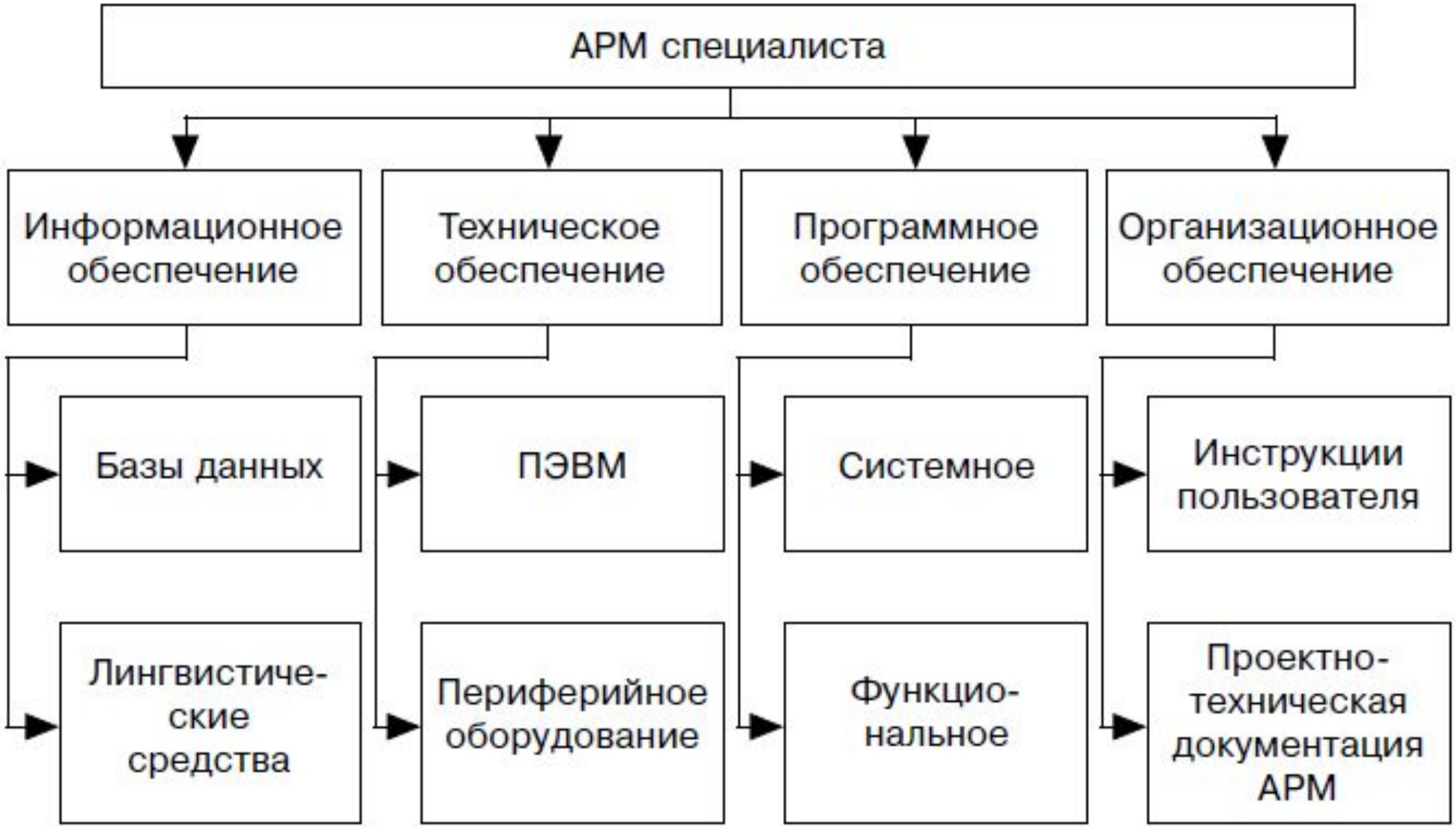
ИТ НА РАБОЧЕМ МЕСТЕ ПОЛЬЗОВАТЕЛЯ

- Качество любой ИТ зависит от того, насколько грамотно она адаптирована к конкретным условиям функционирования. В большей мере эти условия определяются особенностями рабочего места пользователя — специалиста, который применяет ИТ для решения своих профессиональных задач.
- В настоящее время в компьютерной обработке данных широкое распространение получили АРМ пользователей.
- Автоматизированное рабочее место — это размещенный на рабочем месте пользователя персональный компьютер, реализующий совокупность методов, средств и процедур информационного, технического, программного и организационного характера по решению профессиональных задач пользователя.

- За рубежом в первой половине 1970-х гг. появился термин *workstation*, смысл которого во многом совпадает с понятием АРМ. В понятие АРМ обычно входит комплекс программно-технических компьютерных средств, телекоммуникаций и оргтехники, предназначенный для решения конечным пользователем своих функциональных задач, а также процедур генерации, сбора и обработки информации в любой сфере деятельности на рабочем месте.

- Наметилась тенденция создания унифицированных АРМ, обслуживающих несколько предметных областей. Например, АРМ-аналитик, созданный на базе АРМ-статистика, значительно расширяет возможности последнего по решению комплекса функциональных задач. АРМ имеет определенную структуру (см рис.)


АРМ специалиста
Информационное обеспечение
Техническое обеспечение
Программное обеспечение
Организационное обеспечение
Базы данных
ПЭВМ
Системное
Инструкции пользователя
Лингвистиче-ские средства
Периферийное оборудование
Функцио-нальное
Проектно-техническая документация АРМ

- В состав баз данных АРМ входят файлы, документы, показатели, данные, обеспечивающие решение задач по профилю специалиста. Лингвистические средства, как правило, составляют ИПЯ, классификаторы, справочники, методики индексирования документов перед их вводом в ЭВМ, критерии смыслового соответствия и др. В состав системного ПО входят компоненты, рассмотренные ранее. Функциональное ПО содержит программы, обеспечивающие решение функциональных задач по планированию, учету, анализу, контролю, оптимизации, имитации, редактированию и др.

- Общими требованиями, предъявляемыми к АРМ, являются: удобство и простота общения с ними, в том числе настройка АРМ под конкретного пользователя и эргономичность конструкции; оперативность ввода, обработки, размножения и поиска документов; возможность оперативного обмена информацией между персоналом организации с различными лицами и организациями за ее пределами; безопасность для здоровья пользователя. Широкое применение находят АРМ для подготовки текстовых и графических документов, обработки данных, в том числе в табличной форме, создания и использования баз данных, проектирования и программирования. Требования к аппаратным и программным средствам АРМ заключаются в обеспечении технологичности выполняемых процедур, «дружественного» интерфейса и эргономичности.

АРМ конфигурируется с учетом следующих условий:

- содержания задач, решаемых пользователем;
- функциональных обязанностей пользователя;
- режимов работы и решения задач пользователя;
- преемственности в применении ИТ на фирме и мнений пользователя;
- финансовых возможностей фирмы по приобретению ППП;
- перспективности применения ППП и др.;
- схемы взаимодействия программ.

- Основное назначение АРМ состоит в повышении интеллектуального уровня решения задач и автоматизации рутинных процедур специалиста. Выделяют АРМ руководителя, секретаря, специалиста, технического и вспомогательного персонала и др. При этом в АРМ используются различные ОС и прикладные программные средства, зависящие, главным образом, от функциональных задач и видов работ, в частности управленческих, технологических, творческих, административно-организационных и технических.

- Методы, средства и процедуры АРМ основаны на идентичных категориях, которые применяются в ИТ. В область режимов функционирования ИТ входит комплекс условий, который определяется системообразующими признаками ИТ: принципы создания и функционирования, цели, задачи, функции, структура, технология процедур и процессов, развитие ИТ и др.

