

数据库系统概论

数据模型

北京理工大学

数据模型

- ◆ 模型
- ◆ 数据模型（**data model**）是现实世界数据特征的抽象
- ◆ 数据模型不仅要反应数据本身，还要反应出数据之间的关系
- ◆ 数据模型的要求
 - ◆ 真实
 - ◆ 易于理解
 - ◆ 易于通过计算机实现

数据模型的两种类型

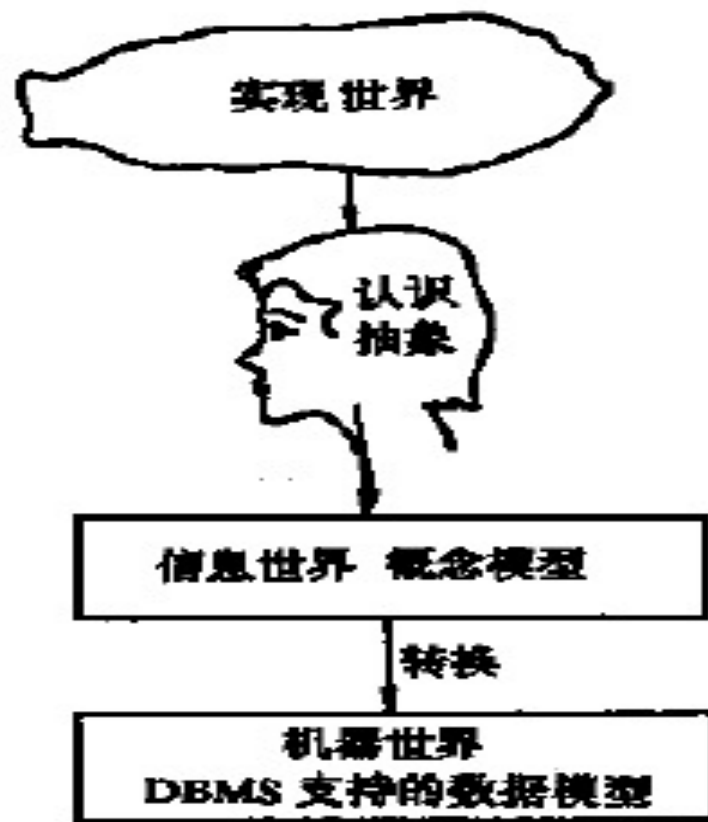
◆根据模型的不同应用目的

◆概念模型（信息模型）

- ◆用户角度，用于数据库设计

◆数据模型

- ◆计算机系统的角度，用于DBMS实现
- ◆网状模型
- ◆层次模型
- ◆关系模型



数据模型的组成要素

- ◆数据结构
- ◆数据操作
- ◆完整性约束

数据结构

- ◆是所研究对象类型的集合
 - ◆数据内容
 - ◆数据和数据之间的联系
- ◆是数据库的静态描述

数据操作

- ◆是对数据库中各个对象允许执行操作的集合以及相关操作的规则
- ◆插入，删除，修改

数据的约束条件

- ◆约束条件是一组完整性规则的集合
- ◆给定数据模型中数据及其联系具有的制约和依存规则

概念模型

- ◆是现实世界到信息世界的第一层抽象
- ◆数据库设计人员进行数据库设计的工具

一些基本概念

- ◆ 实体(**Entity**)-可以相互区别的事物
- ◆ 属性 (**attribute**) 实体的某一方面特性
- ◆ 码(**key**)唯一标识实体的属性集
- ◆ 域(**domain**)属性的取值范围
- ◆ 实体型(**entity type**)用实体名和属性名集合来描述同类实体, 称为实体型
- ◆ 实体集 (**entity set**) 同型实体的集合
- ◆ 关系 (**relationship**) 实体内部或者实体之间的关系

