

Тема №2 Предметная область базы данных и ее модели.

1. Понятие предметной области
2. Информационная модель предметной области базы данных
3. Сущности, атрибуты и домены атрибутов
4. Отношения и связи
5. Подтипы и супертипы
6. Диаграммы «сущность-связь»
7. Контроль качества результатов анализа предметной области

Понятие предметной области

Понятие *предметной области базы данных* является одним из базовых понятий информатики и не имеет точного определения.

Его использование в контексте ИС предполагает существование устойчивого во времени соотношения между именами, понятиями и определенными реалиями внешнего мира, не зависящей от самой ИС и ее круга пользователей.

Таким образом, введение в рассмотрение понятия *предметной области базы данных* ограничивает и делает обозримым пространство информационного поиска в ИС и позволяет выполнять запросы за конечное время.

Понятие предметной области

Совокупность реалий (объектов) внешнего мира - объектов, о которых можно задавать вопросы, - образует *объектное ядро предметной области*, которое имеет онтологический статус.

Нельзя получить в ИС ответ на вопрос о том, что ей неизвестно. Термин объект является первичным, неопределяемым понятием. Синонимами термина "объект" являются "реалия, сущность, вещь".

Выделяемые в предметной области объекты превращаются аналитиками (а не проектировщиками базы данных) в сущности.

Понятие предметной области

Один из подходов к классификации объектов предметной области:



Понятие предметной области

Примерами сущностей (с точки зрения ИС) или объектов (с точки зрения внешнего мира) являются отдельный студент, группа студентов, аудитория, время занятий, слова, числа, символы.

Считается, что быть объектом - это значит быть *дискретным и различимым*.

Примеры "не-объектов" - это мир, время, смысл, хотя и такие категории могут сохраняться в базе данных.

Понятие предметной области

С объектами связано две проблемы: *идентификация и адекватное описание*.

Для идентификации используют имя. При этом предполагается, что происходит отказ от его смысла, который присущ естественному языку. Используется только указательная функция имени.

Имя - это прямой способ идентификации объекта. К косвенным способам идентификации объекта относят определение объекта через его свойства (характеристики или признаки).

Понятие предметной области

Объекты взаимодействуют между собой через свои свойства, что порождает *ситуации*.

Ситуации - это взаимосвязи, выражающие взаимоотношения между объектами.

Ситуации в предметной области описываются посредством высказываний о предметной области с использованием *исчисления высказываний* и *исчисления предикатов*.

Пример: высказывание "Программист и менеджер есть служащие компании" описывает отношение включения.

Понятие предметной области

Пример: Рассмотрим высказывание: Студент Иванов А.А, родился в 1982 году. Оно выражает следующие свойства объекта "Иванов А.А.":

в явном виде - год рождения;

в неявном - принадлежность к студентам.

Первое свойство устанавливает связь между объектами "Иванов А.А." и "Год рождения", а второе - между объектами "Иванов А.А." и "Множество студентов". Формализация этого высказывания представляется как результат присваивания значений переменным, входящим в предикаты:

РОДИЛСЯ (Иванов А.А., 1982)

ЯВЛЯЕТСЯ СТУДЕНТОМ (Иванов А.А.)

Понятие предметной области

Ситуация содержит *высказывание* об объектах предметной области, которому можно *приписать некоторую оценку истинности* и *представить в виде предиката* после введения переменных.

Таким образом, совокупность высказываний о предметной области можно трактовать как определение информационного пространства для базы данных.

Понятие предметной области



Понятие предметной области

Предметная область - это целенаправленная первичная трансформация картины внешнего мира в некоторую умозрительную картину, определенная часть которой фиксируется в ИС в качестве алгоритмической модели фрагмента действительности.

Понятие предметной области было введено в начале 80-х годов прошлого века, когда учеными в области ИС была осознана необходимость использовать семантические модели для представления информации в компьютерных системах.

Информационная модель предметной области базы данных

Информационная модель данных предназначена для представления семантики предметной области в терминах субъективных средств описания - сущностей, атрибутов, *идентификаторов сущностей, супертипов, подтипов* и т.д.