- При создании АРМ разработчики должны исходить из того, что в общем случае АРМ является квинтэссенцией ТПОД. Один из основных принципов построения АРМ — это придание ему свойства адаптивности к изменяющимся условиям в управлении объектом. Выдаваемые администратору сведения должны быть информативными, достоверными, полными, оперативными. Форма отображения данных должна обеспечивать их восприятие, анализ и принятие рациональных решений специалистом в минимальное время. Поэтому основной акцент при создании, эксплуатации и развитии АРМ должен быть направлен на его логическую составляющую.

Для решения задач в АРМ могут быть реализованы следующие функции:

- ввод данных в ЭВМ с первичных документов;
- контроль вводимых данных;
- корректировка информации;
- обработка текстовой, табличной, графической и мультимедиа информации;
- расчет данных по задаваемым формулам;
- накопление данных и их хранение по годам и периодам;
- логико-математическая обработка информации для получения информации по запросам;
- формирование сводных отчетов с выдачей на экран дисплея, на печать, а также с записью на магнитные носители;

- группировка данных по характеристикам предприятий и показателям;
- обработка и представление данных в различных разрезах;
- построение графиков и диаграмм по финансово-экономическому состоянию предприятий;
- пересылка данных абонентам по каналам связи.

- В результате работы АРМ полученная информация может быть выведена на печать, записана на магнитные носители, а также передана в виде файла по каналам связи на вышестоящие (нижестоящие) уровни управления объектом. Организация работы АРМ базируется на применении системной технологии обработки информации, включающей следующие этапы:

- формирование электронных таблиц на основе документов с использованием необходимых расчетов и методов контроля информации;
- создание сводных документов в необходимых разрезах, используя данные электронных таблиц;
- формирование локальных баз данных;
- передача подготовленных сводных отчетов, справок, аналитических записок и других материалов в вышестоящие и другие инстанции;
- получение поручений и запросов на выполнение указанных выше
- видов работ с учетом изменений и дополнений, выработанных на основе анализа переданных материалов.

ИТ ЭЛЕКТРОННОГО ОФИСА

- Информационный обмен между людьми как в рамках предприятия, так и за его пределами происходит очень интенсивно. В современный офис поступает большой поток информации. Причем с каждым годом объем потока заметно возрастает. Значительная часть этого потока поступает в виде разного рода сообщений — электронных писем, факсов, голосовых сообщений. В последнее время к этому добавились еще и видеофрагменты, пересылаемые с помощью электронных средств связи.

- Коммерческий успех предприятия в значительной степени зависит от того, насколько его сотрудникам удается, во-первых, осмысливать и упорядочивать входящую информацию, во-вторых, оперативно отвечать на поступающие запросы. Для решения этих задач крайне желательно использовать технологию, благодаря которой можно было бы обращать внимание только на содержание сообщений, в максимальной степени абстрагируясь от их конкретной формы — будь то факс, электронное письмо или голосовое сообщение. Так родилась идея создания единой среды обмена сообщениями (от англ. *unified messaging).*

- Суть единой среды обмена сообщениями сводится к следующему. Вся входная информация — документы, голосовые и факсимильные сообщения, электронные письма попадает в один почтовый ящик, т.е. в общее хранилище данных. С содержимым этого почтового ящика пользователь может ознакомиться, используя настольный компьютер, телефон или переносной компьютер. С помощью компьютера пользователь может рассмотреть список полученных сообщений и краткие аннотации. Щелкнув мышью на нужном сообщении, можно просмотреть или прослушать его независимо от того, в какой форме оно поступило. В то же время с помощью телефона пользователь получает возможность прослушать голосовые сообщения, переслать факсы на находящийся поблизости факсимильный аппарат. Электронные письма можно либо переслать туда же по факсу, либо прослушать в голосовом виде

- При работе в единой среде обмена сообщениями их физическая форма в принципе может оставаться почти полностью скрытой от пользователя. Эта независимость дает ему возможность всецело сосредоточиться на содержании сообщения. Пользователь всегда может ознакомиться с любыми поступившими на его имя сообщениями независимо от формы сообщений, своего местонахождения и времени суток. Для этого нужен минимум технологических ресурсов.

- При таком подходе часто бывает необходимо, чтобы в БД электронных писем имелись записи, соответствующие голосовым и факсимильным сообщениям. При этом сообщения хранятся отдельно. Были созданы системы, в которых все сообщения независимо от их формы хранятся в единой БД. Обычно такие системы представляют собой надстройку над программой электронной почты или системой поддержки коллективной работы, обеспечивающую интеграцию сообщений разных типов в готовую систему обмена сообщениями.

- В частности, существуют программы, организующие единую среду обмена сообщениями на базе *Microsoft Exchange*, *Lotus Notes* и *Novell Group Wise*. При этом управление базой сообщений осуществляется средствами той программы, на которой базируется надстройка *unified messaging*. В этом случае значительно удобнее выполнять преобразование между разными типами сообщений

- При условии применения в качестве аппаратной базы для систем голосовой почты учрежденческих АТС всегда существовала проблема обмена информацией между вычислительной и телефонной сетями. Без организации такого обмена режим *unified messaging* построить невозможно. Проблема упрощается с появлением компьютерно-телефонного оборудования на базе открытых стандартов — плат расширения для компьютера. Посредством этих плат можно строить большие и достаточно развитые в логическом плане системы. Системы, построенные на базе плат компьютерной телефонии, работают под управлением обычных программ ЭВМ. Это упрощает получение нужной информации.

- Компания Octel Communications предлагает своим клиентам программу *Octel Unified Messenger,* представляющую собой надстройку над системой *Microsoft Exchange*. Программа помещает все голосовые сообщения в *Exchange Inbox* (входной почтовый ящик обмена). Эта программа обеспечивает единое управление электронными письмами и голосовой почтой. Тип сообщения (голосовой фрагмент или электронное письмо) указывается в поле заголовка.

- Как и другие современные продукты, система *Octel Unified Messenger* позволяет знакомиться с содержанием голосовых сообщений и электронных писем в дистанционном режиме с использованием компьютера или телефона. В последнем случае (чтение электронного письма по телефону) в технологию включается модуль преобразования «текст-речь».

- В офисной работе постоянно приходится обращаться к правовой информации. Это необходимо для того, чтобы принимаемые решения не противоречили правовым законодательным нормам. Эта задача в информационном плане решается на базе АИПС Базы данных документальных АИПС могут содержать редуцированные (сокращенные) характеристики или полные тексты нормативных документов — законов, указов, постановлений, инструкций и других видов документов органов власти и управления. К редуцированным основным характеристикам какого-либо определенного нормативного акта в данном случае могут быть отнесены поисковый образ документа, его аннотация или реферат и адрес хранения самого документа. В случае необходимости обращения к тексту документов пользователь находит нужный документ по его адресу. При условии поиска документов в полнотекстовой базе данных пользователю на его запрос выдаются полные тексты всех релевантных документов.

- Для актуализации и целостности БД правовых АИПС фирмы- разработчики обеспечивают сопровождение этих систем, как правило, на договорных началах. В России наиболее известными являются информационно-правовые справочные системы «Консультант Плюс», «Гарант-Сервис», «Кодекс», «1С: Электронный справочник бухгалтера», «Консультант-Бухгалтер».

- Одной из рациональных форм построения ИТ электронного офиса являются пакеты программ, реализующие интеграцию офисных задач. Примером может служить известный пакет*Microsoft Office.* В этом пакете реализованы технологии текстовых редакторов, электронных таблиц, машинной графики, мультимедиа, СУБД и др.

- Использование Интернета позволило создать разновидность электронного офиса, получившую название «виртуальный офис». В подобных случаях основные функции информационного обслуживания управленческой деятельности и информационные ресурсы не сосредоточены в реальном офисе с его атрибутами — помещением, оборудованием, персоналом и т. п. Эти функции и ресурсы пространственно распределяются в различных узлах информационной сети.

ТЕХНОЛОГИИ ОБРАБОТКИ ГРАФИЧЕСКОЙ ИНФОРМАЦИИ

- В последние годы в работе пользователей широкое распространение получили технологии обработки графической информации, или графических образов. Обычно под графической информацией можно понимать графический образ как модель объекта — рисунок, чертеж, фотографию, иллюстрации в изданиях или картину живописи, изображение на экране телевизора и т.д.