联系

◆ 1: 1（一对一）

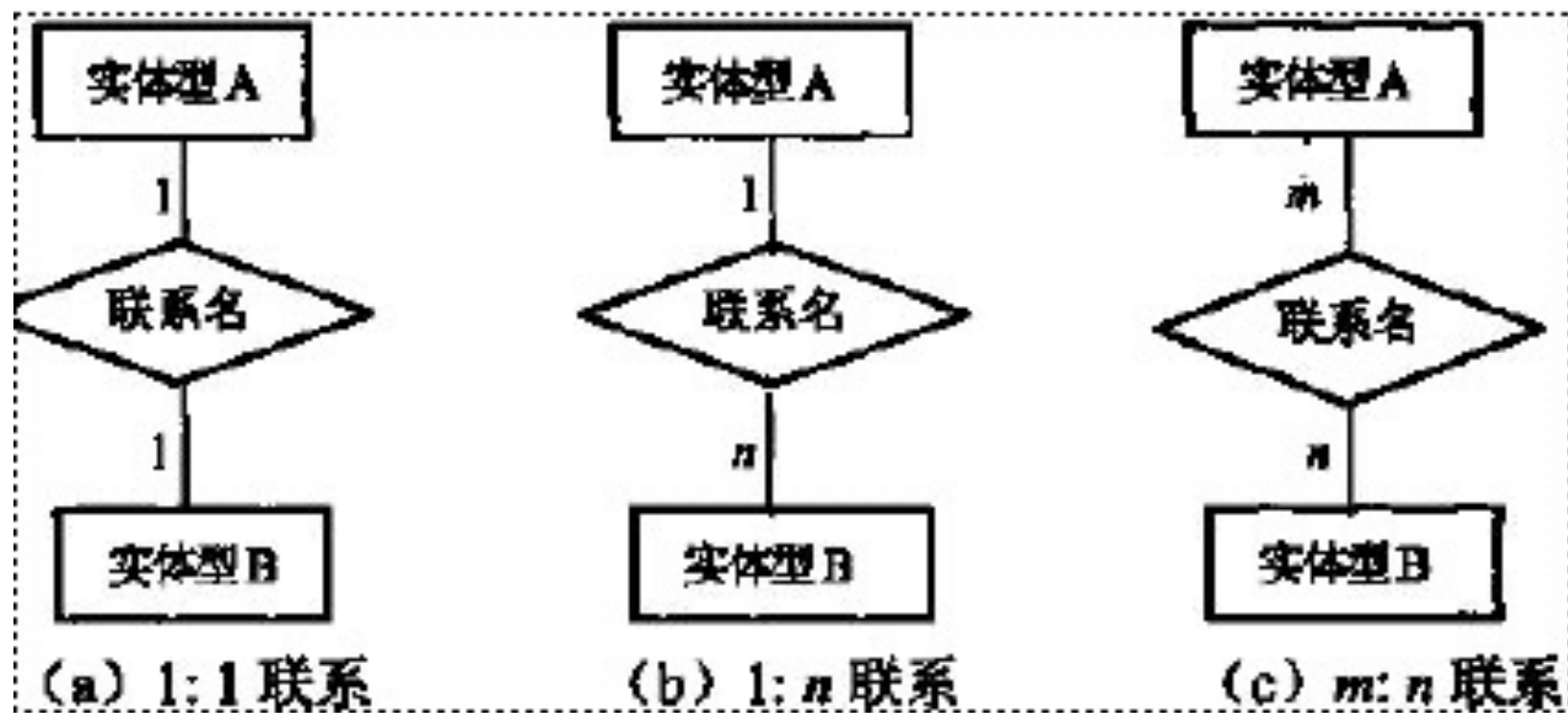
- ◆ 实体集A中每一个实体在实体集B中至多有1个实体与之联系，反之亦然