Информационная модель предметной области базы данных содержит следующие основные конструкции:

- диаграммы "сущность-связь" (Entity - Relationship Diagrams);
- определения сущностей;
- уникальные *идентификаторы сущностей*;
- определения *атрибутов сущностей*;
- отношения между сущностями;
- супертипы и подтипы.

Сущности, атрибуты и домены атрибутов

Предметом информационной модели является абстрагирование объектов или явлений реального мира в рамках предметной области, в результате которого выявляются сущности (entity) предметной области. Они обозначаются *именем существительным* естественного языка.

Сущность описывается с помощью данных, именуемых свойствами или *атрибутами (attributes) сущности*. Атрибуты обозначаются *именами существительными* естественного языка.

Сущности вступают в связи друг с другом через свои атрибуты.

Сущности, атрибуты и домены атрибутов

Каждая группа атрибутов, описывающих одно реальное проявление сущности, представляет собой *экземпляр* (instance) сущности.

Экземпляры сущности - это реализации сущности, отличающиеся друг от друга и допускающие однозначную идентификацию.

Сущности, атрибуты и домены атрибутов

При представлении сущности в базе данных хранятся только ее атрибуты.

Одним из основных компьютерных способов распознавания сущностей в базе данных является присвоение сущностям идентификаторов (Entity identifier).

Часто идентификатор сущности называют ключом.

Сущности, атрибуты и домены атрибутов

Некоторые сущности имеют естественные идентификаторы.

Пример: естественным идентификатором счета-фактуры является его номер.

Идентификаторы сущности могут быть *составными* - составленными из нескольких атрибутов и *атомарными* - составленными из одного атрибута сущности.

Задача проектировщика баз данных - обеспечить при сохранении экземпляров сущности в базе данных наличие у каждого ее нового экземпляра уникального идентификатора.

Сущности, атрибуты и домены атрибутов

Уникальный идентификатор сущности - это атрибут сущности, позволяющий отличать одну сущность от другой.

Если сущность имеет несколько уникальных идентификаторов, так называемых возможных ключей, то проектировщик должен выбрать *первичный ключ сущности*.

Различают *однозначные* и *многозначные* атрибуты. Однозначными являются атрибуты, которые в пределах конкретного экземпляра сущности имеют только одно значение. В противном случае они считаются многозначными.

Сущности, атрибуты и домены атрибутов

Каждый *атрибут сущности* имеет домен (domain). Домен - это выражение, определяющее значения, разрешенные для данного атрибута. Иными словами, **домен - это область значений атрибута.**

Проектировщик базы данных должен проконтролировать, чтобы в *информационной модели предметной области* для каждого *атрибута сущностей* был определен домен.

На уровне информационного моделирования данных назначение домена атрибуту носит общий характер. Например, атрибут текстовый, числовой, бинарный, дата или "не определен".

Отношения и связи

Сущности не существуют отдельно друг от друга. Между ними имеются реальные отношения (Relationship), и они должны быть отражены в *информационной модели предметной области*.

При выделении отношений акцент делается на фиксацию связей и их характеристик.

Отношение (связь) представляет собой соединение (взаимоотношение) между двумя или более сущностями.

Каждая связь реализуется через значения *атрибутов сущностей*. Обычно связь обозначается глаголом. Каждая связь также должна иметь свой уникальный идентификатор связи.

Отношения и связи

Связи характеризуются *степенью связи* и *классом принадлежности* сущности к связи.

Степень (мощность) связи - это отношение числа сущностей, участвующих в образовании связи. Например, "один-к-одному", "один-ко-многим", "многие-ко-многим".

На уровне информационной модели допускается неопределенная или неразрешенная связь.

Класс принадлежности сущности - это характер участия сущности в связи.

Отношения и связи

Различают *обязательные* и *необязательные* классы принадлежности сущности к связи. Обязательным является такой класс принадлежности, когда экземпляры сущности участвуют в установлении связи в обязательном порядке. В противном случае сущность принадлежит к *необязательному* классу принадлежности.

Для необязательного класса принадлежности сущности степень связи может быть равна нулю, т.е. экземпляр сущности можно связать с 0, 1 или несколькими экземплярами другой сущности.