- Технология обработки графических образов — это разновидность ИТ по обработке и получению графических изображений и выдаче их на устройства отображения. Этими устройствами могут быть дисплей, принтер, плазменный экран, плоттер и др. Любое компьютерное изображение состоит из набора графических примитивов, которые отражают некоторый графический элемент. Примитивами могут также быть алфавитно-цифровые и любые другие символы. Совокупность графических примитивов, которой можно манипулировать, называют сегментом отображаемой информации. Наряду с сегментом часто используется понятие «графический объект». Графический объект — это множество примитивов, обладающих одинаковыми визуальными свойствами и статусом, идентифицированных одним именем.

- Графическая информация на экране монитора компьютера образуется из точек. В графическом режиме экран монитора представляет совокупность светящихся точек — пикселей (от англ. *pixel, picture element*). Суммарное количество точек на экране называют разрешающей способностью монитора, которая зависит также от его типа и режима работы. Единицей измерения в этом случае является количество точек на дюйм (dpi)

- Преимущество использования графического интерфейса пользователя в ОС заключается в том, что он позволяет создавать одинаковые графические изображения для всех устройств, поддерживаемых ОС. Реализуется принцип WYSIWYG (*What You See Is What You Get* — «что видим, то и получаем»).

- Различают растровую, векторную и фрактальную компьютерную графику. Эти виды различаются принципами формирования изображения. Для каждого из них используется свой способ кодирования. Использование цветных изображений связано с тем, что каждый пиксель должен представлять комбинацию битов, определяющую его цвет. При растровом методе такую комбинацию битов часто называют битовой картой (от англ. *bit map).* Она представляет карту или схему исходного изображения.

- Файлы *растровой* (или битовой) *графики* содержат в определенной последовательности совокупность отдельных точек растровых изображений (от англ. *bitmap images*). В качестве графических редакторов, работающих с растровой графикой, используют *Paint*, *Adobe Photoshop* и др. Форматы файлов растровой графики (BMP, PCX, GIF, TIF и JPEG) предусматривают собственные способы кодирования информации о пикселях и другой присущей компьютерным изображениям информации. Кроме того, графические редакторы предлагают собственные форматы графических данных, например EPS, PSD, PDD, CDR, CMX, которые могут преобразовываться в другие графические форматы с помощью специальных конверторов.

- К недостаткам растровых изображений относят их большой объем и невозможность сильного увеличения рисунка, так как видны составляющие его точки. Имеет место эффект, обозначаемый как *пикселизация.*

- *Векторное изображение* представляет графический объект, состоящий из линий — элементарных отрезков и дуг. Положение этих элементарных объектов определяется координатами точек и длиной радиуса. При этом основным элементом векторного изображения является не точка, а линия. Следовательно, *линия* — элементарный объект векторного изображения. Для каждой линии указываются ее характер (сплошная, пунктирная, штрих-пунктирная), толщина и цвет. К другим свойствам линии относят ее форму. Замкнутые линии можно заполнить каким-нибудь цветом, текстурой или картой. Любая простая линия имеет две вершины, называемые *узлами.* Информация о векторном изображении кодируется как обычная буквенно-цифровая и обрабатывается специальными программами. При каждом отображении векторное изображение перерисовывается компьютером, что несколько замедляет работу, но позволяет получать изображения с высоким разрешением.

- В векторной графике объем памяти, занимаемый линией, не зависит от ее размеров, так как линия представляется формулой или ее параметрами. Сложные объекты (ломаные линии, различные геометрические фигуры) представляются в виде совокупности элементарных графических объектов. Любой объект состоит из совокупности связанных линий. Это обстоятельство определило еще одно название данного явления — *объектно-ориентированная графика.* На экран компьютера изображение выводится в виде точек. При этом программа перед выводом изображения производит вычисление координат экранных точек отображаемого объекта. Аналогичные вычисления производятся и при выводе объектов на печать. В связи с этим возникло еще одно название данного метода — *вычисляемая графика.*

- Векторная графика предназначена для создания иллюстраций и широко используется в рекламном деле, дизайне, редакционно-издательском деле. Оформительские работы, основанные на применении шрифтов и простых геометрических элементов, проще выполняются с помощью векторной графики. При этом размер символов может изменяться в широких пределах. Такие шрифты называют масштабируемыми. Например, технология*True Type*, разработанная компаниями Microsoft и Apple Computer, описывает способ отображения символов в тексте.

- Векторные методы также широко применяются в автоматизированных системах проектирования (от англ. *computer+aided design*, CAD), используемых для работы со сложными трехмерными объектами. Однако векторная технология не позволяет достичь фотографического качества изображений объектов, как при использовании растровых методов. Работать с векторными рисунками можно с помощью редактора *Corel DRAW* и др. Наиболее популярны векторные форматы WMF, CDR, DXF.

- *Фрактальные графические изображения* создаются автоматически с помощью специальных математических вычислений, т.е. путем программирования, а не рисования. Фрактальная графика обычно используется в оформительских работах и развлекательных программах. Для просмотра, масштабирования и конвертирования графических файлов используются различные программы. Наиболее популярной из них считается *ACD See* компании ACD System.

- Потребность использования графиков, диаграмм, схем, рисунков, этикеток в произвольном тексте или документе вызвала необходимость создания графических процессоров. Графические процессоры представляют собой инструментальные средства, позволяющие создавать и модифицировать графические образы с использованием следующих типов ИТ: коммерческой, иллюстративной, научной и когнитивной графики.

- ИТ *коммерческой (или деловой) графики* обеспечивают отображение информации, хранящейся в табличных процессорах, базах данных и отдельных локальных файлах в виде двух- или трехмерных графиков, в частности круговой диаграммы, столбиковой гистограммы, линейных графиков и др. Они включаются в состав офисных приложений, многих интегрированных технологий и систем.

- ИТ *иллюстративной графики* позволяют создавать иллюстрации — деловые схемы, эскизы, географические карты для различных текстовых документов в виде регулярных структур, т.е. различные геометрические фигуры (векторная графика), и нерегулярных структур (растровая графика). Процессоры, реализующие иллюстративную растровую графику, дают возможность пользователю выбрать толщину и цвет линий, палитру заливки, шрифт для записи и наложения текста, включить созданные ранее графические образы.

