

Лекция 5

Информационное обеспечение компьютерных технологий

Содержание лекции

- Основные понятия и положения информационного обеспечения архитектурной деятельности.
- Понятия информационного файла, библиотеки, базы данных, базы знаний.
- Простейшие информационные системы: документальные и фактографические. Особенности поиска данных.
- Простейшие структуры данных.
- Модели данных, особенности их организации и поиска в них данных.
- Процесс создания баз данных.

5.1. Основные понятия и положения информационного обеспечения архитектурной деятельности

Информационным обеспечением любого процесса является совокупность сведений, представленных в заданной форме, необходимых для автоматизированного выполнения этого процесса. Основу информационного обеспечения архитектурной деятельности составляют данные, необходимые для получения результатов этой деятельности.

Данные - это описание явления или процесса, представленное в определенном формализованном виде. Данные соответствуют дискретным зарегистрированным фактам относительно явлений.

Информационное обеспечение решения архитектурных задач с помощью компьютеров предназначено для организации, использования (получения) и поддержания в актуальном и корректном состоянии всех данных, необходимых для решения профессиональной (проектной, административно-управленческой, учебной, просветительской) задачи.

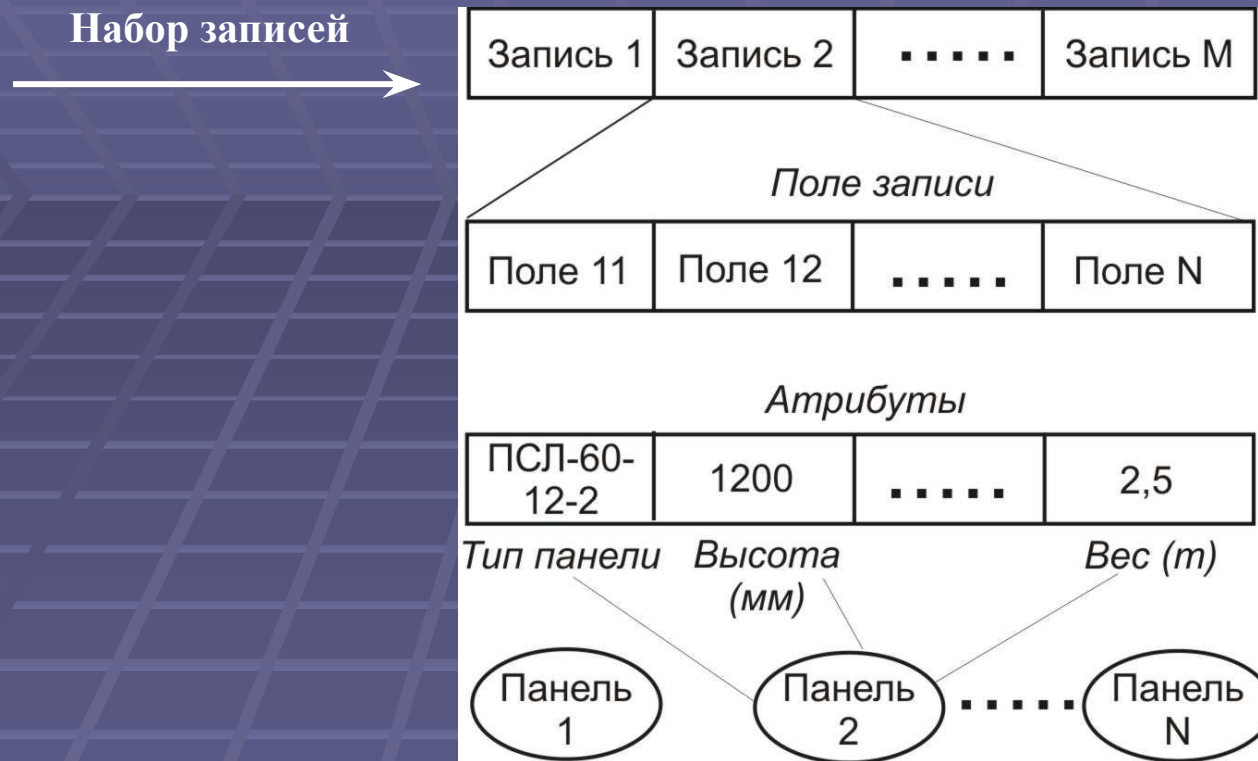
В информационном плане любой рассматриваемый объект называется *сущностью*.