Для обязательного класса принадлежности степень связи не может равняться нулю.

Отношения и связи

Отношения, связывающие сущность саму с собой, называются *рефлексивными*.

Типичным примером рефлексивных отношений является определение структуры подчиненности в отношении "Сотрудники".

Рефлексивные отношения чаще всего отражают иерархические отношения внутри структуры данных.

Отношения и связи

С точки зрения отношений различают *слабые* (weak) сущности.

Слабые сущности - это сущности, которые не могут присутствовать в базе данных, пока не существует связанного с ней экземпляра другой сущности.

Пример: является заказ, который не может существовать без клиента.

Слабые сущности имеют обязательный класс принадлежности, и степень связи такой сущности не может равняться нулю. Связь "заказ-клиент" является обязательной.

Подтипы и супертипы

Иногда выделенная сущность несет в себе отношение включения или часть-целое. При этом существует некоторый атрибут, значения которого порождают разбиение множества экземпляров сущности на непересекающиеся подмножества - *категории сущности*.

Категории сущности называются *подтипами* и выделяют в подчиненную в рамках отношения сущность, которая является категорией исходной сущности.

Подтипы и супертипы

Иногда из исходной сущности выделяются общие для полученных категорий атрибуты, и таким образом выделяется сущность, которая становится *супертипом*.

За выделенной сущностью-супертипом обычно оставляют наименование исходной сущности, хотя ее семантический смысл меняется.

Супертип с порожденными им *подтипами* является примером так называемой составной сущности.

Диаграммы «сущность-связь»

Типичной формой документирования *информационной модели предметной области* являются *диаграммы "сущность-связь"* (ER-диаграммы).

ER-диаграмма позволяет графически представить все элементы информационной модели согласно простым, интуитивно понятным, но строго определенным правилам - *нотациям*.

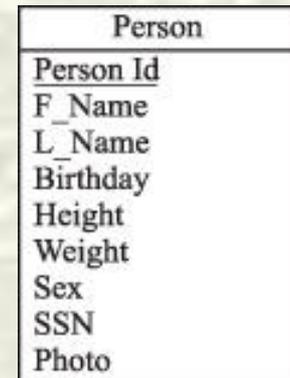
Диаграммы «сущность-связь»

Сущность на ER-диаграмме представляется прямоугольником с именем в верхней части.

Представление сущности Person (персонал) на ER-диаграмме :



Представление сущности Person с атрибутами и уникальным идентификатором сущности:



Диаграммы «сущность-связь»

Домены назначаются аналитиками и фиксируются в специальном документе - словаре данных (Data Dictionary).

На стадиях разработки логической и физической моделей реляционной базы данных домены уточняются в сущностях на ER-диаграмме.

Проектировщик базы данных должен тщательным образом изучить домены каждого атрибута с точки зрения их реализуемости в СУБД, с участием аналитиков внести в них изменения, если условие реализуемости не выполняется.

Диаграммы «сущность-связь»

При определении доменов проектировщик руководствуется следующим:

- для реализации реляционной базы данных требуется использовать реляционную СУБД, например Oracle;
- в большинстве реляционных СУБД в качестве языка манипулирования и описания данных используется диалект SQL, поддерживающий определенные стандарты, например ANSI SQL-92.

Диаграммы «сущность-связь»

Пример: Изначально домен атрибута Photo (Фотография) сущности Person (Персона) был определен как Image (Рисунок).

Проектировщик базы данных должен изменить значение домена на LONG RAW (СУБД Oracle) или BLOB (двоичный большой объект) (SQL-92).



PERSON	
<u>PERSON ID</u>	<u>INTEGER</u>
F_NAME	CHAR(30)
L_NAME	CHAR(20)
BIRTHDAY	DATE
HEIGHT	HUMBER(5,2)
WEIGHT	HUMBER(4,1)
SEX	HUMBER(1)
SSN	CHAR(30)
PHOTO	LONG RAW

Диаграммы «сущность-связь»