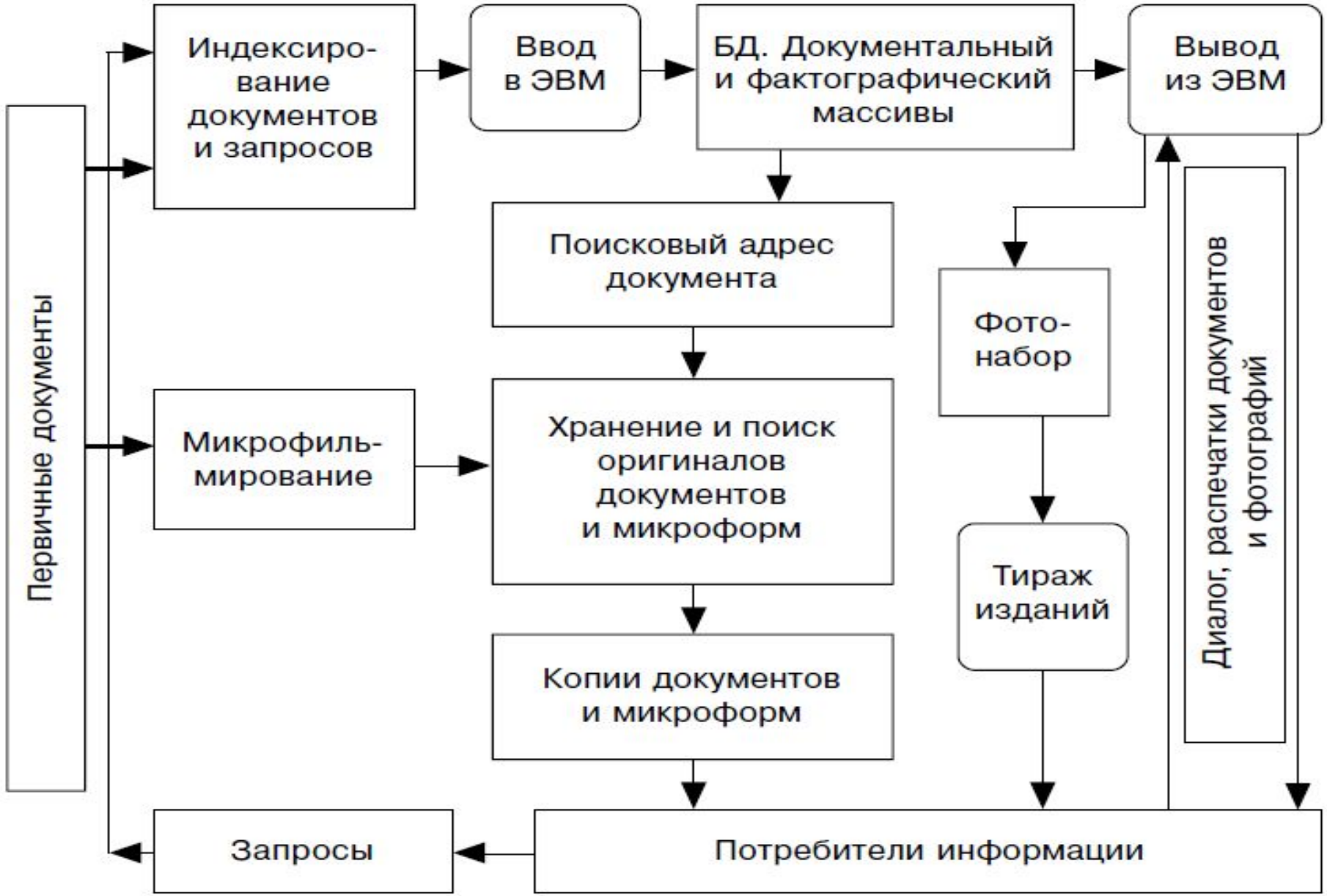
- ИТ *научной графики* предназначены для оформления научных расчетов, содержащих химические, математические и прочие формулы, а также могут быть использованы в картографии и других сферах. Для их реализации используются средства векторной и когнитивной графики.

- *Когнитивная графика* — совокупность приемов и методов образного представления условий задачи, которая позволяет сразу увидеть решение либо получить подсказку для его нахождения. Она позволяет образно представить различные математические формулы и закономерности для доказательства сложных теорем, открывает новые возможности для познания законов функционирования сознания — этой наиболее сложной и сокровенной тайны мироздания

- Когнитивные компьютерные средства представляют собой комплекс виртуальных устройств, программ и систем, реализующих совокупную обработку зрительной информации в виде образов, процессов, структур, позволяющих средствами диалога реализовать методы и приемы представления условий задачи или подсказки решения в виде зрительных образов. Когнитивная графика используется в информационном моделировании, интеллектуальных ИТ, системах поддержки принятия управленческих решений и др.

ТЕХНОЛОГИИ ИНФОРМАЦИОННОГО ПОИСКА

- В процессе решения задач пользователь довольно часто вынужден осуществлять поиск необходимой информации по соответствующему запросу. Обычно технологии информационного поиска являются основой для создания АИПС. Пример такой АИПС представлен на рисунке. На корпоративном уровне АИПС представляют собой развитую ИС с довольно сложной структурой и расширенным перечнем функций — индексирования документов, ввода их в ЭВМ, обработки, поиска, хранения, выдачи, копирования и тиражирования информации и др.


Индексирование документов и запросов
Ввод в ЭВМ
БД. Документальный и фактографический массивы
Вывод из ЭВМ
Первичные документы
Поисковый адрес документа
Фото-набор
Микрофильмирование
Хранение и поиск оригиналов документов и микроформ
Диалог, распечатки документов и фотографий
Тираж изданий
Копии документов и микроформ
Запросы
Потребители информации

- В составе диверсифицированных баз данных АИПС хранятся обширные объемы информации различного содержания и форм. Для обеспечения доступа широких категорий пользователей к корпоративным электронным информационным массивам осуществляются кооперация и интеграция этих ресурсов. Обычно развитые БД содержат разнородную информацию, в частности документальные, фактографические, аудио-, видео-, мультимедийные массивы информации.

Базы данных АИПС

- *База данных АИПС* (автоматизированных информационно-поисковых систем) — это совокупность данных, организованных по определенным признакам, имеющим общие принципы и методы описания, хранения и манипулирования данными, обеспечивающих их независимость от прикладных программ.

- По количеству форм представления данных различаются одноконтурные и многоконтурные БД. Основная форма представления БД — двухконтурная. Первый контур хранится на внешнем накопителе ЭВМ — жестком магнитном диске, магнитной ленте и др., а второй контур, как страховой, может быть представлен на флоппи и (или) CD и других носителях. Могут быть и трехконтурные БД, когда третий контур представлен и сохраняется на традиционных бумажных документах. БД АИС четвертого контура может быть представлена в форме микрофильмированной ленты и (или) ее отдельных отрезков (кляссеров)

- По характеру содержащейся информации различают фактографические, документальные и смешанные БД.
- Фактографическая БД отображает конкретные сведения, необходимые пользователю, — факты, показатели, свойства продукции, формулы расчета какой-либо величины, фрагмент текста документа, документ полностью и др. Документальная БД содержит только сведения о документах — библиографическое описание документа, аннотацию, реферат, идентификатор документа, адрес его хранения в БД и т.д.

- Сам документ, например «Методика расчета амортизации оборудования металлургического производства», хранится, как правило, во внешнем контуре БД

- При условии отсутствия дефицита внешней памяти и производительности ЭВМ документальные БД объединяют во внешней памяти ЭВМ первый и второй контуры. В смешанных базах данных представлены как фактографические, так и документальные массивы информации.

Файлы АИС

- Опорным структурным элементом БД является *файл* — поименованная область внешней памяти ЭВМ, содержащая данные. Файл может содержать различные данные: текстовый документ, рисунок, музыкальное произведение, программу ЭВМ и др. Каждый файл записывается и хранится во внешней памяти ЭВМ и имеет собственное имя, идентифицирующее его в комплексе файлов, находящихся в БД.

- *Таблица* — способ формализованного представления данных в виде двумерного массива. Таблица состоит из строк и столбцов. Строки таблицы обозначаются записями. *Запись* — это единица обмена данными между программой и внешней памятью ЭВМ. Запись может содержать данные о различных объектах — отдельном человеке, например Иванов И.И., устройстве, процессе и др. Записи состоят из полей, содержащих отдельные данные об объекте, его отдельном экземпляре.

- Элементом записи является поле. *Поле записи* — часть записи файла, имеющая функционально самостоятельное значение и обрабатываемая в программе как отдельный элемент данных. Поле представляется данным, которое описывает какой-либо определенный аспект (атрибут) объекта. Поля имеют имена. При представлении файла в виде таблицы столбцам соответствуют атрибуты или поля, а строкам — записи или объекты. Поля определяют свойства, характеристики, признаки, атрибуты объектов, например год рождения человека, его пол, профессию и др. В таблице каждый столбец относится к определенному полю записи.

- Для обеспечения доступа к записям и поиска нужной информации в БД, размещенной в памяти ЭВМ, применяются так называемые ключи. *Ключ* — это совокупность знаков, используемая для идентификации записи в файле и быстрого доступа к ней. Ключ представляет уникальный номер записи в БД, ее фрагменте, файле, присваиваемый каждой записи при ее загрузке в БД. Существует несколько типов ключей. По характеру выполняемых функций различают основной (первичный), неосновной (вторичный), исходный (внешний) ключи. Первичный ключ — это ключ, который однозначно идентифицирует запись в файле. Вторичный ключ в файле вместе с несколькими ключами не обеспечивает (в отличие от основного ключа) однозначной идентификации записи в целом. В этот ключ могут входить все поля записи, кроме полей, составляющих основной ключ.