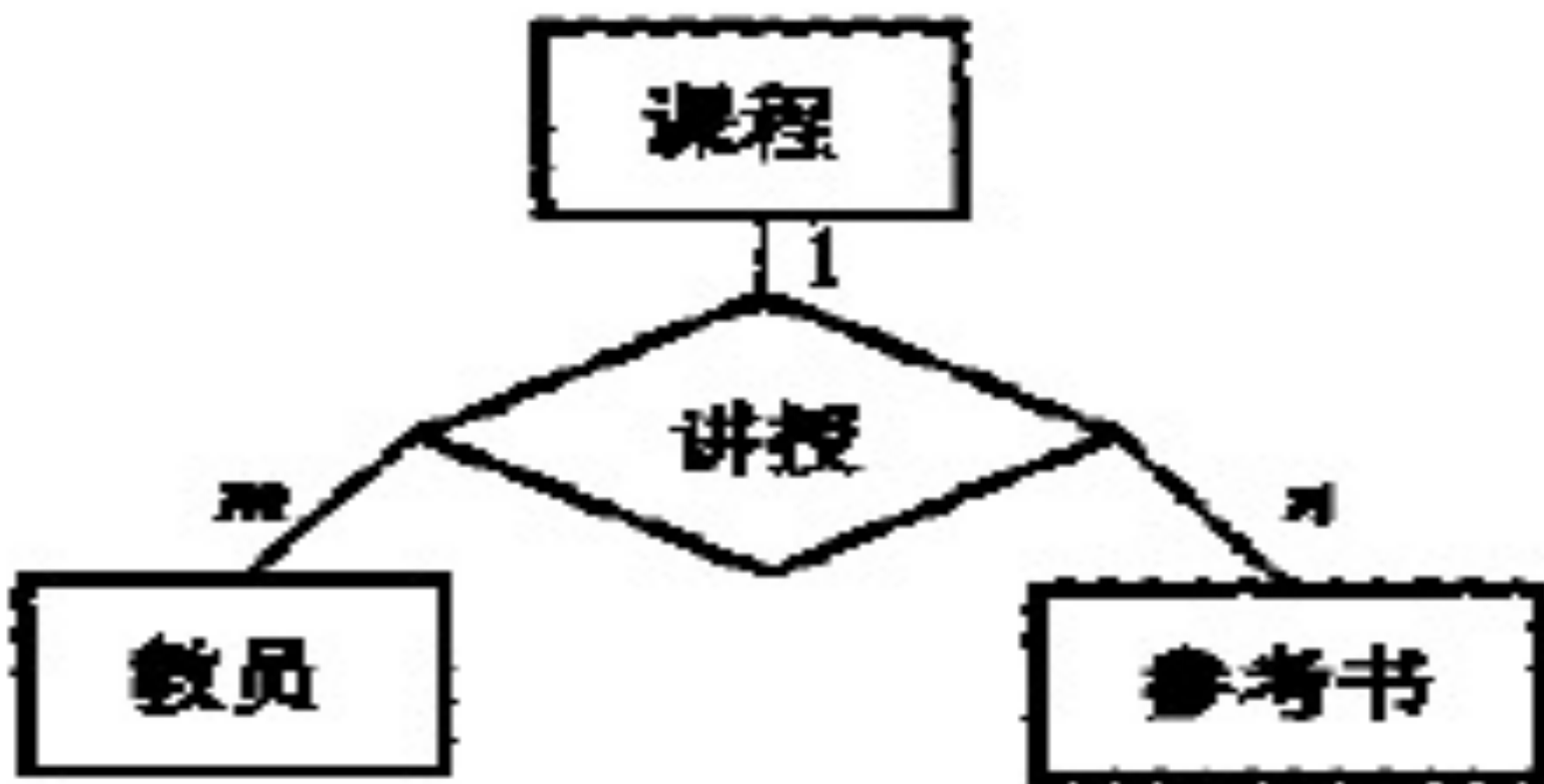
◆ 1: n（一对多）

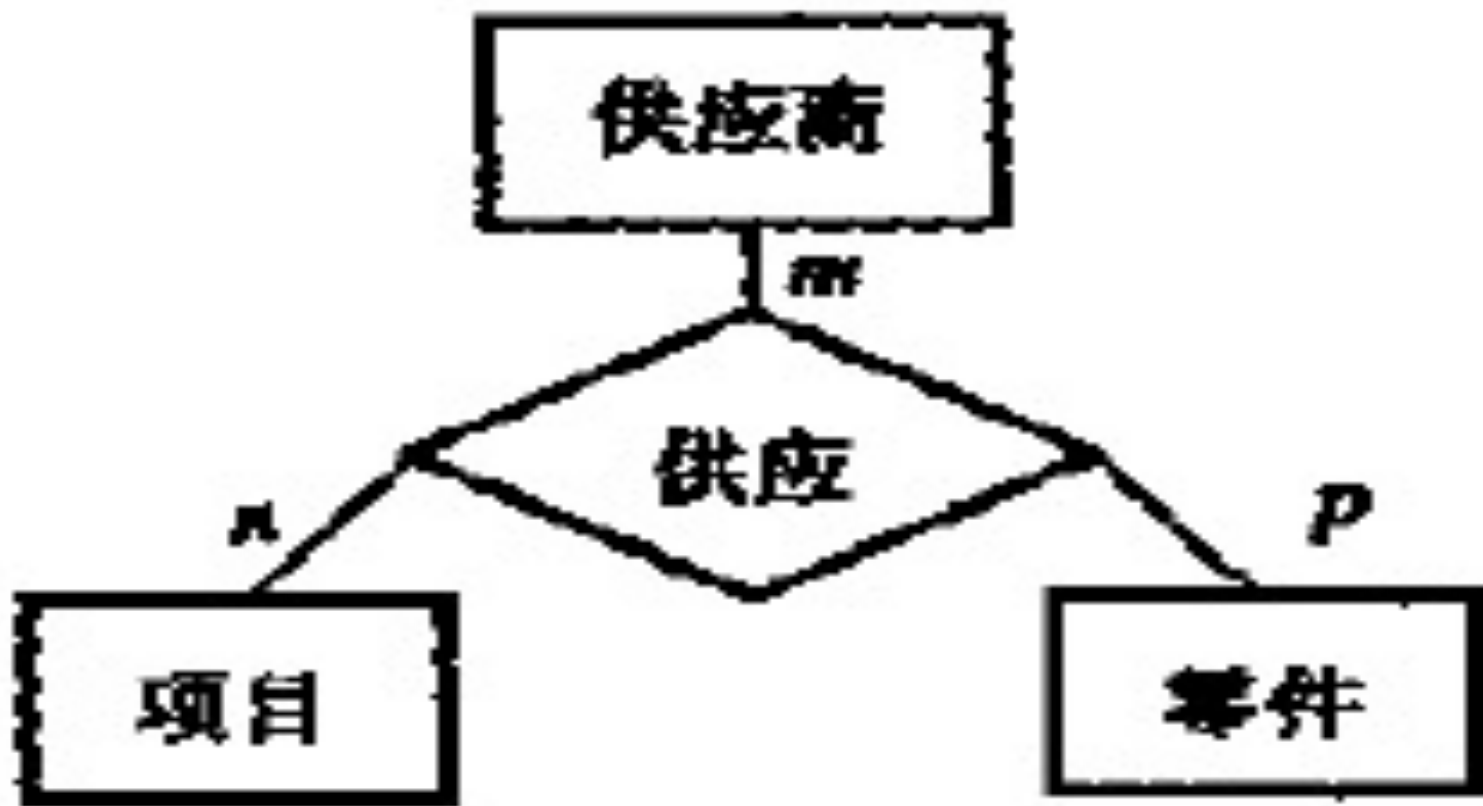
- ◆ 实体集A中每一个实体在实体集B中至有n个实体与之联系，反之是1: 1关系

