

第三章



空间数据的 采集与管理



第一节

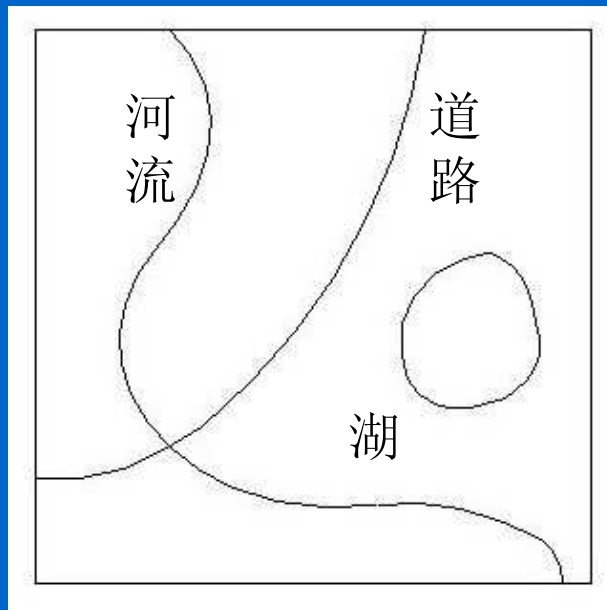
地理信息系统的 数据结构



数据结构的概念

- 数据结构是指数据的组织形式，是适合于计算机存储、管理和处理的数据逻辑结构。
- 地理数据相当复杂，必须包括坐标位置、可能的拓扑关系、地物的特征等。

GIS的数据结构



基于矢量(图形)的数据结构

		2			1			
			2		1			
		2			1			
		2			1			
	2				1	3	3	
	2			1		3	3	
	2		1			3	3	
	1	2						
1			2	2	2	2	2	
								2

基于栅格(图像)的数据结构

点没有大小 点怎样表示? 用固定大小的面单元表示点

栅格数据结构(栅格模型)

- 栅格结构是最简单、最直观的空间结构，来自遥感、数字摄影、扫描地图等方面的数据都是基于栅格形式的。
- 在栅格图像中，地理实体的位置和状态是由其占据的栅格行、列定义的。栅格单元也称为像素。栅格单元的大小代表所定义空间的分辨率。
- 栅格模型的空间数据和属性数据储存为一体。

栅格数据结构(栅格模型)

- 多个栅格能够方便地互相叠合，是栅格模型的一个明显优点。
- 不管栅格单元多么小，总会使信息丢失。
- 缩小单个栅格单元的面积，能够提高分辨率从而提高了数据精度。但同时因单元个数增加而使数据量急剧增加，并且影响计算机的处理速度。

矢量数据结构 (矢量模型)

- 矢量数据模型是通过记录坐标的方式，用点、线、面等基本要素尽可能精确地表示各种地理实体的一种数据抽象方法。
- 矢量数据表示的坐标空间是连续的，因此可以精确地定义地理实体的空间位置、长度、面积等。

矢量模型的点对象

- 空间中不可再分的地理对象
- 没有大小、形状，但可以用一定大小、有特定形状的点状符号绘制到地图上
- 点对象在矢量的数据结构中，除了点的坐标以外，还应当储存一些与点对象有关的属性数据来描述，例如点对象的类型、制图符号、显示要求等

矢量模型的线对象

- 用一串有序的坐标表示，也可以看成由若干小线段来逼近一条曲线
- 描述线对象需要储存其相应的属性数据，如各线段两端点的坐标、显示符的线型等
- 对于需要精确描述、小线段特别多的线对象，可能还需要在属性记录中加上一个指示字，以便启动相关的内插函数来加密数据点

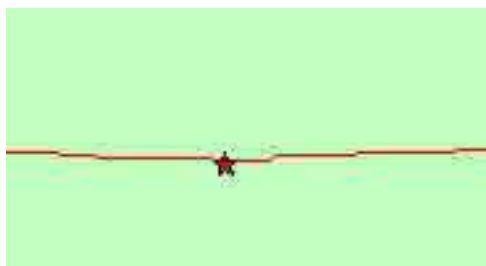
矢量模型的面对象

- 由封闭的线链组成，其数据结构需要记录线链各顶点的坐标，注意第一点和最后一点的坐标相同
- 在需要记录相关属性的同时，更重要的是要记录其空间的拓扑关系，即形状、邻域、层次结构等。