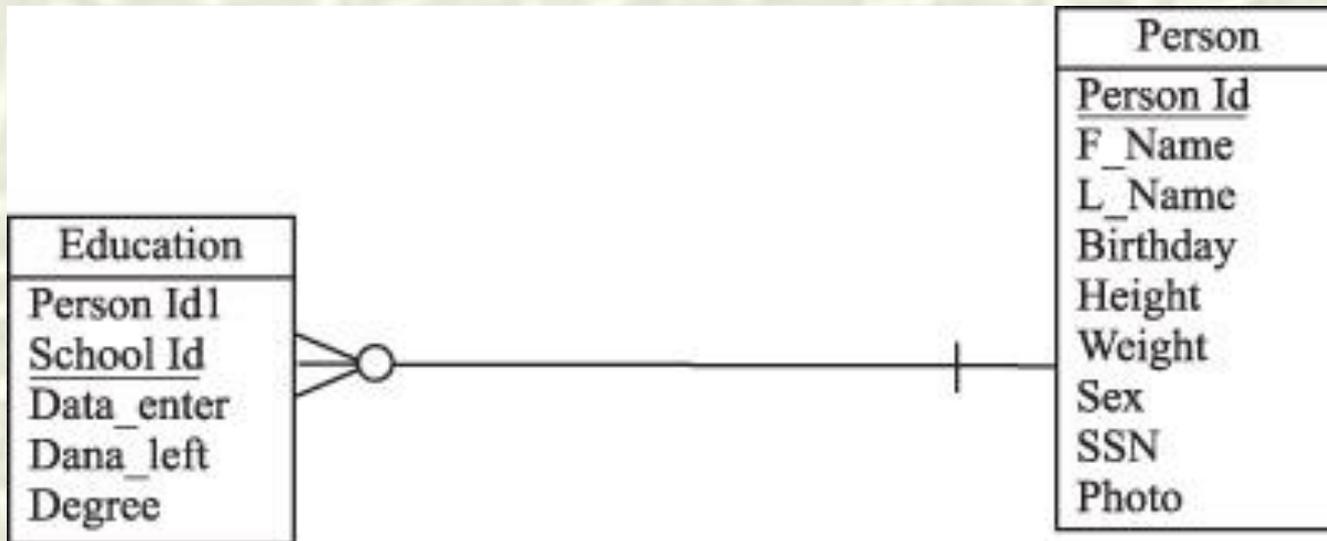
Отношение (связь) сущностей на ER-диаграмме изображается линией, соединяющей эти сущности.

Степень связи изображается с помощью символа "птичья лапка", указывающего на то, что в связи участвует много (N) экземпляров сущности, и одинарной горизонтальной чертой, указывающей на то, что в связи участвует один экземпляр сущности.

Необязательный класс принадлежности изображается с помощью кружочка на линии отношения рядом с сущностью, обязательный класс принадлежности - с помощью вертикальной черты на линии отношения рядом с сущностью.

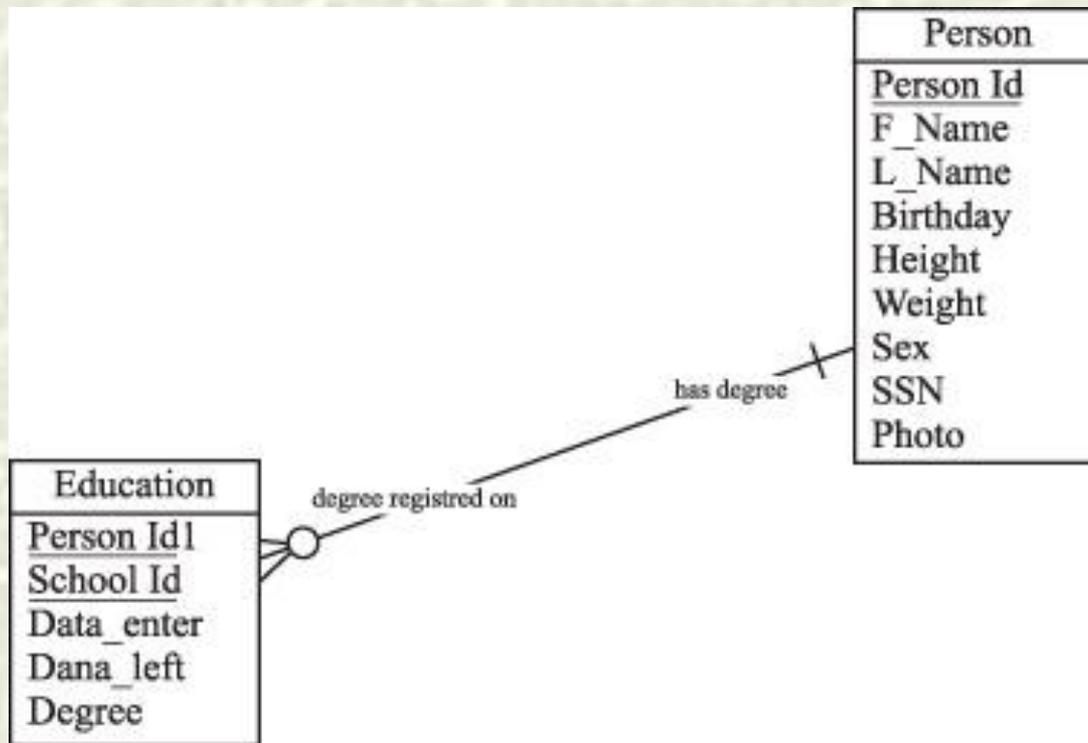
Диаграммы «сущность-связь»