◆ M: n（多对多）

- ◆ 实体集A中每一个实体在实体集B中至有n个实体与之联系，反之亦然









概念模型的表示方法

- ◆ **P.P.S.chen** 1976年提出的实体—联系方法
- ◆ **E—R模型** 反应实体型 属性和联系
 - ◆ 实体型 矩形 内部写实体名
 - ◆ 属性 椭圆 用无向边和对应的实体联系起来
 - ◆ 联系 菱形 内写联系名 无向边标识联系类型

E-R模型实例

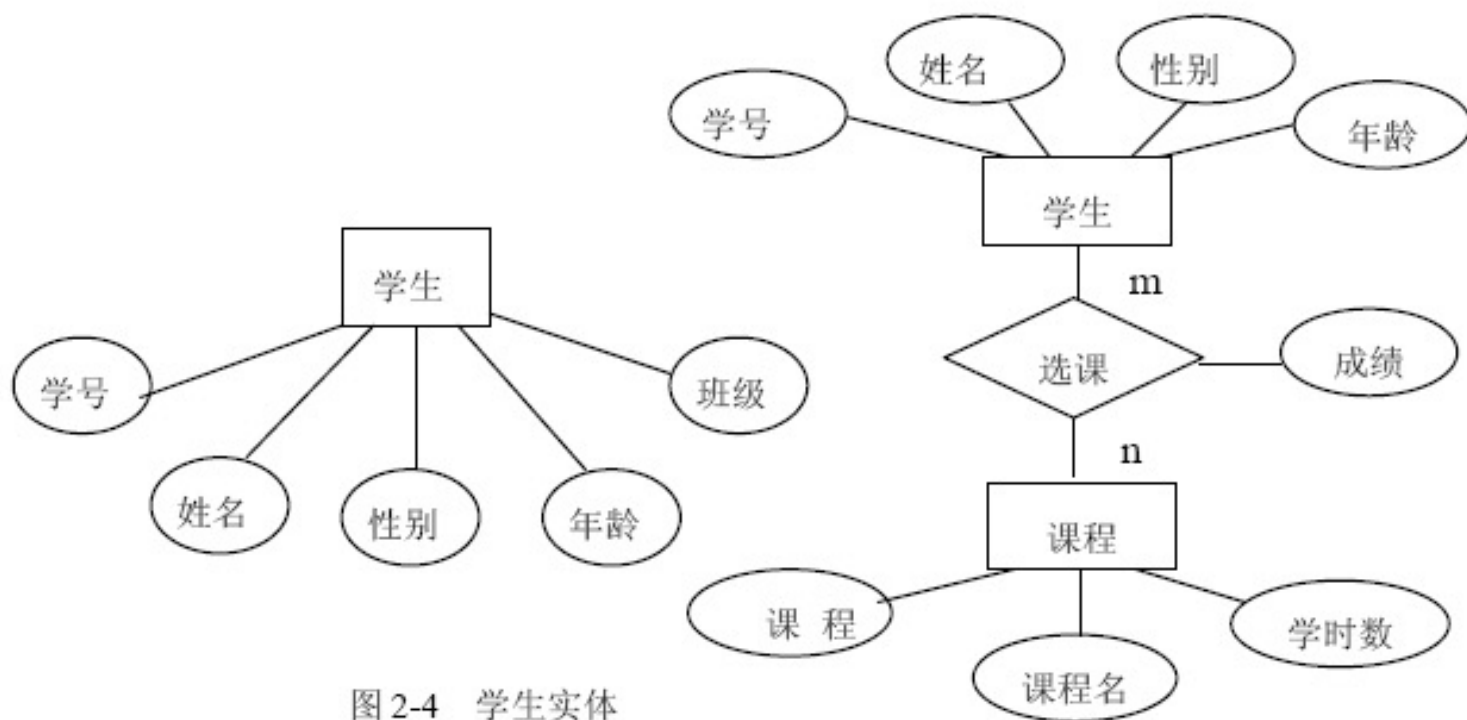
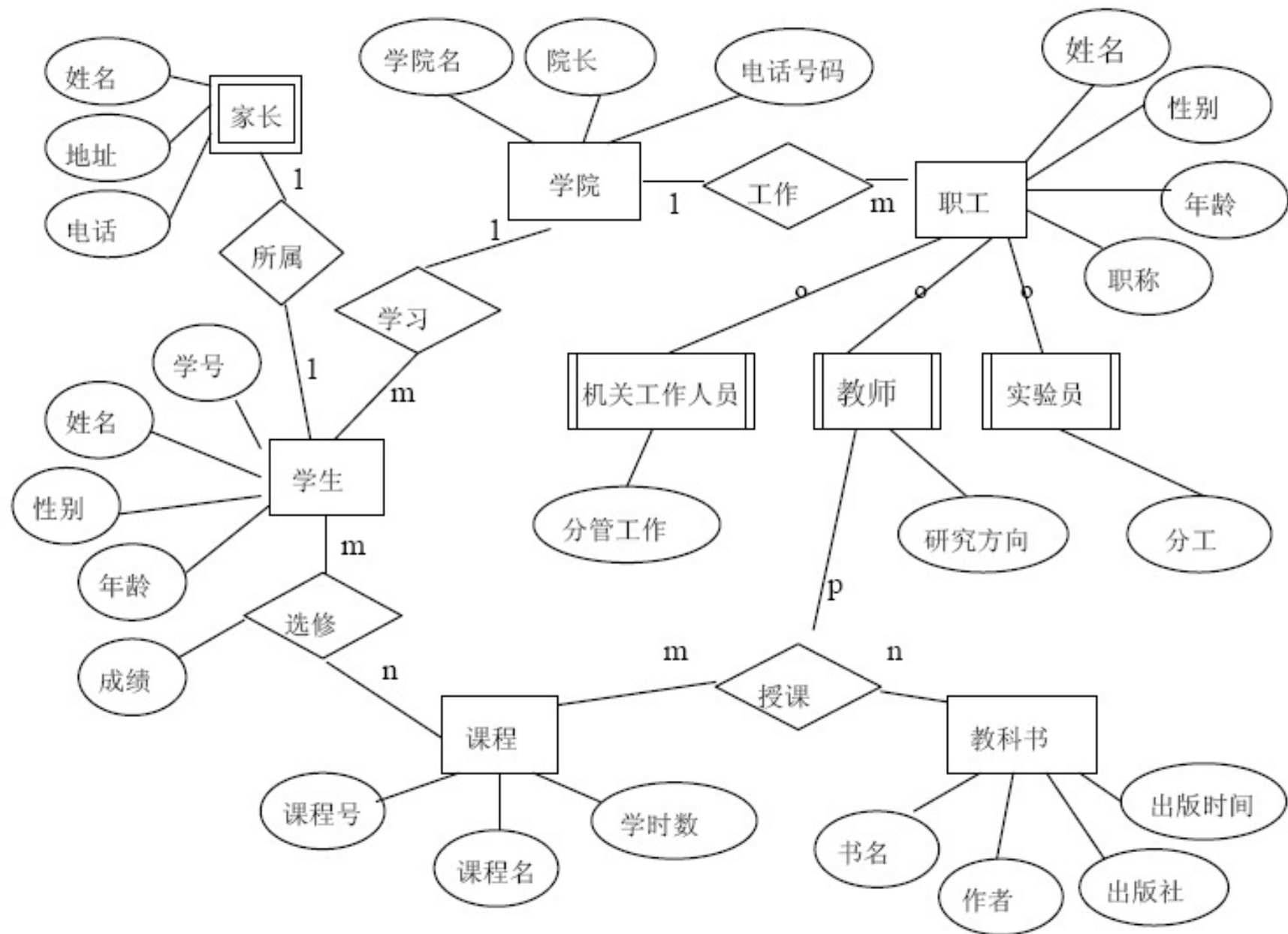
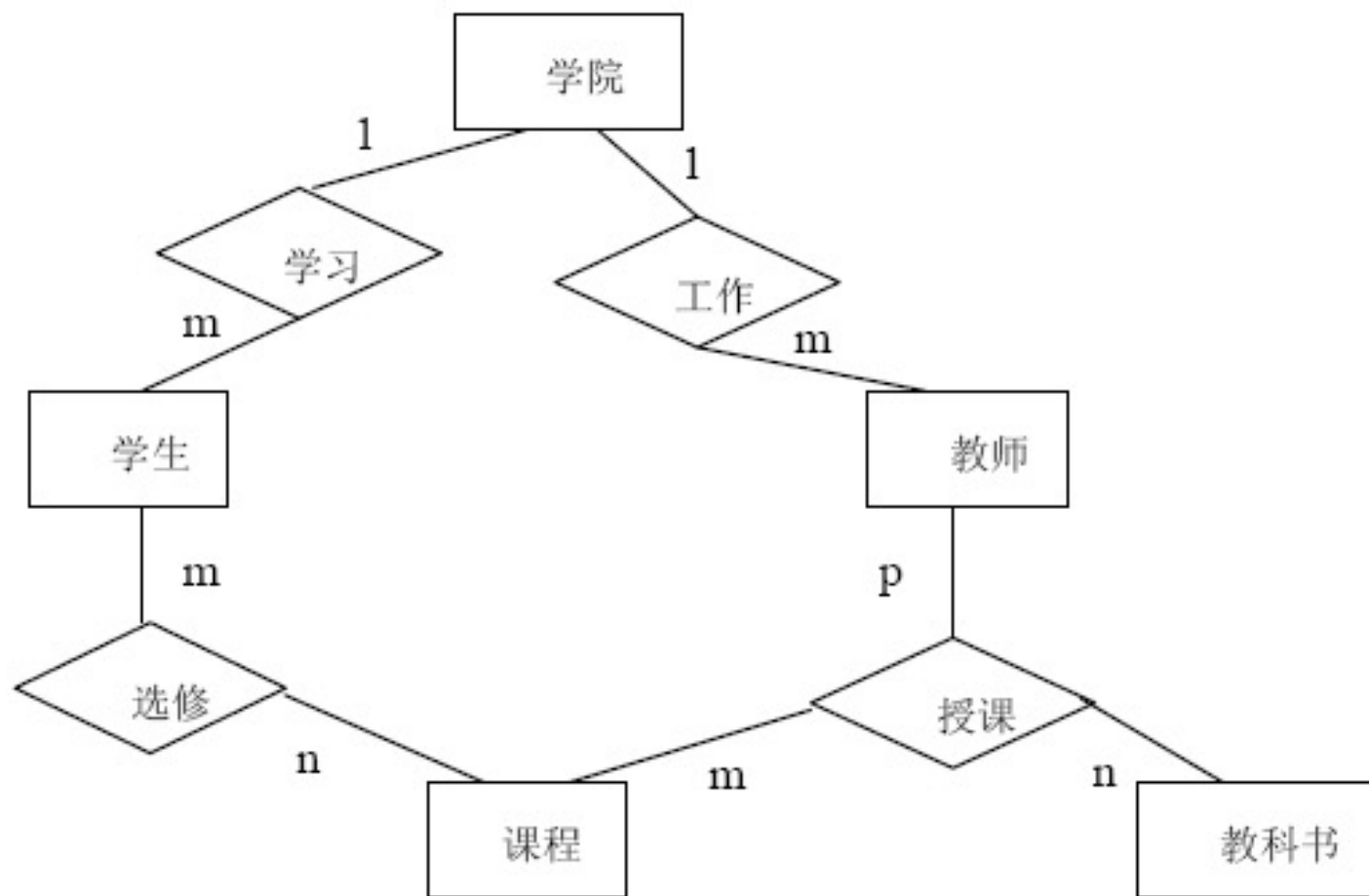