矢量模型的拓扑结构

- 拓扑学 (Topology) 是几何学的一个分支，主要研究在拓扑变换下能够保持不变的几何属性——拓扑属性。
- 拓扑变换在各类空间研究中有着广泛的应用。

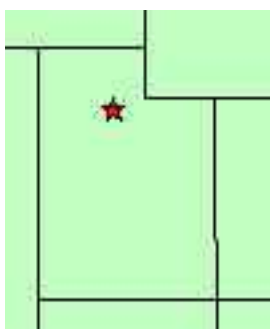
矢量模型的拓扑结构



点—线



线—线



点—面



线—面



面—面

矢量模型的拓扑结构

- 地理空间研究主要用到下面的拓扑概念：
 - 连接性
 - 多边形区域定义
 - 邻接性
 - 方向性
 - 包含性

空间地物的相互关系

- 两个相邻的面对象的公共边界是否需要重复输入是矢量模型的特殊问题。
- 采用环状数据结构，公共边界需要重新输入，修改边界比较麻烦，而且不易保证图形数据的质量。
- 采用拓扑型的数据结构，公共边界只需要输入一次，移动、修改都很方便，容易检验数据的质量。
- 拓扑结构可以在线和线相交处通过节点连通；自动判断两个多边形的公共边界的左右侧分别是谁。

两种数据结构的比较

	矢量数据模型	栅格数据模型
优点	数据结构紧密，数据量少 提供有效的拓扑编码 图形输出美观 数据精度高	数据结构简单 叠置操作简单 能有效表达空间的可变性 数学模拟方便 技术开发费用低
缺点	数据结构复杂 表达空间变化能力差 数学模拟比较困难 叠置分析时难以与栅格图组合	难以表达拓扑关系 图形不够美观 数据量大,降低分辨率时，信息缺失严重
适用对象	大比例、边界明确的事物	小比例、边界模糊的事物



第二节

空间数据的转换



栅格数据和矢量数据的转换

- 矢量结构向栅格结构的转换
 - 转换技术比较成熟
 - 在转换前，一般都要根据要素的属性进行分类
- 栅格数据向矢量数据的转换的基本步骤
 - 多边形边界点提取
 - 边界线追踪
 - 去除多余点并将曲线光滑
 - 拓扑关系生成



空间坐标的转换

- 平面坐标系的变换
 - 平移
 - 缩放
 - 旋转
- 坐标系的投影变换

地图投影

- 由于要把地球曲面展开成地图平面，又要顾及面积、距离、角度等与实际相符，因此必须使用投影理论制作地图
- 地图投影是在球面和平面之间建立点与点之间函数关系的数学方法
- 由于地球是不规则的椭球，不同地区往往根据各地的具体情况采用不同的地球定义数据和不同的大地参照系

地图投影

- 用于地图绘制上的投影方法
 - 等角投影(墨卡托投影)
 - 等积投影
 - 高斯—克吕格投影
 - 等差分纬度圆锥投影
 - 格灵顿投影
 - 约二十多种
- 不同的投影方法产生的误差各有不同，要根据不同需要选用不同的投影方法去制作地图。

地图投影

- 当根据实际需要生成另一种投影类型的地图时，就要涉及投影变换
- GIS软件一般都配备有根据投影转换数学模型执行投影转换的功能
- 注意投影转换产生的误差

我国地形图采用的投影方案

地形图比例	投影方案
1:10000 和大于 1:10000	3°分带的高斯—克吕格投影
1:25000 ~ 1:500000	6°分带的高斯—克吕格投影
1:1000000	双标准纬线等角圆锥投影



第三节



空间数据的 来源与特点

数据来源

- 获取信息是建立地理信息应用系统的基本步骤，非数字化的信息必须转换成数字媒体才能被计算机接受。
- 地理信息系统数据的主要来源
 - 图形数据源
 - 属性数据源

数据来源

- 图形数据源
 - 地图
 - 野外实地测量
 - 摄影测量与遥感
- 属性数据源
 - 现场考察
 - 社会调查
 - 统计数据
 - 文字报告和文件
- 利用已有资料

数据来源类别

- 基础制图数据
- 自然资源数据
- 统计调查数据
- 数字高程数据
 - 数字地形模型(Digital Terrain Model, DTM)
 - 数字高程模型(Digital Elevation Model, DEM)
- 文档数据
 - 元数据文档、元数据标准
 - 与GIS有关的国家标准
- 已有的系统数据