Пример: каждая специальность по образованию должна быть зарегистрирована за определенным физическим лицом (персоной), физическое лицо может иметь ноль или более специальностей по образованию.



Диаграммы «сущность-связь»

Как правило, отношения на *ER-диаграммах* именуются с обеих сторон.

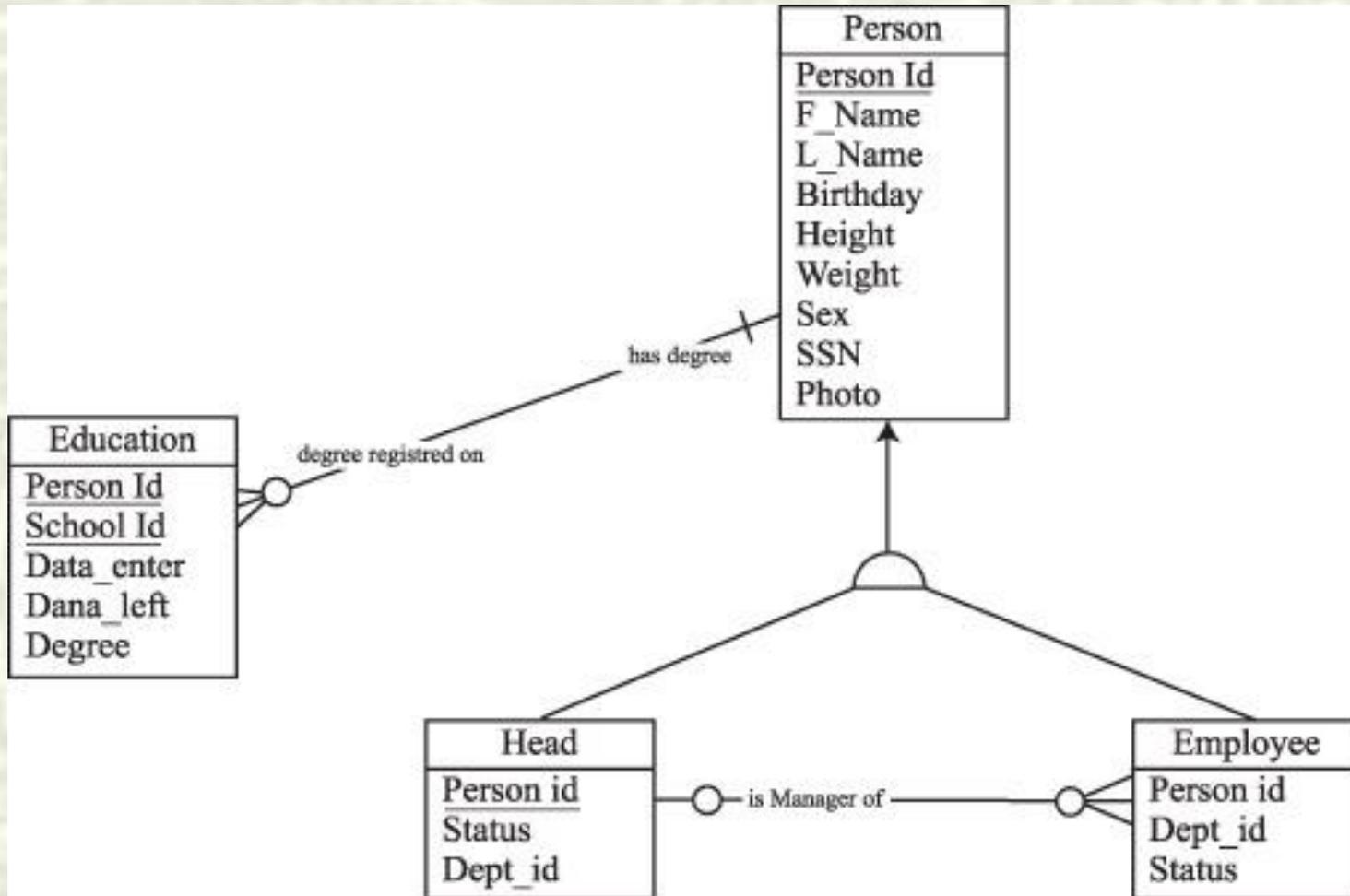


Диаграммы «сущность-связь»

Супертипы и *подтипы*, так же как и сущности, обозначаются на ER-диаграмме с помощью прямоугольников. Отношения между ними изображаются с помощью "вилки", имеющей в точке ветвления полукруг.

Подтипы содержат только атрибуты, характерные для выделенных категорий.

Диаграммы «сущность-связь»



Контроль качества результатов анализа предметной области

Если проектировщик баз данных получил от аналитиков набор материалов с результатами анализа предметной области базы данных, то его *задача - произвести контроль качества предоставленных результатов анализа в целях обеспечения их полноты и достоверности.*

Первое, что необходимо сделать, - составить перечень полученных документов и проверить, все ли необходимые документы присутствуют.

Контроль качества результатов анализа предметной области

Проектировщику должны быть представлены:

- информационная модель предметной области базы данных
- совокупность частных моделей, которые относятся к функциональной модели предметной области базы данных
- общесистемные требования и решения.

Контроль качества результатов анализа предметной области

На диаграммах "сущность-связь" должны быть представлены сущности, связи, степени связи, классы принадлежности сущности, неопределенные связи, супертипы/подтипы сущности.