图 2-4 学生实体

图 2-5 学生选课的 E-R 图



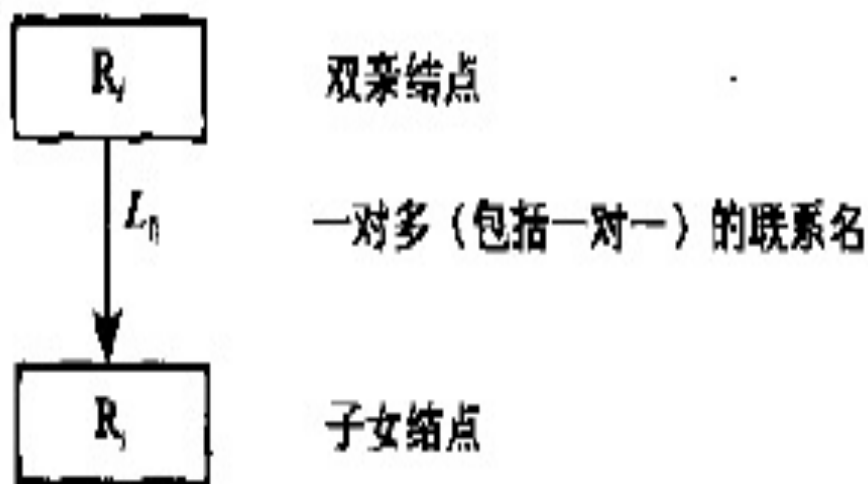


常用的数据模型

- ◆ 层次模型
- ◆ 网状模型
- ◆ 关系模型
- ◆ 面向对象模型

非关系模型

- ◆基本层次联系 指两个记录以及它们之间的联系



层次模型

- ◆用树型结构表示实体和实体之间的联系
- ◆有且只有一个节点没有双亲节点，根节点
- ◆其他节点有且仅有一个双亲节点
- ◆有相同双亲节点的节点称为兄弟节点
- ◆没有子女节点的节点称为叶节点
- ◆每个节点是一个数据类型
- ◆每个节点由若干个字段构成