数据特点

- 以统一的坐标系统进行空间定位
- 空间数据(定位数据)与非空间数据(属性数据)相结合
- 空间数据具有时效性
 - 沿着时间序列的变化
 - 定位数据不变, 属性数据发生了变化
 - 属性数据不变, 定位数据发生了变化
 - 定位数据和属性数据都发生了变化
- 空间信息的层次性

数据质量与数据精度

- 判断空间数据质量应根据其用途确定其标准。
- 评价空间数据质量的代价随空间数据精度而变化。

数据质量与数据精度

- 影响空间数据质量的微观因素
 - 定位精度
 - 属性精度
 - 逻辑的不一致性
 - 分辨率
- 影响空间数据质量的宏观因素
 - 完整性
 - 时限性
 - 适用性
 - 数据档案完整性
- 影响数据精度的人为因素



第四节

数据的采集、组织、 加工与整理



数据采集

- 数字化 通过图形—数字转换装置将各类地图图形离散化为数据
- 规范化 对于不同来源的数据，统一坐标、统一记录格式
- 数据编码 根据数据结构和目标属性特性，进行编码以方便存贮
- 数据存储 将数据记录在计算机的存贮介质中
- 数据编辑 GIS系统软件为用户提供修改、增加、删除等更新数据的方法

空间数据的输入

- 数字化仪输入
- 扫描仪输入
- 直接从测量仪器获取数据
- 外部数据文件的输入

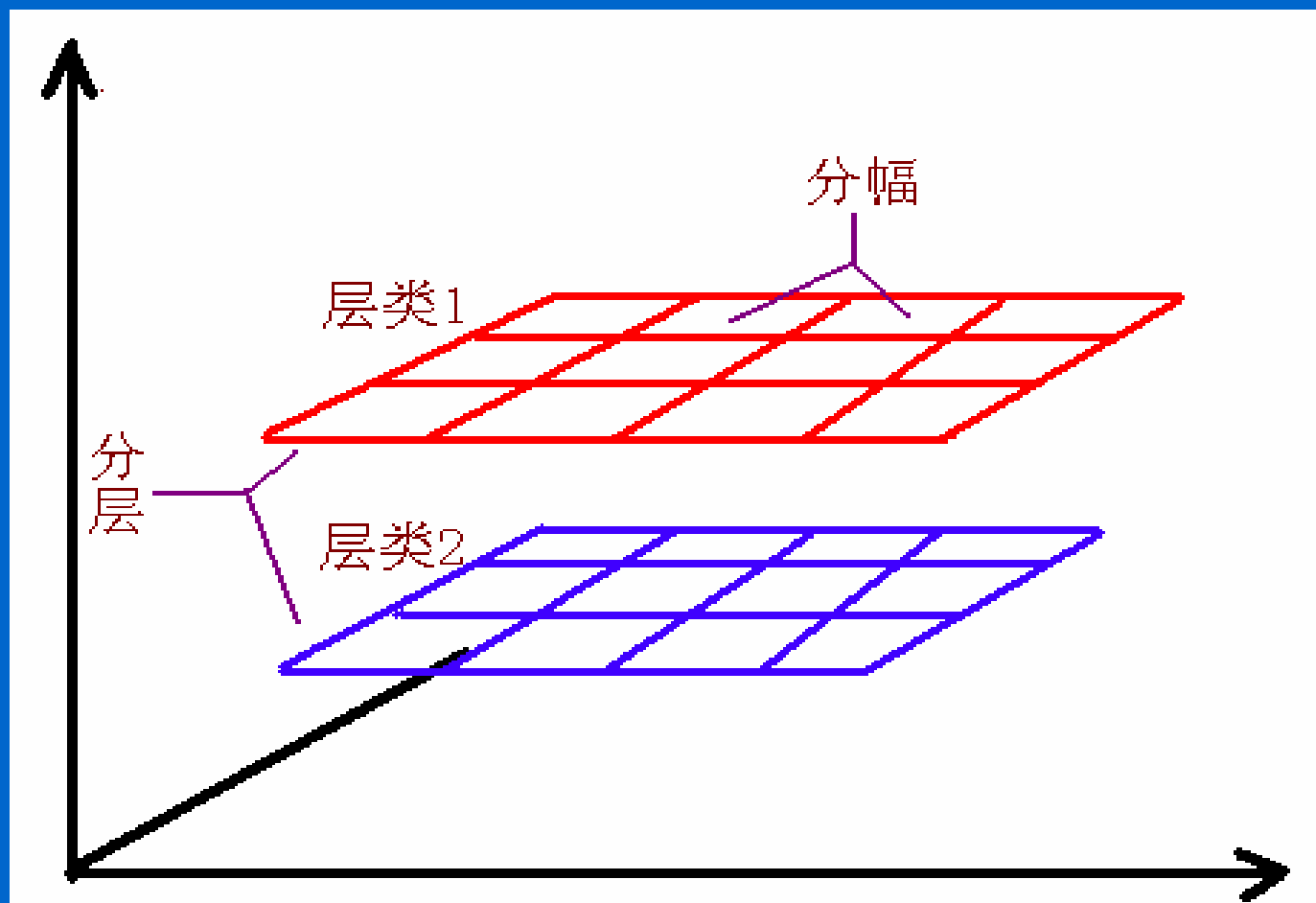
属性数据的输入

- 属性是对地理目标的数量和质量的描述指标
 - 属性数据即空间实体的特征数据，一般包括名称、等级、数量、代码等多种形式。
 - 直接记录在栅格或矢量数据文件中的属性数据
 - 输入数据库存储为属性文件的属性数据
- 属性数据一般由关系数据库来管理
- 属性数据和空间数据的连接要通过属性数据库和空间数据库中的共同字段来建立

空间数据的组织

- GIS应用系统基础数据的特点
 - 大量的基础数据在建立系统时必须录入
 - 基础数据中含有大量的空间数据(图形数据)
- 空间数据的组织
 - 空间数据分层
 - 空间数据分幅
 - 按自然单元分幅
 - 按行政单元分幅
 - 按规格网分幅