Любая сущность обязательно характеризуется *атрибутами*. Атрибуты могут принимать численные (например, длина и ширина окна, его площадь), буквенные (архитектурный стиль объекта, материал стен, фамилия автора проекта), логические (да/нет; присутствует/отсутствует; имеет перестройки или нет) и значения типа «дата» (например, год постройки или реконструкции, даты жизни автора и пр.).

Основными структурными элементами информационного фонда являются записи и наборы записей.

Запись состоит из отдельных полей, в которые проставляются значения атрибутов сущности (объекта).

Соответствие атрибутов и полей записей



Набор записей представляет собой совокупность записей об объекте для всех экземпляров объектов одного типа.

Информационный файл - совокупность наборов записей, соответствующих множеству объектов проектирования (или множеству частей объекта).

Компоненты *внутримашинной информации* можно *классифицировать* по организации и по носителям записи.

По организации компонентов выделяют файлы (или наборы данных), библиотеки, базы данных, базы знаний, и интегрированные банки данных и знаний, геоинформационные системы.

Библиотека представляет собой каталогизированную совокупность информационных компонентов (текстовых и /или графических данных, решений, чертежей и т.п.), в которой организован простейший поиск объекта (с помощью системы меню). Библиотеки легко дополняются, видоизменяются; их структура не зависит от базового графического пакета.

База данных (БД) - это хранилище специально организованных и логически взаимосвязанных информационных элементов (файлов), отображающих состояние объектов и их отношений в рассматриваемой предметной области.

Функционирование БД обеспечивает **система управления базой данных (СУБД)**, основными функциями которой являются: определение данных, подлежащих хранению (определение структур данных и структур хранения информации); первоначальная загрузка информации в БД - так называемое создание БД; обновление данных; доступ к данным по различным запросам пользователей, отбор и извлечение некоторой части БД, редактирование извлеченной части и выдача результатов пользователю, обеспечение секретности - защиты от несанкционированного доступа и целостности данных - защиты от непредсказуемого воздействия, а также от сбоев оборудования.

База знаний (БЗ) - совокупность систематизированных основополагающих сведений, относящихся к определенной области знания, хранящихся в памяти компьютера, объем которых необходим и достаточен для решения заданного круга теоретических и практических задач.

База знаний является содержательной частью *банка знаний* (системы баз знаний), который следует рассматривать как хранилище специальных знаний в совокупности со средствами, с помощью которых осуществляется их накопление, сохранение, обновление и использование, а также со средствами управления всеми этими процессами. В системе управления базой знаний используются методы искусственного интеллекта, специальные языки описания знаний, интеллектуальный интерфейс.

Экспертными системами называются базы знаний, позволяющие получать, накапливать, корректировать знания (в основном экспертов) из определенной предметной области, выводить новые знания, решать на основе этих знаний практические задачи и объяснять их решения.

Интегрированный банк данных (ИБнД) человеко-машинная система, состоящая из БД, БЗ, средств поиска решений и сервиса, осуществляющего взаимосвязь между пользователем и элементами банка данных.

Для ИБнД необходимы внешние и внутренние пользователи.

Внешние (конечные) пользователи являются основными потребителями информации. *Внутренние* пользователи — это по существу обслуживающий персонал ИБнД.

5.2. Простейшие автоматизированные информационные системы

Информационными системами (ИС) называются системы обработки данных о какой-либо предметной области со средствами накопления, хранения, обновления, поиска и выдачи данных. По средствам выполнения информационной задачи различают информационные системы: традиционные и автоматизированные.

Традиционные ИС используются только для просмотра информации в БД. В них не используется автоматизированный поиск данных, а полученная информация применяется для решения задачи традиционным методом (без применения компьютера).

Автоматизированные информационные системы (АИС) используют компьютеры на этапах ввода, обработки и выдачи информации по различным запросам потребителей.

АИС делятся на две основные группы:

- системы информационного обеспечения (являются важнейшими компонентами САПР, ВМ-систем и автоматизированных систем научных исследований (АСНИ));
- системы, имеющие самостоятельное целевое назначение и область применения (информационно-поисковые (ИПС) и информационно-справочные (ИСС)).