- Исходный, или внешний, ключ отображает значение ключевого поля записи, уникально идентифицирующее ее в файле. Поле ключа и его значение определяет лицо, создающее файл. В качестве этого ключа может быть задействовано более одного поля записи. Значения ключей расположены в специально предназначенных для этого полях записи.
- Ключи служат для индексной организации данных в форме индексированных файлов. *Индексированный файл* — это файл, снабженный системой индексов, обеспечивающей быстрый доступ к записям файла. В зависимости от применяемых ключей различают файлы прямого и последовательного доступа.

- *Прямой файл* — это вид файла, доступ к записям которого осуществляется по адресу либо последовательно путем поиска по ключу. В плане прямого доступа следует указать индексный файл, упорядоченный по значениям одного или нескольких полей базы данных, список указателей на ее записи. В качестве указателей здесь используются логические или физические адреса. Логическим адресом записи является внутренний номер — ключ БД, уникально идентифицирующий запись в БД или ее фрагменте (области), который присваивается записи в процессе ее загрузки в БД. Физическим адресом выступает число, идентифицирующее ячейку или область физической памяти ЭВМ.

- Смежным с индексным файлом является индексно-последовательный файл, у которого каждая запись снабжена своим ключом. Этим обеспечиваются прямой доступ к записи по ключу, а также последовательный доступ в соответствии с упорядоченностью записей по ключам. Применяются также связанные файлы, записи которых объединены в цепной список. Для организации этого списка могут быть применены индексно-последовательный и прямой файлы.

- Противоположным прямому файлу по доступу является *последовательный файл* — файл, к записям которого доступ обеспечивается только посредством последовательного чтения записей. В данном случае обращение к записи и ее идентификация производятся в соответствии с упорядоченностью этих записей.

- Отметим также *инвентированный файл* — файл, в котором записи упорядочены по неключевому полю. Кроме того, существует полностью инвентированный файл, который имеет индексы по всем вторичным ключам, и частично инвентированный файл, имеющий индексы по отдельным, но не по всем вторичным ключам.

- Близким по значению записи является *агрегат* — структурированная совокупность информационных объектов, определяемая как единый тип данных. В моделях данных агрегаты обозначаются как вид абстракции, когда программа рассматривается как сложная система, представляющая совокупность конечного числа объектов. Это позволяет описывать программы до этапа непосредственного программирования, т.е. на уровне алгоритмов.

- В структуре БД, кроме файлов, присутствуют и другие единицы информации, например массивы информации. *Массив информации* — это поименованная совокупность однотипных (логически однородных) элементов, упорядоченных по индексам, определяющих положение элементов в массиве. Элементами массива могут быть документы, файлы, записи и др. Одним из весомых параметров массива является его измерение, которое можно обозначить как градацию размерности массива. Такими градациями могут быть одномерные массивы, имеющие одно измерение, например запись файла, двумерные массивы, например строки и столбцы таблицы и др. Индексы играют важную роль в организации данных. Они позиционируют элементы и указывают его адрес в массиве. Так, индекс может выступать в роли адресной константы, используемой для модификации адреса путем суммирования ее значения с вычисленным в программе адресом.

- БД имеют определенные способы построения. Эти способы построения определяются моделями БД: иерархические, сетевые, реляционные, объектно-ориентированные и эволюционные. Модель БД определяет принципиальный способ представления данных в памяти ЭВМ

- Иерархическая модель БД. Построена по принципу древовидного графа, в котором информационные элементы представлены по уровням их соподчиненности (иерархии). Например, на первом уровне расположены сведения об объекте «конкуренты», на втором уровне — о продукции, которую они поставляют на рынок, на третьем уровне — цена продукции и т.д. Таким образом, в структуре иерархии каждый порожденный узел не может иметь более одного порождающего (выходного) узла. Корень дерева является здесь не порожденным, а порождающим узлом. Узлы, не имеющие выхода, носят названия листьев.

- При поиске необходимых данных происходит чтение записей от корня к листьям дерева, т.е. снизу вверх. Достоинством является то, что подобная структура БД обеспечивает более быстрый доступ и выдачу данных пользователю. Вместе с тем недостатком представляется жесткость иерархической структуры. Отсутствует информационная гибкость в поиске, так как за один проход невозможно получить данные, например, о ценах одного товара разных поставщиков. В иерархической модели реализована связь между данными по схеме «один ко многим». При необходимости реализации связи типа «многие ко многим» следует строить сетевую модель БД.

- Сетевая модель БД. Имеет независимые типы данных, т.е. «конкуренты», и зависимые типы данных — продукция и цены на продукцию. В сетевых моделях возможны любые виды связей между данными (записями), как прямые, так и обратные. Существует ограничение — каждая связь должна включать основную и зависимую записи. К достоинству сетевой модели можно отнести более развитую гибкость организации и доступа к данным относительно иерархической модели. Как недостаток можно указать, что сохраняется относительная жесткость в построении структуры БД. Это влечет необходимость в определенных условиях реструктурирования БД, препятствует реализации гибкой стратегии поиска данных. Более совершенной моделью является реляционная БД.

- Реляционная модель БД. Этот класс БД строится на применении отношения типа «сущность — связь». В основе модели лежит деление реального мира на отдельные различимые сущности. Сущности находятся в определенной связи друг с другом. Эта БД имеет независимую организацию взаимосвязи логических и физических записей. Отношения между данными построены в виде двумерных таблиц. Отношения наделены определенными признаками. Каждый элемент таблицы отображает одно данное. Элементы столбца таблицы имеют одинаковую природу, отображая одно свойство (признак) в строке (записи) таблицы. Пример — таблица, в которой имеются столбцы (фамилия, имя, год рождения, место работы, домашний адрес), и строки, куда записываются сведения о работниках фирмы. Эти данные составляют ядро реляционной БД.

- Достоинства реляционной модели объясняются тем, что в ее основе лежит строгий аппарат реляционной алгебры. В этой модели реализованы простота доступа к данным, гибкость поиска и защиты данных, независимость данных, относительная простота построения языка манипулирования данными. Язык запроса в соответствии с реляционной алгеброй имеет следующие основные операции: проекция, соединение, пересечение и объединение. Язык позволяет описывать характер поиска данных без указания последовательности действий, необходимых для получения ответа на запрос.

- Качество реляционных БД зависит от нормализации данных, придания им нормальной формы. *Нормальная форма* — требование, предъявляемое к структуре таблиц для устранения из БД избыточных функциональных зависимостей между атрибутами (полями таблиц). Понятие нормальной формы было введено Э. Коддом — создателем реляционной модели БД. Процесс преобразования БД к виду, отвечающему нормальным формам, называется *нормализацией.* Нормализация позволяет обезопасить БД от логических и структурных проблем, называемых аномалиями данных.

- Например, когда существует несколько одинаковых записей в таблице, есть риск нарушения *целостности данных* при обновлении таблицы. Таблица, прошедшая нормализацию, менее подвержена таким проблемам, так как ее структура предполагает определение связей между данными, что исключает необходимость в существовании записей с повторяющейся информацией.

- Модель объектно-ориентированной БД (ООБД). Эта модель является примером реализации БД более высокого логического уровня. ООБД возникли на концептуальной основе объектно-ориентированного программирования (ООП). В отличие от структурного ООП базируется не на процедурных (программных) категориях (циклы, декларации, условия и др.), а на более широкой категории — объектах. Объектом можно объявить все, что представляет интерес для обработки данных на ЭВМ, — завод, подразделение, рабочий, программа ЭВМ, запись БД, пиктограмма экранного окна и т.д.

- Объект — программно связанный набор процедур, методов и средств, реализующих определенную задачу. Процедура — это совокупность операций, которые может выполнять объект. Метод — это способ, прием, которым пользуется объект при выполнении процедур. Свойство — это характеристика (признак), с помощью которой описывается объект. Например, работник как объект обладает свойствами, характеризующими его функциональные способности, технологическими процедурами, выполняемыми им в процессе трудовой деятельности, и методами, посредством которых реализуются технологические процедуры.