空间数据的组织



属性数据处理要注意的问题

- 定性信息定量化
- 数据的规格化

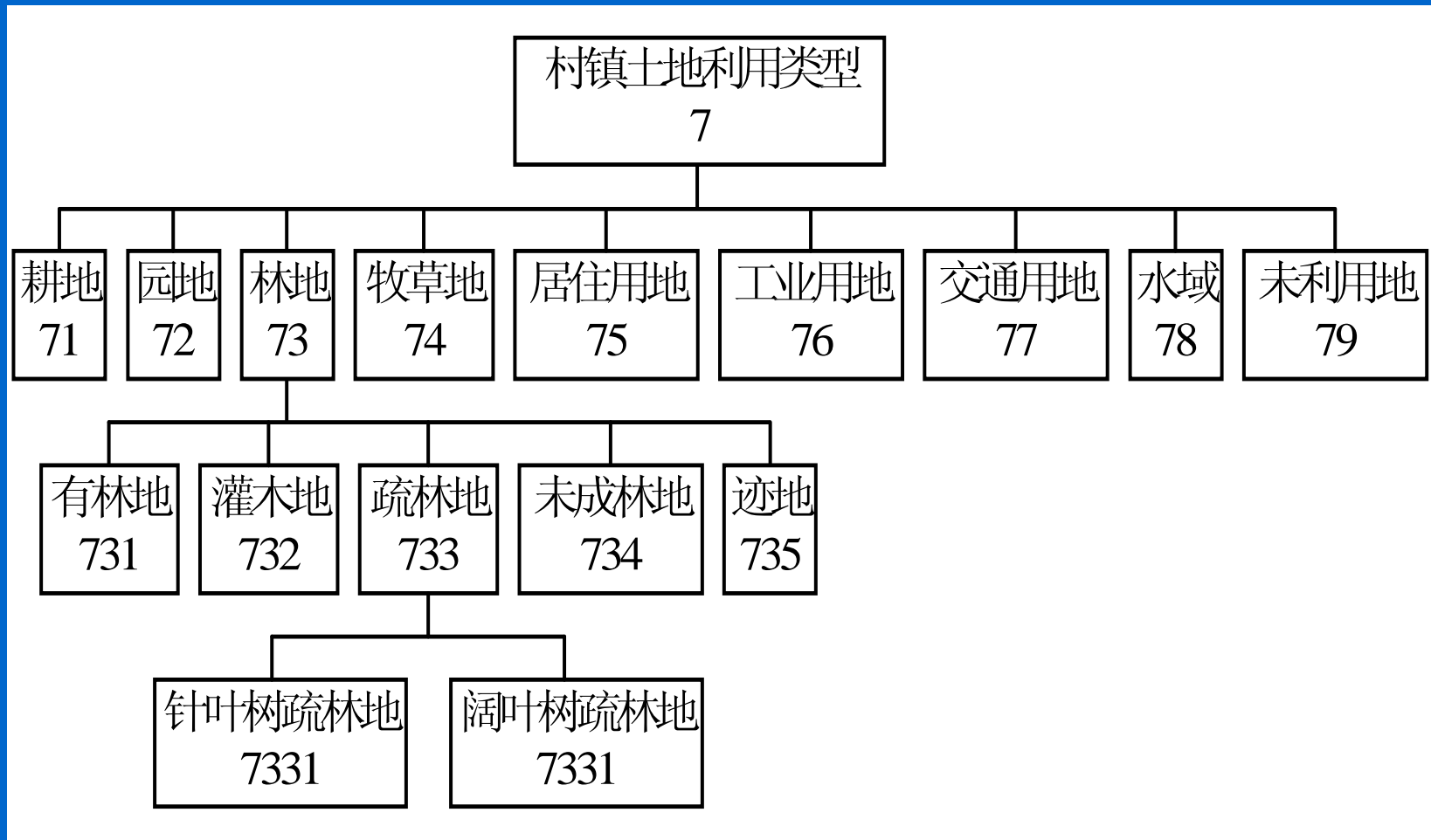
属性数据的分类编码

- 使用特征码法来处理与图形要素有着比较密切的联系属性数据
- 属性数据编码的几个原则
 - 科学性与系统性
 - 标准化、通用性和一致性
 - 完整性与可扩充性
 - 简捷性

常用的编码方法

- 层次分类编码法
 - 以分类对象的从属和层次关系为排列顺序的一种编码方法
 - 优点是能够明确表示出分类对象的类别，代码结构有严格的隶属关系
- 多源分类编码法
 - 对一个特定的分类目标，根据诸多不同的分类依据进行编码，各位数字代码之间并没有隶属关系
 - 优点是信息载量大

层次分类编码法



多源分类编码法

分 类		标 志 编 号					31214				
		I	II	III	IV	V					
河流所在的地貌	平原河	1									
	过渡河	2									
	山地河	3									
水流状况	常年河		1								
	时令河		2								
	消失河		3								
通航状况	通航河			1							
	不通航河			2							
河型	顺直型				1						
	弯曲型				2						
	分汊型				3						
	游荡型				4						
干流或支流	干流：一级					1					
	支流：二级					2					
	三级					3					
	四级					4					
	五级					5					
	六级					6					



