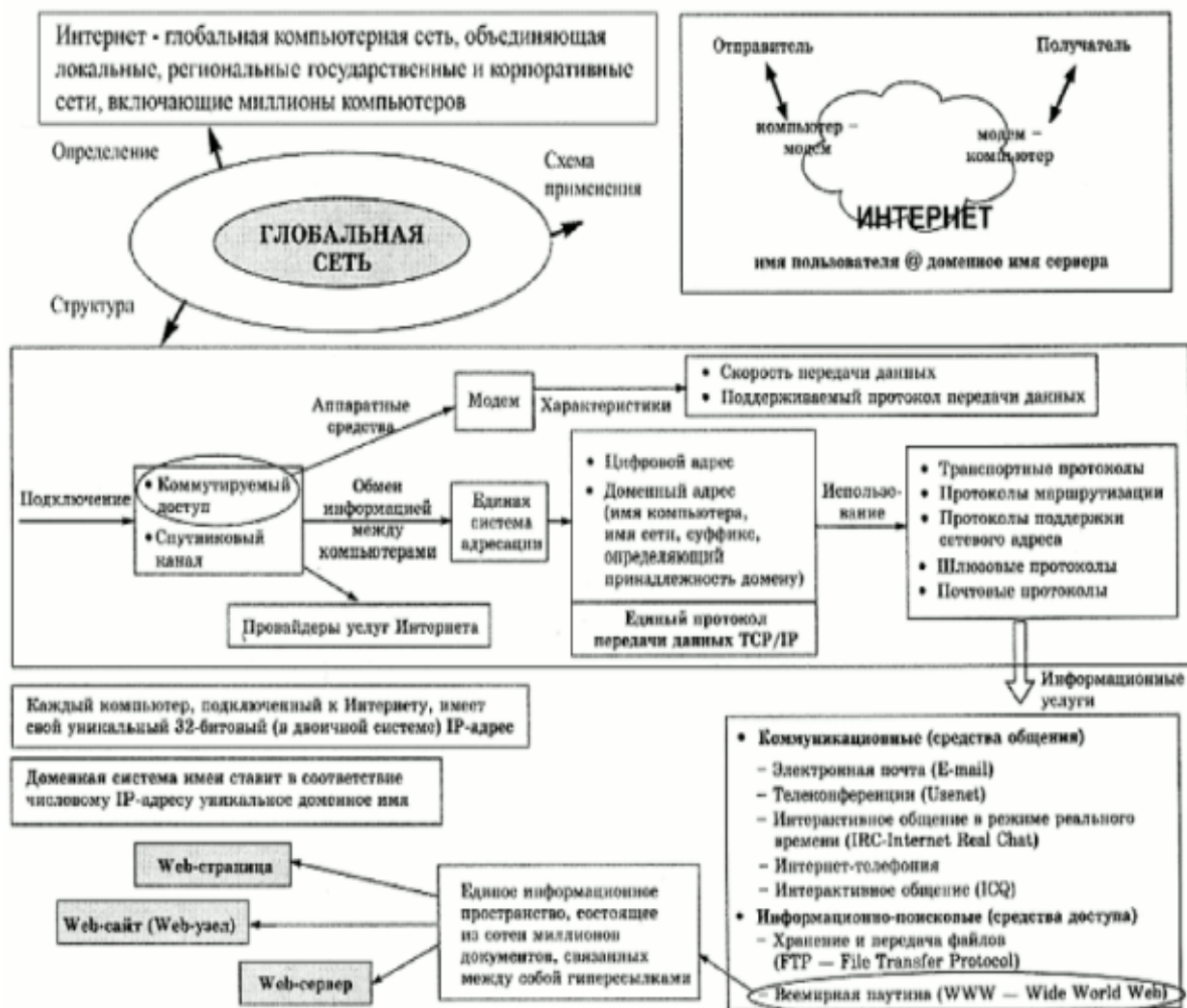
Основная метафора Web-гипертекста - это *"электронная книга"* с автоматически поддерживаемыми мгновенными переходами по ссылкам. Сам же термин *гипертекст* был впервые предложен Тедом Нельсоном в 1965 году, а первую работающую гипертекстовую систему создал в 1968 году Дуг Энгельбард.