Информационно-поисковые системы представляют собой совокупность языковых, алгоритмических и технических средств, предназначенных для хранения, поиска и выдачи необходимой информации. ИПС позволяет получить из входящих в ее состав БД и/или БЗ данные по запросам пользователей (системы, позволяющие получить рекомендации в зависимости от конкретных проектных условий).

Информационно-справочные системы предназначены только для получения справочной информации (Консультант, Гарант).

В зависимости от режима организации поиска ИПС и ИСС разделяются на фактографические, документальные и документально-фактографические.

БД *фактографических* ИПС и ИСС состояются из формализованных (форматированных) записей.

БД *документальных* ИПС и ИСС могут иметь записи, содержащие различные информационные документы (статьи, письма, сообщения, бланки и т.п.).

Документально-фактографические — содержат как документальные, так и фактографические данные.

В простейшем случае *фактографическая АИС* имеет в качестве базы один файл с записями фиксированного формата (все записи имеют одну и ту же структуру и длину). Примером может являться библиографическая система.

Среди атрибутов форматированных записей обычно существует атрибут, однозначно идентифицирующий запись - *основной* (или первичный) *ключ*.

Атрибуты, которыми задаются свойства, принадлежащие не одной, а нескольким записям называются *дополнительными* (вторичными) *ключами*.

Поиск нужных записей по дополнительному ключу обычно разбивают на два этапа:

- 1) выявление основного ключа у записей с заданными дополнительными ключами;
- 2) нахождение по основному ключу адреса записи, а затем и самой записи.

Для быстрого поиска основного ключа служат *инвертированные списки*. Каждый список состоит из пар значений дополнительного ключа и соответствующего ему множества значений основного ключа.

Первоначальный список БД «Отдел кадров»

Учетный №	1	2	3	4
Фамилия	Иванов И.	Петров П.	Сидоров С.	Семенов А.
Год рождения	1968	1952	1968	1958
№ спец-ти	270100	290100	270300	290100

Список, инвертированный по ключу «Год рождения»

Год рождения	1968	1952	1958
Учетный №	1,3	2	4

Основной задачей *документальной АИС* является поиск документов по их содержанию. Наиболее доступным методом является поиск по дескрипторам.

Под *дескриптором* понимают некоторое фиксированное множество слов (в том числе профессиональных терминов), которые в наибольшей степени характеризуют содержание документального фонда.

Алгоритм использования фиксированного словаря дескрипторов :

- 1) просмотр текстов запросов;
- 2) фиксация всех встречающихся дескрипторов;
- 3) просмотр всех текстов и выделение тех, которые содержат все необходимые дескрипторы.

Информационный шум — это неотфильтрованный поток информации, в котором полезность полученных данных уменьшается прямо пропорционально количеству этих данных.

5.3. Структурирование данных

Структура данных (в информационном смысле) — это представление пользователя о данных, не зависящее от способа их хранения. Структура данных характеризует типы данных и правила их взаимосвязи (отношений).

Тип данных образует множество значений, которые могут принимать соответствующие ему данные.

Отношение — одно из основных понятий современной информатики. Бывают: аналитические, ассоциативные, парадигматические, синтагматические, подчинения.

Структурирование данных задают прежде всего с помощью различного рода отношений порядка (упорядоченности).

Простейший вид упорядоченности задают обычной нумерацией данных с помощью последовательности чисел.

Например, X_i , где i пробегает целые числа от m до n , идентифицирует упорядоченный набор данных, называемый обычно *одномерным массивом*.

Двухиндексный идентификатор X_{ij} идентифицирует двумерный массив и т.д.

В упорядоченных таким образом массивах возникают отношения следования. Так, следующим по индексу j для элемента X_{ij} будет элемент $X_{i,j+1}$, а предыдущим – $X_{i,j-1}$. Если индекс j пробегает значения от m до n , то для m не существует предыдущего, а для n – следующего значения индекса.

Если значения индексов задаются константами, то это прямоугольный массив. Если задано соотношение $m \leq i \leq j \leq n$, то массив треугольный.