第五节

数据库与 数据库管理系统



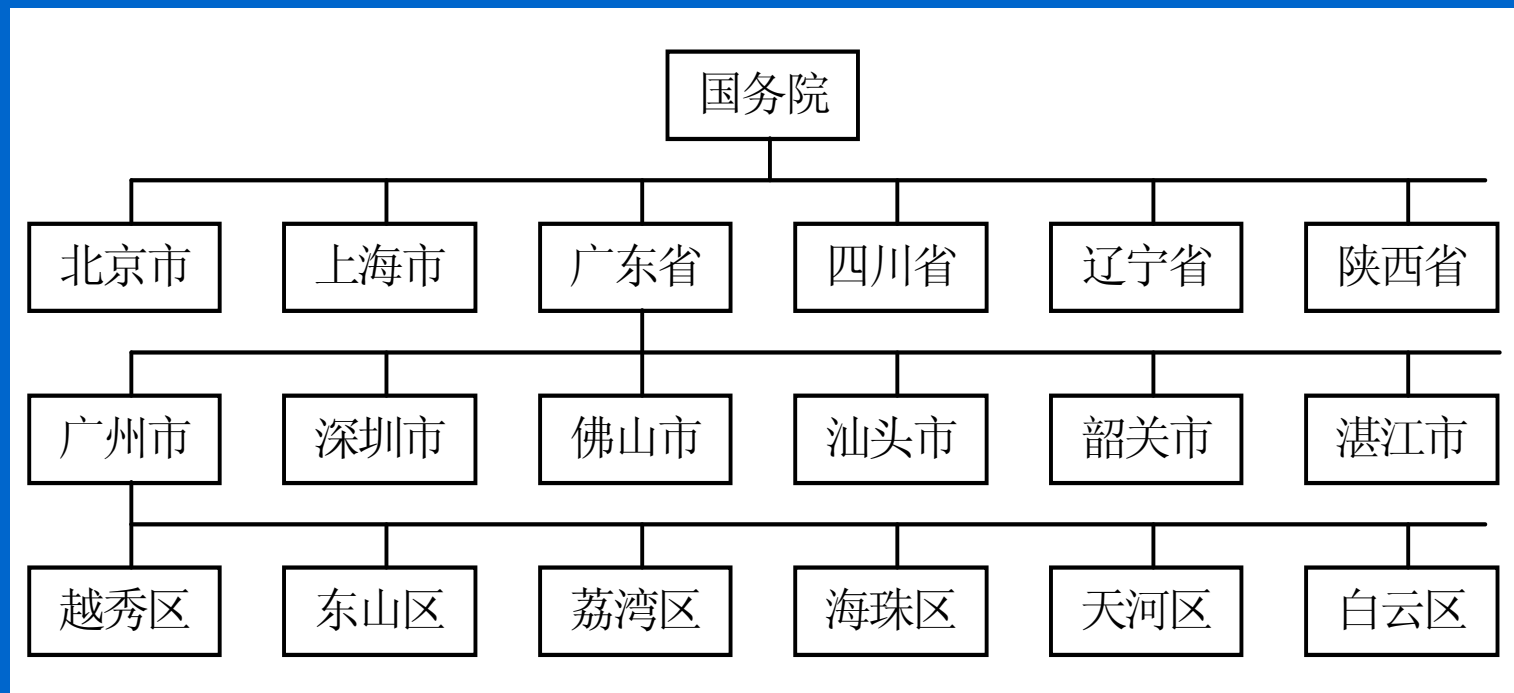
数据库及数据库管理系统

- 数据库 (Data Base) 是以一定的组织方式存储在一起相互关联的数据集合, 能以最佳方式、最少重复(冗余)为多种目的服务的数据组织方式
- 数据库管理系统 (Data Base Management System) 是处理数据存储、进行各种管理的软件系统
- 应用程序对数据库的操作全部通过DBMS进行
- 著名的数据库管理系统主要有: Visual FoxPro、Oracle、System9、SyBase等