- Организация ООБД имеет несколько стадий:
- концептуальная модель, когда множество объектов БД прошли описание по соответствующим правилам;
- логическая модель, когда определены свойства объектов и указаны логические взаимосвязи объектов;
- физическая модель, когда определены адреса и проведено размещение объектов в памяти ЭВМ.

- Модель эволюционных БД. Кардинальным направлением улучшения качества удовлетворения информационных потребностей является поиск новых путей развития БД и электронных хранилищ информации. Ведутся поиски нового подхода к организации БД. Перспективным представляются эволюционные БД. Они базируются на модели представления данных так называемого миварного пространства. Здесь предлагается новый подход к созданию модели представления данных и знаний в системах искусственного интеллекта. Эта модель основана на использовании динамического многомерного объектно-системного дискретного пространства представления данных и правил. Основой миварного подхода является такое описание предметной области, при котором сущности (объекты), свойства и отношения могут переходить в друг друга в зависимости от предмета изучения.

- Таким образом, сущность может быть свойством другой сущности или сущность может быть отношением других сущностей, и наоборот. Наименьшим элементом пространства данных здесь является мивар. Каждый мивар представлен в базе данных в системе декартова трехмерного пространства. Проведенные эксперименты программной реализации миварной базы данных показали перспективность этого подхода.

- Лингвистические средства. В реализации технологии информационного поиска применяются *лингвистические средства* — совокупность информационно-поисковых языков, методик индексирования и критерия смыслового соответствия. В составе лингвистических средств содержатся следующие компоненты.
- Информационно-поисковые языки.
- Методики индексирования документов.
- Форматы документов с указанием структуры их «шапок» и «боковиков».
- Критерий смыслового соответствия или критерий выдачи документов и (или) поисковых образов документов (ПОД) по различным классам документальной информации, содержащейся в БД.

- В решении задач информационного поиска связующим звеном между пользователем и ЭВМ является информационно-поисковый язык (ИПЯ). *Информационно-поисковый язык* — это упорядоченное множество понятий, терминов определенной предметной области, предназначенное для отображения содержания документов и запросов с целью обеспечения ввода документов и запросов в ЭВМ и осуществления последующего поиска данных. Словарной единицей ИПЯ является ключевое слово, которое может быть как отдельным словом, так и словосочетанием.

- При условии устранения неоднозначности (омонимии) отдельных слов ключевые слова обозначаются как дескрипторы ИПЯ. Посредством ИПЯ в технологии обработки осуществляется индексирование документов и запросов. *Индексирование* — это совокупность логических операций по отображению содержания документов и запросов средствами ИПЯ.
- По уровню применения технических средств индексирование бывает ручное и автоматическое. При ручном индексировании процессы анализа документов и запросов выполняются без применения ЭВМ. При автоматическом индексировании ЭВМ выполняет функции анализа текстов документов и запросов, определения их значимости (весомости) и формирования состава дескрипторов поисковых образов документов и запросов.

- При автоматическом индексировании ЭВМ поручаются функции дериватного, приписного индексирования и автоматической классификации. Так, дери- ватное индексирование, или индексирование извлечением, представляет собой метод автоматического индексирования документов, при котором система анализирует лексический состав текстов и выбирает из них те слова и их сочетания, которые удовлетворяют заданным критериям. Одним из таких критериев может быть критерий поиска, рассматриваемый ниже. Программы автоматического индексирования довольно сложны и относятся к высокоинтеллектуальным продуктам. Автоматическое индексирование имеет относительно высокую стоимость и применяется в АИС, где это функционально и экономически оправдано.

- В результате индексирования получаются поисковые образы документов и поисковые образы запросов. *Поисковый образ документа (ПОД)* — это совокупность ключевых слов, кодовых обозначений, отображающих содержание документа, адрес хранения и его системный номер (идентификатор). *Поисковый образ запроса (ПОЗ)* — это совокупность ключевых слов, отображающих содержание запроса и условия поиска документов. Следует различать ИПЯ классификационного и дескрипторного типов. Наибольший удельный вес в задачах поиска занимают ИПЯ классификационного типа — классификаторы и кодификаторы

- Сюда входят классификаторы категорий, по которым строится принципиальная схема управления объектом. В современных технологиях поиска в основном базовым ИПЯ являются классификаторы. *Классификатор* — это систематизированная совокупность наименований и кодов языковых элементов определенной предметной области. Классификаторы строятся по иерархическому принципу. Исходное множество элементов делится на группировки следующего уровня деления и образуют древовидную систему группировок. Для выделения группировок применяется соответствующий признак (основание) деления. Каждому элементу классификатора по принципу однозначного соответствия проставляется код. Код может быть цифровым, буквенным, комбинированным (алфавитно-цифровым).

- По применяемому способу кодирования классификаторы имеют следующие основные разновидности:
- десятичные классификации;
- библиотечно-библиографические классификации;
- фасетные классификации.

- В десятичных классификациях множество объектов делится на десять частей, каждая из которых в свою очередь также делится на десять частей и т.д. Представителем десятичной классификации является универсальная десятичная классификация (УДК). УДК в современной информатике является международной классификацией, охватывающей все отрасли знаний, в том числе и экономику, строится по десятичному принципу и используется в современных ИС для индексирования документов и их последующего поиска в справочно-информационных фондах, автоматизированных БД и др. УДК впервые была введена в СССР в качестве обязательного классификатора в 1962 г. для индексирования документов по естественным, точным и техническим наукам

- Библиотечно-библиографическая классификация (ББК) основана на порядке следования букв в том или ином алфавите. В России действует ББК, разработанная Всероссийской государственной библиотекой. Вся область знаний разбита на количество разделов (подобластей), равное количеству букв русского алфавита, за исключением «неудобных», в частности «ь», «ъ», «й». На втором и последующем уровнях деления каждый уровень также разделяется на такое же количество подразделов и т.д. В системе научной информации ББК применяется с 1962 г. для обозначения документов по гуманитарным областям знаний. В научно-технических библиотеках для систематизации и кодирования определенной группы документов применяются обе системы классификации

- Фасетная классификация (ФК) является разновидностью системы классификации, в которой реализована возможность классификации объектов параллельно по нескольким различным признакам. Возможность параллельной классификации не означает принципиальную обязательность этого условия для всех ФК. Так, поставщик комплектующих изделий при определенных условиях может быть идентифицирован как кредитор и (или) дебитор. Элементарной (неделимой) рубрикой в ФК является так называемый фокус, например «склад холодильников»

- В соответствии с принципом ФК в России применяется Единая система классификации и кодирования (ЕСКК).
- 1 Общегосударственные классификаторы разрабатываются в централизованном порядке и являются едиными и обязательными для применения по всей стране.
- 2 Отраслевые классификаторы разрабатываются соответствующими отраслями для решения задач. Они в определенных случаях могут быть задействованы и в АИС других отраслей, например шифры счетов бухгалтерского учета применяются во многих отраслях деятельности.
- 3 Локальные классификаторы разрабатываются предприятиями на номенклатуры, относящиеся только к данному предприятию, например коды предоставляемых услуг, коды (табельные номера) сотрудников и др.

Отраслевые классификаторы

- разрабатываются соответствующими отраслями для решения задач. Они в определенных случаях могут быть задействованы и в АИС других отраслей, например шифры счетов бухгалтерского учета применяются во многих отраслях деятельности.

Локальные классификаторы

- разрабатываются предприятиями на номенклатуры, относящиеся только к данному предприятию, например коды предоставляемых услуг, коды (табельные номера) сотрудников и др

- В начале 1970-х гг. для лингвистического обеспечения АИС в стране начали разрабатываться общегосударственные классификаторы. В настоящее время насчитывается свыше сорока классификаторов такого уровня.
- Систему классификаторов можно разделить на следующие разновидности.