В 1991 году был создан первый *браузер* (Browser) - компьютерная программа просмотра гипертекста, - работавший в режиме командной

строки. Его применение позволило уже в 1992 году успешно реализовать предложенный проект, направленный в конечном итоге на создание *"бесшовного информационного пространства"* (Seamless Informational Area), охватывающего всю планету.

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

В сентябре 1994 года Оливер Мак-Брайан (Oliver McBryan) из Колорадского университета (США) разработал одно из первых автоматических средств составления предметного указателя для WWW, названное WWW-Worm. За несколько минут Worm формировал базу данных из 300000 мультимедийных объектов, которые можно было находить по ключевым словам. Можно считать, что с этого момента информационное пространство World Wide Web было в принципе сформировано. Дальнейшее развитие шло по линии совершенствования технологий поиска, передачи, обеспечения безопасности, разработки и стандартизации различных Web-интерфейсов, повышающих комфорт использования Web-технологий. С середины 90-х годов эти технологии стали находить все более широкое применение во многих сферах человеческой деятельности.



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

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

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

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

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

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

- служба пересылки и приема сообщений (E-mail);
- служба гипертекстовой среды (WWW);

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

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

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

Инструментальные средства разработки Internet-приложений разнообразны и включают:

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

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

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

Средства мультимедиа предназначены для создания звукового и музыкального сопровождения, анимационных и видеороликов. Часто

воспроизведение файлов мультимедиа осуществляется клиентскими расширениями или специальными Helper-программами.

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

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

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

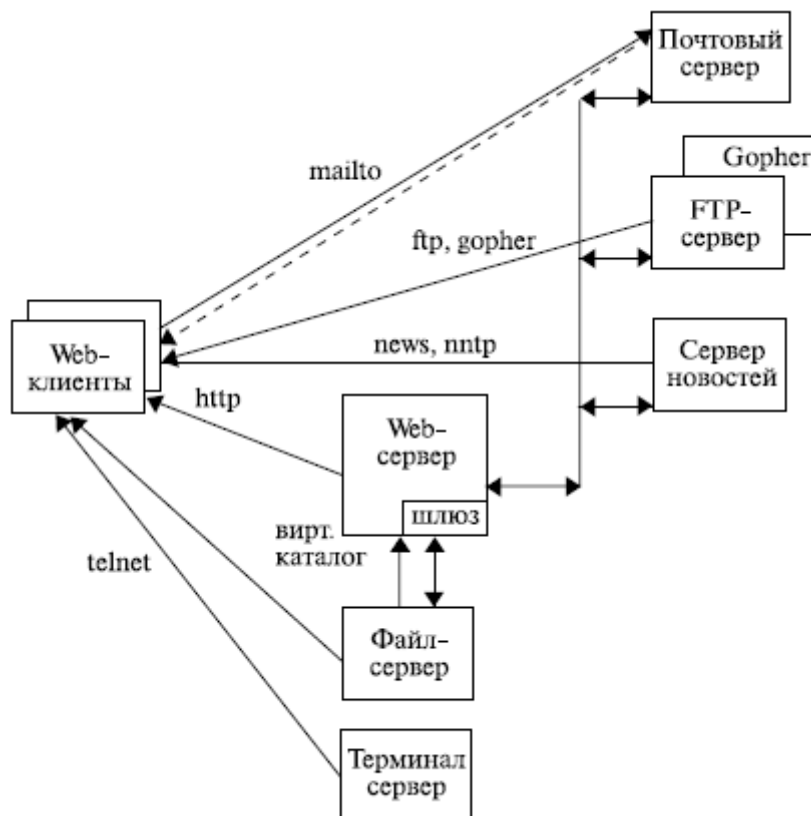
Средства безопасности могут быть встроены в программное обеспечение Internet-серверов или представлены в виде дополнительных компонентов: комплексов Firewall и Proxy-серверов, выполняющих фильтрацию данных на различных уровнях.

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

через специальный графический интерфейс (Computer Graphical Interface - CGI).

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

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



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

Пользователь работает с данными в привычном интерфейсе, пользуясь средствами Microsoft Office для доступа к сетевым данным.