数据库中的数据模型

- 层次模型
- 网状模型
- 关系模型

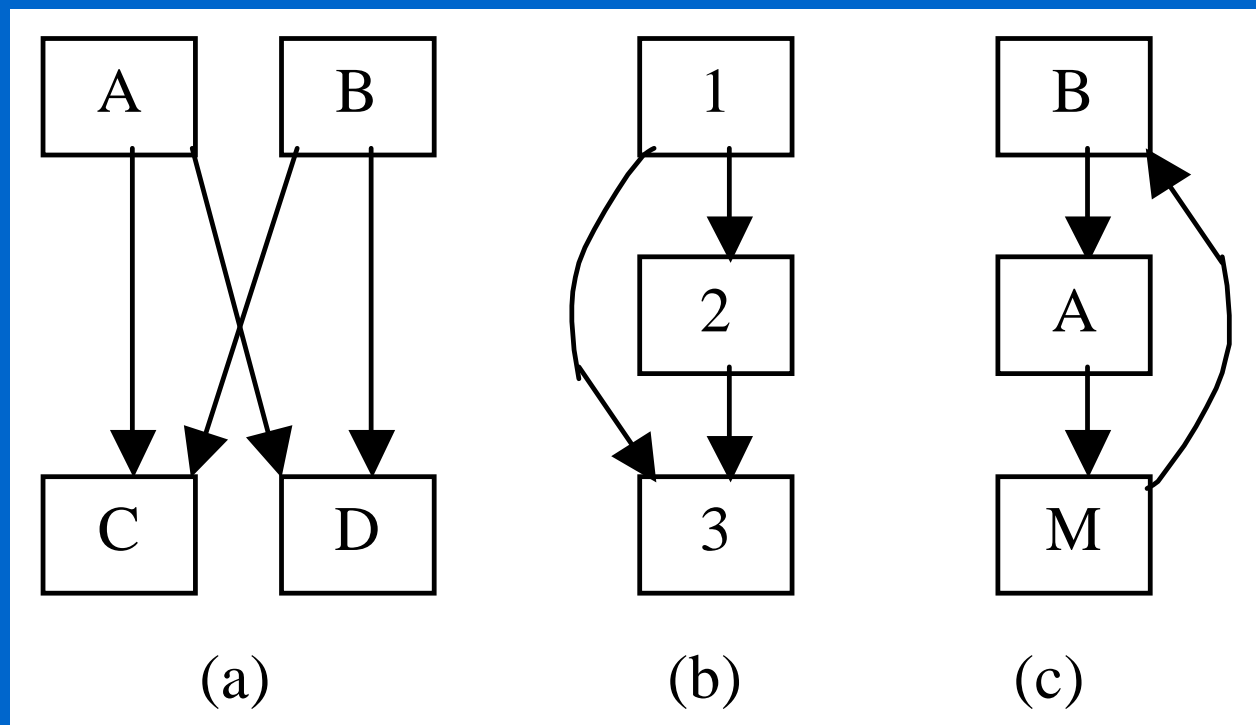
层次模型



层次模型

- 用树形结构(层次结构)来表示实体之间联系的模型称之为层次模型
- 层次模型的基本特征
 - 有而且只有一个结点没有父结点，这个结点被称为根结点
 - 除根结点外，所有结点都有而且只有一个父结点
- 层次模型难以反映地物之间的拓扑关系，无法顾及公共点或者线数据共享等问题

网状模型



网状模型