Классификаторы структуры отраслей народного хозяйства

- В эту группу входят, например, Общегосударственный классификатор отраслей народного хозяйства (ОКОНХ), Система обозначений органов государственного управления (СООГУ), Система обозначений административно-территориальных объектов (СОАТО), Общегосударственный классификатор предприятий и организаций (ОКПО), Общегосударственный классификатор форм собственности (ОКФС).

- Классификаторы продукции, например, Общегосударственный классификатор промышленной и сельскохозяйственной продукции (ОКП)

- Классификаторы ресурсов, например, Общегосударственный классификатор профессий рабочих, должностей служащих и тарифных разрядов (ОКПДТР)

- Классификаторы информационных единиц, например, Общегосударственный классификатор технико-экономических показателей (ОКТЭП), Общегосударственный классификатор управленческой документации (ОКУД)

- Эффективность автоматизированной обработки информации требует предварительного представления информации в удобной и компактной форме, что достигается в процессе ее кодирования посредством применения определенных кодов. *Код* — это элемент системы условных обозначений объекта или элементов информационной совокупности в виде знака или группы знаков, выраженных цифрами, буквами, символами и различными сигналами.
- Процесс присвоения объектам кодовых обозначений называется кодированием. Основная цель кодирования состоит в однозначном обозначении объектов, а также в обеспечении необходимой достоверности кодируемой информации. При проектировании кодов к ним предъявляется ряд требований:

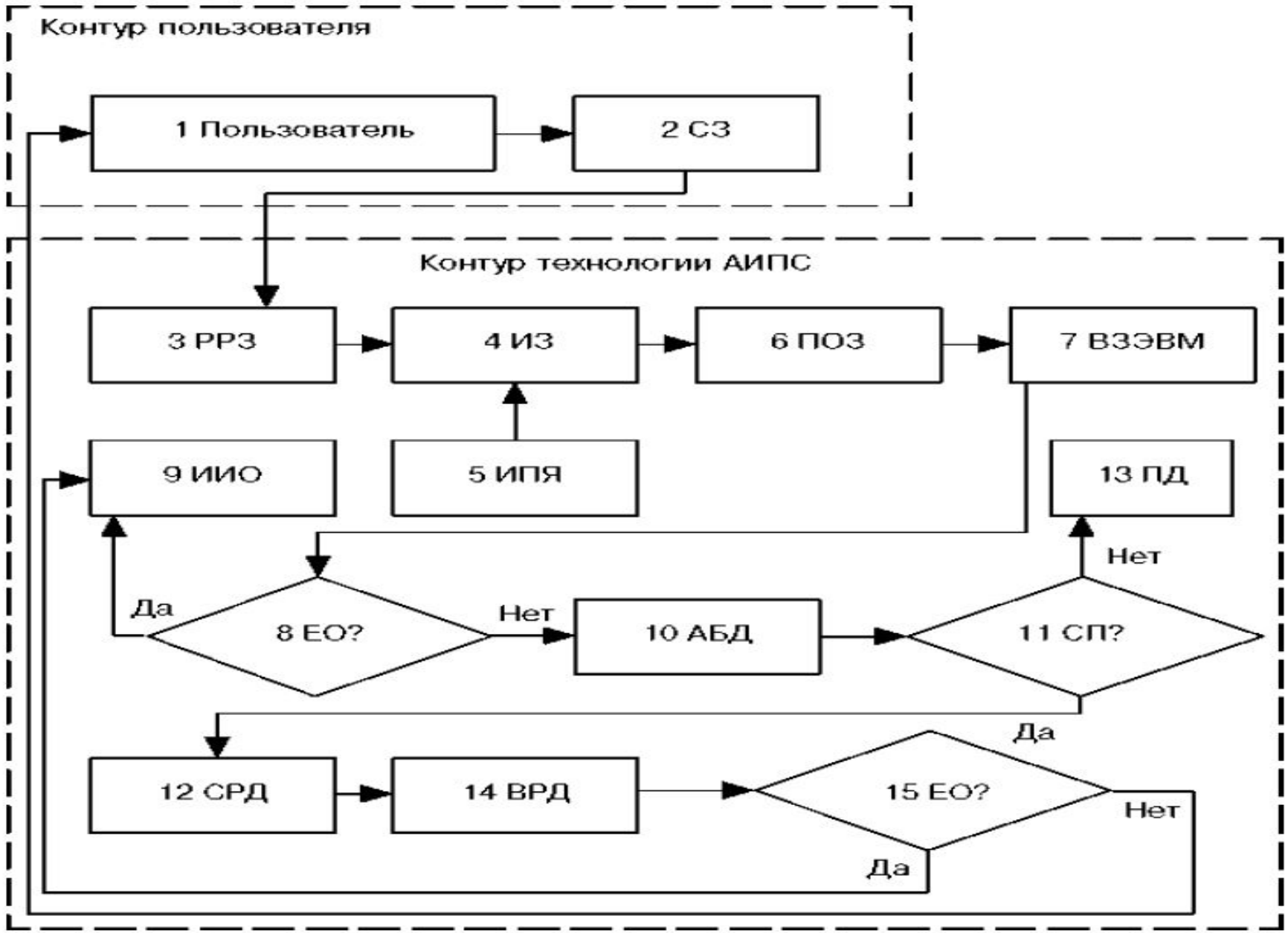
- охват всех объектов, подлежащих кодированию, и их однозначное обозначение;
- возможность расширения объектов кодирования без изменения правил их обозначения;
- удобство восприятия и запоминания кодовых обозначений индексировщиком, обеспечивающее простоту заполнения, чтения и обработки статистического отчета;
- максимальная информативность кода при его минимальном формате (значности) с целью эффективной обработки информации;
- возможность использования кодов для автоматического получения сводных итогов;
- возможность автоматического контроля кодовых обозначений с целью обнаружения ошибок.

- Для организации документальных БД и реализации поиска в них предназначены дескрипторные языки. *Дескрипторный язык* АИПС — это разновидность ИПЯ (информационно-поискового языка), применяемого в АИПС для поиска необходимых документов по тематике, связанной с решением экономических задач. Методика их разработки существенно не отличается от методики дескрипторных языков и основана на идее координатного индексирования.

- В основе координатного индексирования лежит представление о том, что содержание любого документа или текста можно отобразить с достаточной степенью полноты и точности набором так называемых ключевых слов. Ключевое слово в среде дескрипторного ИПЯ понимается как наиболее существенные для этой цели понятия, термины, словосочетания, имена собственные, хронологические данные, величины измерения и т.п., которые явно или в скрытом виде содержатся в индексируемом документе.

- Поиск нужных документов АИПС может осуществляться при условии обеспечения единообразия индексирования документов и запросов. Существенными компонентами лингвистических средств являются методы индексирования. В рамках АИПС индексированию подвергаются документы, вводимые в БД, и запросы на поиск данных, а также входные документы АИПС, содержащие документальную и фактографическую информацию. Аналитико-синтетическая переработка документов, содержащих документальную информацию, строится по традиционной схеме путем выделения в документах формальных и содержательных признаков, ключевых слов дескрипторного языка и фиксирования их в формате ПОД (поисковый образ документа)

- Индексирование документов, содержащих фактографическую информацию, выполняется посредством применения в основном языков классификационного типа. Каждая классификационная рубрика (реквизит-признак) снабжается соответствующим шифром (кодом) классификатора. Как правило, коды классификационных рубрик фиксируются в специальных субполях на титульной странице документа — баланса, ведомости и др. Отметим, что процедуры индексирования документов могут быть реализованы традиционным (ручным) и автоматическим способами. Автоматическое индексирование, например, фактографической информации может выполняться на этапе ввода в ЭВМ, распознавания лексем рубрик классификаторов и последующего определения кода соответствующего реквизита входного документа.
- Технология прохождения запроса на поиск информации отображена в виде принципиальной схемы функционирования АИПС


Контур пользователя
1 Пользователь
2 СЗ
Контур технологии АИПС
3 РРЗ
4 ИЗ
6 ПОЗ
7 ВЗЭВМ
9 ИИО
5 ИПЯ
13 ПД
Нет
Да
8 ЕО?
Нет
10 АБД
11 СП?
Да
12 СРД
14 ВРД
15 ЕО?
Нет
Да

Обозначения: СЗ — составление запроса, РРЗ — регистрация и редактирование запроса, ИЗ — индексирование запроса, ИПЯ — информационно-поисковый язык, ПОЗ — поисковый образ запроса, ВЗЭВМ — ввод запроса в ЭВМ, ЕО? — есть ошибки?, ИИО — идентификация и исправление ошибок, АБД — анализ базы данных, СП — совпадение признаков ПОЗ и ПОД, СРД — селекция релевантных документов, ПД— пропуск документа, ВРД— вывод релевантных документов.