Массивы, состоящие из элементов одного и того же типа, называются *однородными*. Одномерные однородные массивы называются *векторами*, двумерные – *матрицами*.

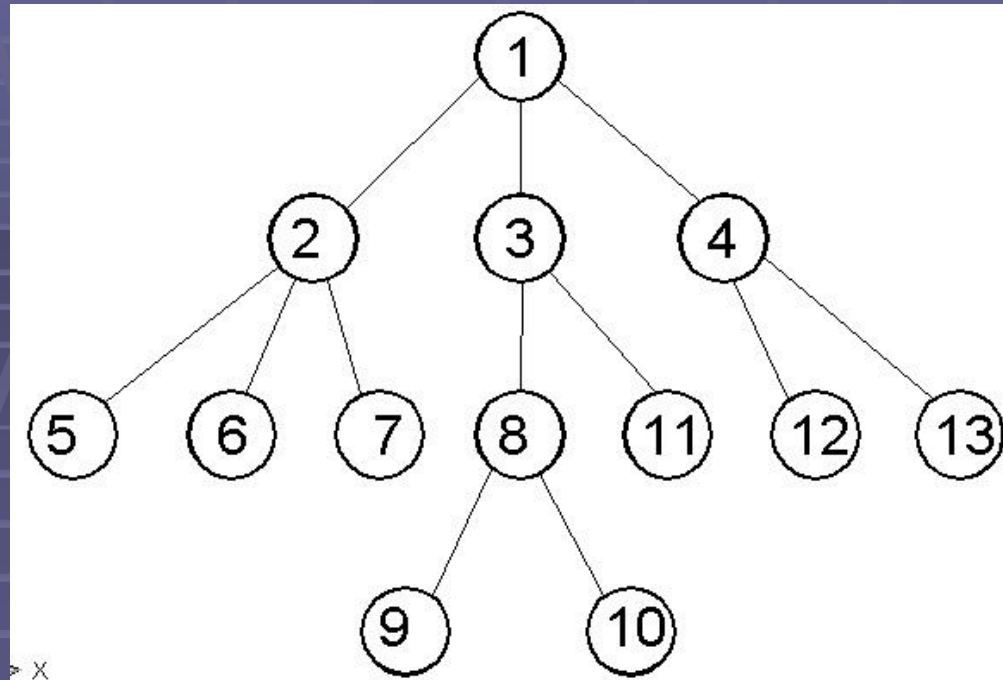
5.4. Модели данных, особенности их организации и поиска в них данных

Более сложные, составные структуры данных, представленные в виде фиксированной системы понятий и правил для описания структуры, называются *моделями данных*.

Распространены графовые и реляционные модели данных.

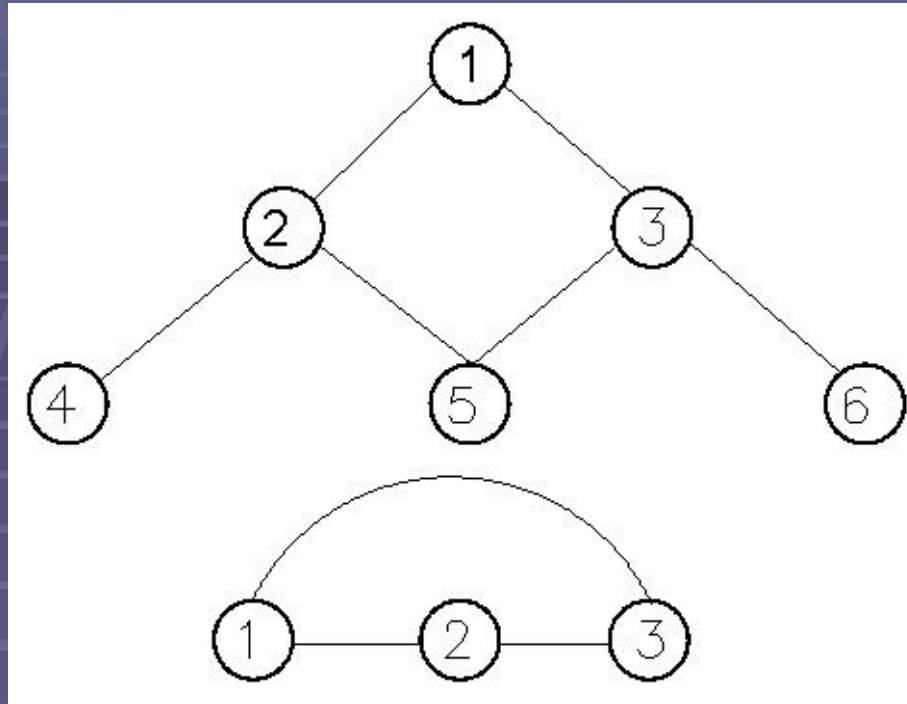
В *графовых моделях* структуру данных изображают и описывают в виде графа, в котором узлами (вершинами) являются типы данных, а дугами (связями) - отношения между ними. Графовый подход используют для описания иерархических и сетевых моделей данных.