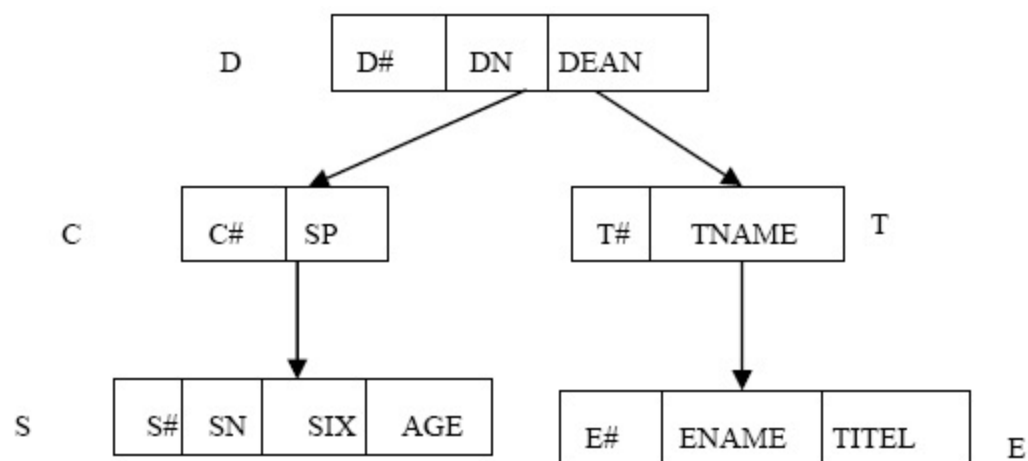
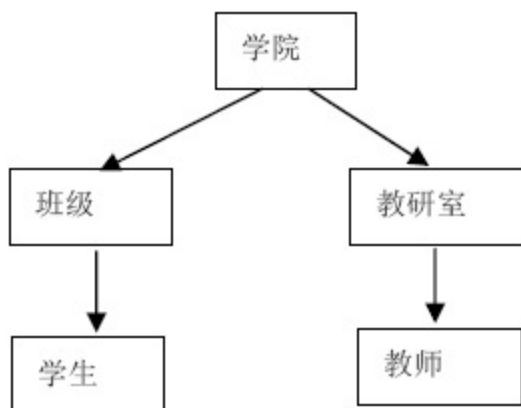
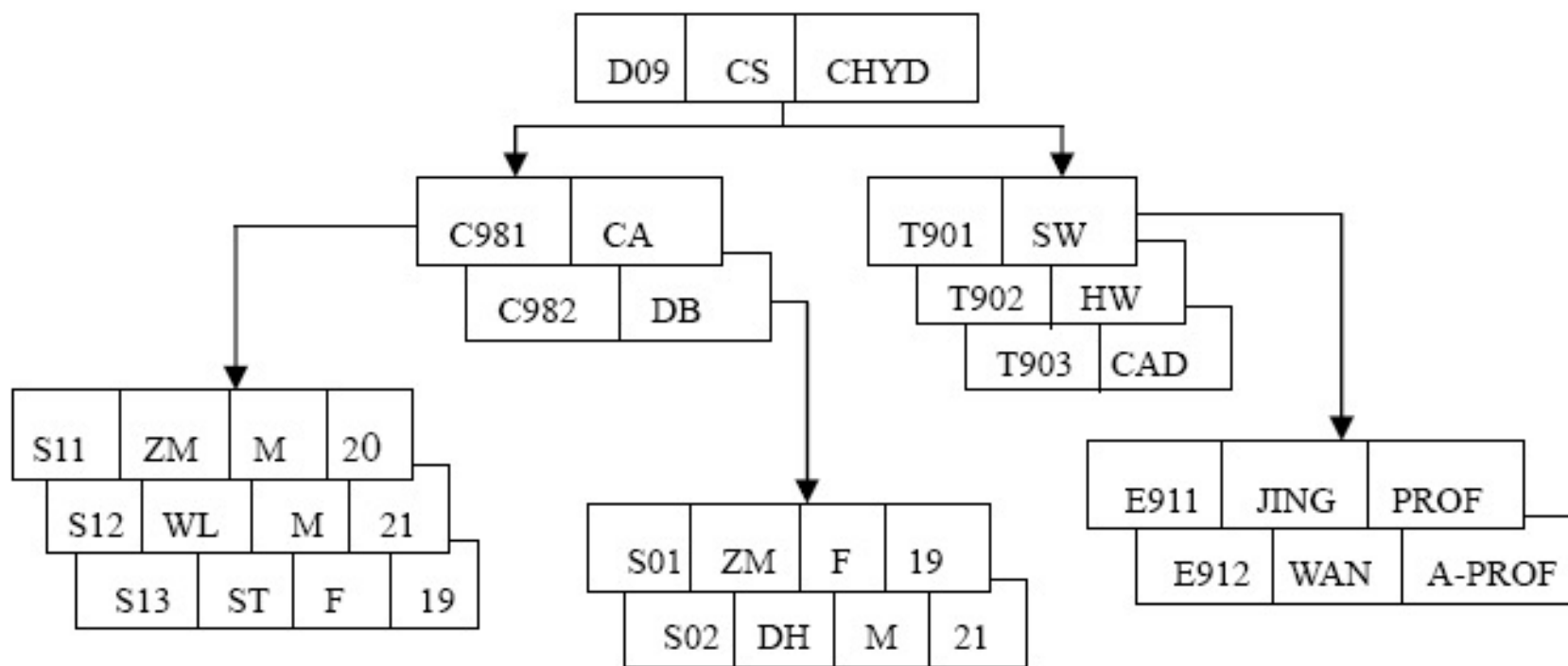