- С целью получения необходимых документов или данных пользователь составляет запрос в произвольной форме на естественном языке (блок 2). Свой запрос он направляет в контур АИПС в службу индексирования. Его запрос регистрируется и редактируется (блок 3). Запрос редактируется в части расширения его полноты и уточнения содержания с позиций предметной области решаемой экономической задачи. Затем запрос индексируется, т.е. проводится аналитико-синтетическая переработка текста запроса (блок 4).
- Индексирование запроса состоит в анализе набора ключевых слов запроса и унификации этого набора в виде дескрипторов соответствующего дескрипторного ИПЯ (блок 5). Затем в этом блоке проводится перенос результатов индексирования запроса на специальный формат (бланк) — «поисковое предписание», или ПОЗ (блок 6). В этом предписании содержатся дескрипторы, а также логические правила, по которым будет выполняться поиск нужной информации. Затем ПОЗ передается на участок ввода документов в ЭВМ. В зависимости от режимов загрузки ЭВМ этот ПОЗ обычно вводится сразу или по мере накопления комплекта определенного объема (блок 7)

- При вводе в ЭВМ проводятся лексический, синтаксический, логический и арифметический виды контроля в зависимости от характера ПОЗ (блок 8). При обнаружении ошибки проводятся ее идентификация и исправление (блок 9).
- Сообщения об ошибке оператору могут быть выведены на дисплей или принтер в форме протокола ввода ПОЗ в ЭВМ. При отсутствии замеченных ошибок проводится вызов программы поиска. Эта программа проводит обращение к структурированной БД и последующий анализ документов (блок 10).

- Технология поиска информации по запросу в значительной мере определяется характером БД. При поиске в документальной БД проводится сравнение дескрипторов поисковых образов документов (ПОД) с дескрипторами ПОЗ на предмет их совпадения (блок 11). При условии совпадения определенного количества дескрипторов ПОД и ПОЗ программа проводит селекцию ПОД из базы данных в отдельный массив ПОД (блок 12). Программа проводит селекцию при условии релевантности ПОД. Программа поиска определяет релевантность на основе заданного критерия смыслового соответствия. Таким образом, обеспечиваются полнота и точность выданных документов по запросу. В определенной мере полноту и точность при поиске обеспечивают условия логики поиска, задаваемые в ПОЗ. Так, из списка релевантных ПОД могут быть изъяты документы, содержащие информацию о матричных принтерах, если в ПОЗ стоит логическая метка, обозначающая условие «кроме матричных принтеров». При условии несовпадения программа пропускает ПОД и переходит к анализу следующего ПОД в БД (блок 13)

- Поиск в фактографической БД проходит несколько иначе. Анализу подвергается каждая запись документа, содержащая идентификаторы в соответствии с ИПЯ задач, базирующихся на фактографической информации, т.е. здесь имеется в виду ИПЯ классификационного типа — кодификаторы, словари, справочники. При условии совпадения идентификаторов (кодов) искомой единицы информации (показателя, записи, документа и т.д.) и идентификаторов ПОЗ программа проводит селекцию релевантной единицы фактографической информации. Следует отметить, что селекция единицы информации проводится при условии только полного совпадения единицы информации и ПОЗ. Таким образом, применение критериев выдачи, полноты и точности относительно поиска фактографических единиц информации теряет смысл.

- Результаты поиска могут быть выданы на видеотерминал или распечатаны на принтере (блок 14). Затем проводится контроль качества поиска и распечатки результатов (блок 15). При условии низкого качества поиска как по содержанию, так и по оформлению проводятся идентификация и исправление дефектов поиска (блок 9). Если дефектов не обнаружено, то результаты поиска передаются пользователю (блок 1). Получив результаты поиска, пользователь проводит анализ и оценку качества поиска. При условии удовлетворительного качества он применяет полученную информацию в процессе решения соответствующей задачи. Если же результаты поиска не удовлетворяют пользователя, то в данном случае он корректирует свой запрос и процедура поиска повторяется. Подобные технологические итерации могут продолжаться до тех пор, пока пользователь не будет полностью удовлетворен результатами поиска.

- Эффективный способ повышения релевантности поиска — использование так называемого языка запросов. С помощью языка запросов можно конкретизировать запрос к поисковой системе. Каждая поисковая система имеет собственный язык запросов, однако существуют универсальные конструкции, «понятные» большинству поисковых систем. Как правило, в большинстве решений АИПС используются следующие символы: «*», «-», «&», «|», «~», «(», «)». Для так называемого поиска по маске используются символы «*» и «?». Знаком «?» в ключевом слове запроса заменяют один символ, на место которого может быть подставлена любая произвольная буква, а знаком «*» — целая цепочка символов произвольной длины. Символы «+» и «-», введенные без пробела перед каким-либо ключевым словом, позволяют обеспечить его обязательные присутствие или отсутствие в найденном документе соответственно.

- каждая поисковая система «понимает» свой язык запросов, поэтому пользователям нужно помнить, что инструкцию по использованию языка запросов всегда можно найти в справочном разделе каждой поисковой системы. Чтобы научиться формулировать грамотные запросы, необходимо изучать документацию АИПС.

- К наиболее известным в мире системам веб-поиска относятся *Google, Alta Vista, Norbern Light, Yahoo!, Magellan, Excite, HotBot, Infoseek (Go), Lycos, Open Text, Web Crawler, WWW Worm.* Основными полнотекстовыми поисковыми системами, представленными в русскоязычном секторе Интернета, считаются *Yandex, Rambler, Aport, List.ru, Russia on the Net, FTP-Search.*

- По размерам баз данных (индексов) российские поисковые системы заметно проигрывают известным зарубежным, однако здесь собрана и классифицирована обширная информация на русском языке. Масштабы информационных ресурсов и их количество в Интернете огромными темпами расширяются, и централизованная база данных поисковых машин на сегодняшний день уже не является удовлетворительной. Кроме того, несмотря на то что работа поисковых машин считается вполне успешной, все современные системы имеют серьезные недостатки:

- поиск по ключевым словам дает слишком много ссылок, и многие из них оказываются бесполезными;
- огромное количество поисковых машин с разными пользовательскими интерфейсами порождает у пользователя проблему информационной перегрузки;
- методы индексирования баз данных, как правило, семантически не связаны с их информационным содержанием; неадекватные стратегии поддержки каталогов часто приводят к тому, что пользователю выдаются ссылки на информацию, которой уже нет в Интернете;

- поисковые машины еще не столь совершенны, чтобы понимать естественный язык;
- по тому представлению результатов, которое обеспечивают современные поисковые машины, невозможно сделать логически обоснованный вывод о полезности каждого источника.

- В последнее время потребности в интеллектуальной помощи быстро растут, помощь необходима для продуктивного поиска информации, для нахождения в необъятном Интернете или корпоративной сети специализированной информации. Это привело к появлению интеллектуальных агентов. Как правило, интеллектуальные агенты, в которых для поиска и сортировки информации используются технологии искусственного интеллекта, являются составной частью поисковой машины и способны «понимать», какая именно информация нужна пользователю (пользователь предварительно обучает агента). Получить информацию по теме «Интеллектуальные агенты в Интернете», касающуюся конкретных примеров интеллектуальных агентов, можно, например, на портале сообщества экспертов *Report.ru*.