Каждая сущность и отношение должны быть поименованы и занесены в словарь данных; каждая сущность может появиться на диаграмме только один раз; каждая сущность должна вступать в отношение (отсутствие "висящих" сущностей), и каждое отношение должно иметь сущности (отсутствие "висящих" отношений).

Связи между сущностями должны осуществляться через точно указанные атрибуты.

Контроль качества результатов анализа предметной области

Проектировщик базы данных должен проконтролировать, чтобы в информационной модели предметной области для каждого атрибута сущностей был определен домен.

Определение домена, даваемое аналитиками, носит самый общий характер (число, текст, дата, графика и т.д.), и будет далее уточнено при проектировании.

На этапе контроля качества важно, чтобы домен был определен.

Контроль качества результатов анализа предметной области

Формальные процедуры проверки качества моделей позволяют проектировщику базы данных убедиться в их "формальной" достоверности, но не позволяют оценить, насколько они адекватно отражают реальную ситуацию внешнего мира в рамках предметной области.

Поэтому часто требуется прибегать к неформальным процедурам проверки результатов моделирования предметной области.

Контроль качества результатов анализа предметной области

Неформальные процедуры заключаются в проведении личных бесед с аналитиками, постановщиками задач, конечными пользователями и руководителями всех автоматизируемых подразделений, проведении семинаров и совещаний всех участников проекта, а также в изучении материалов анализа предметной области.

В ходе неформальных процедур могут быть выявлены существенные ошибки (например, потеря сущности предметной области), которые могут привести к пересмотру некоторых результатов проектирования и реализации баз данных, что в конечном счете может стать причиной срыва запланированных сроков выполнения проекта.