Особенности иерархических моделей данных



Верхний узел называется *корнем*, *родовым* или *старшим узлом*. Узлы, из которых выходят дуги на нижний уровень, называются *исходными*, в которые входят дуги с высшего уровня — *порожденными*. Вершины нижнего уровня, не имеющие порожденных узлов, называются *листьями*. Характерно отношение 1:М (один ко многим).

Особенности сетевых моделей данных



В сетевых моделях данных порожденный узел может иметь более одного исходного, т.е. Один тип записи является членом более чем одного типа набора. Для сетей характерны отношения М:1 и М:М.

Реляционные модели данных предложены в 1970 г. Основаны на представлении данных в виде отношений, которые могут подвергаться нормализации – пошаговому процессу приведения их к двумерной табличной форме. К такой же форме могут быть приведены и иерархические и сетевые отношения.

В таблице каждая строка которых соответствует значениям свойств (атрибутов), которыми обладает объект данного типа; каждый из столбцов соответствует множеству значений, которые принимает некоторый атрибут этого типа.

Отношение есть множество векторов из n элементов – *кортежей* $(X_1, \dots, X_n)$, где n – (число столбцов) называемое *степенью отношения*. Совокупность значений одного атрибута (соответствующая столбцу таблицы) называется *доменом*.

Для описания отношений и манипуляций над ними в реляционной модели данных используется строгий математический язык, основанный на алгебре отношений (*реляционной алгебре*) и исчислении отношений (реляционное исчисление).

Операции реляционной алгебры позволяют вырезать отдельные домены из отношения, объединять отношения, причем, в результирующем отношении вырезаются совпадающие строки и др.

Обозначение панели	Вид отделки	Количество бетона	Количество стали

Поиск данных в реляционной БД происходит с помощью индексных файлов. *Индексный файл*, как правило, представляет собой отдельную таблицу, содержащую ключ каждой записи и ее адрес на запоминающем устройстве.

К числу СУБД реляционного типа относятся хорошо известные системы: Access, dBase, Clipper, FoxBASE, R:BASE, Paradox, FoxPro, Oracle, MySQL и т.д. Во всех этих СУБД записи и поля имеют ограничение на число записей (128...10024) и длину (обычно 4000...5000 байт). Исключение составляет поле Memo, в котором хранят текстовую информацию большого объема.

5.6. Процесс создания баз данных

Концепцию, в рамках которой удобно и полезно рассматривать развитие системы БД во времени, называют *жизненным циклом*.

Жизненный цикл БД делят на две самостоятельные фазы: 1) анализ и проектирование; 2) реализация и функционирование.

В течение первой фазы осуществляется сбор требований пользователей и проектирование БД, в течение второй – машинная реализация и использование БД для решения прикладных задач.

Основная цель проектирования базы данных – обеспечение пользователей точными данными за приемлемое время.

Фаза анализа и проектирования БД



Фаза реализации и функционирования

Реализация БД подразумевает создание (наполнение) базы и разработку прикладных программ, а также загрузку БД. Загрузка связана с преобразованием имеющихся данных из формы логической и физической структуры в новую форму, соответствующую результатам проектирования БД.

Анализ функционирования и поддержка осуществляются для регистрации (сбора) и статистической обработки данных о функционировании системы.

Этап модернизации и адаптации предусматривает внесение в проект изменений, возникающих при появлении новых требований, полученных в результате анализа.

Реорганизация БД – это осуществление любых действий, связанных с изменением ее логической и физической структуры.