图 2-12 层次模型结构



多对多关系的表示

◆冗余节点法

- ◆结构清晰允许节点改变存储位置
- ◆有数据冗余，存在潜在不一致性

◆虚拟节点法

- ◆减少存储空间
- ◆但可能需要移动指针的操作

层次模型的数据操纵和约束

- ◆没有双亲不能进行插入操作
- ◆删除双亲则子女节点也同时删除
- ◆修改需要修改所有相应的记录

层次模型的存储结构

◆邻接法

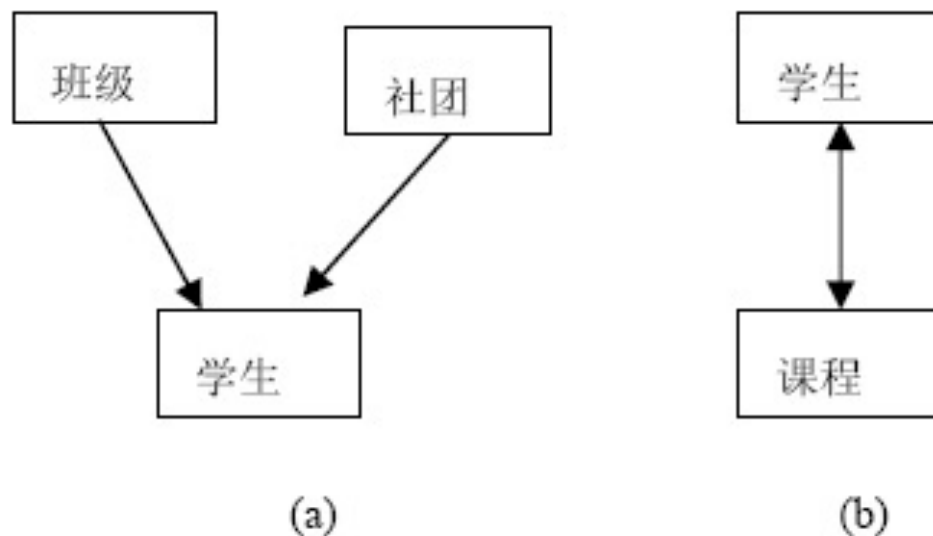
- ◆通过存储的顺序相互关系

◆链接法

- ◆通过指针链接

网状模型

- ◆允许1个以上的节点无双亲
- ◆一个节点可以有多个双亲



网状模型的数据操作和完整性约束

- ◆支持记录码的概念，不允许出现重复值
- ◆保证一个联系中，双亲记录和子女记录之间是1对多的关系
- ◆支持双亲记录和子女记录之间的约束条件

网状模型的存储结构

◆链接法

- ◆单向链接
- ◆双向链接
- ◆环状链接
- ◆向首链接

网状模型的优缺点

◆优点

- ◆很好的表示多对多的关系
- ◆性能较高

◆缺点

- ◆结构复杂
- ◆DDL DML语言复杂

关系模型

- ◆目前最重要的数据库数据模型，关系型数据库采用关系模型做为数据的组织形式
- ◆1970年由E.F.Codd首次提出，因此获得1981年图灵奖
- ◆本课程重点也是关系型数据库

关系型数据库的数据结构

- ◆建立在严格关系运算基础上的，在关系数据库中数据的逻辑结构是一张二维表

学生登记表

学 号	姓 名	年 龄	性 别	系 名	年 级
95004	王小明	19	女	社会学	95
95006	黄大鹏	20	男	商品学	95
95008	张文斌	18	女	法律学	95
...

关系数据库的基本概念

- ◆ 关系(**relation**) 对应一张二维表
- ◆ 元组 (**tuple**) 表中的一行称为一个元组
- ◆ 属性 (**attribute**) 表中的一列称为一个属性，属性的名称叫属性名
- ◆ 主码 (**key**) 某个属性组，可以唯一的确定一个元组
- ◆ 域 (**domain**) 属性的取值范围
- ◆ 分量：元组的一个属性值
- ◆ 关系模式：对关系的描述 关系名 (属性1, 属性2, 属性3)

实体和实体联系的表示方法

- ◆在关系模型中实体和实体联系的表示方法也是用关系来表示的：
- ◆学生（学号，姓名，年龄。。。）
- ◆课程（课程号，课程名，学分）
- ◆选课（学号，课程号，成绩）

关系模型的基本要求

- ◆关系中的每个分量不可再分，即不允许表中还包含表

职工号	姓名	职 称	工 资			扣 除		实 发
			基 本	工 龄	职 务	房 租	水 电	
86051	陈 平	讲 师	805	20	50	60	12	803
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

关系模型的操纵和完整性约束

- ◆关系模型中数据操纵包括：查询，插入，删除，和修改，所有的操纵是针对关系表进行
- ◆关系的完整性主要包含以下三个方面
 - ◆实体完整性
 - ◆参照完整性
 - ◆用户定义完整性

关系模型的存储结构

- ◆在物理存储中，数据以表格的形式存在
 - ◆实体
 - ◆实体和实体之间的联系

关系型数据库的优缺点

◆ 优点

- ◆ 理论基础完善
- ◆ 模型单一，无论实体和实体关系统一用表的形式表示
- ◆ 一体化的数据子语言
- ◆ 存取方式对用户透明，具有更好的数据独立性
- ◆ 面向集合的存取
- ◆ 有利于应用的扩展

◆ 缺点

- ◆ 查询效率较低
- ◆ 需要用户对查询进行优化

面向对象的模型

- ◆数据库技术和**OOP OOA**技术结合产物
- ◆数据单元可以是自己定义的具有复杂结构的对象
- ◆数据与在数据上的操作封装在一起
- ◆支持数据之间的继承关系