- 以有向图表示的网络结构，每个结点依然表示数据库中的一个记录类型(实体)
- 网状模型的基本特征
 - 可以有零个或多个结点是没有父结点的
 - 至少有一个结点有多于一个父结点
 - 允许两个结点之间有两种或多种联系

关系模型

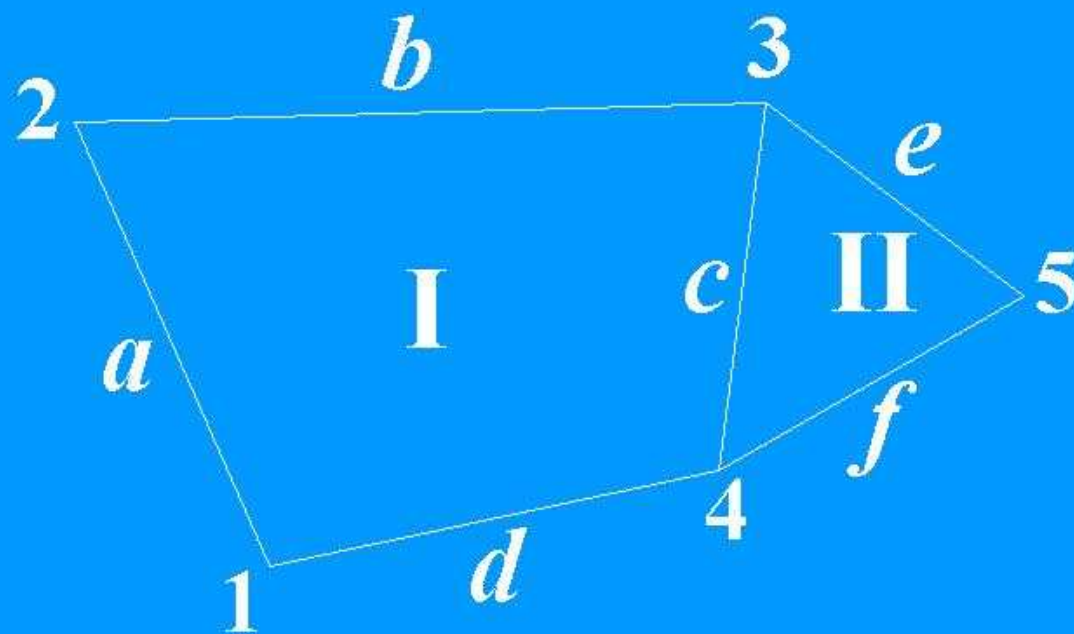
地块编号	土地使用	建筑密度	容积率
6732	住宅	0.4	1.5
7894	商业	0.8	2.1
4605	办公	0.7	3.0
6917	工业	0.3	0.8

关系模型

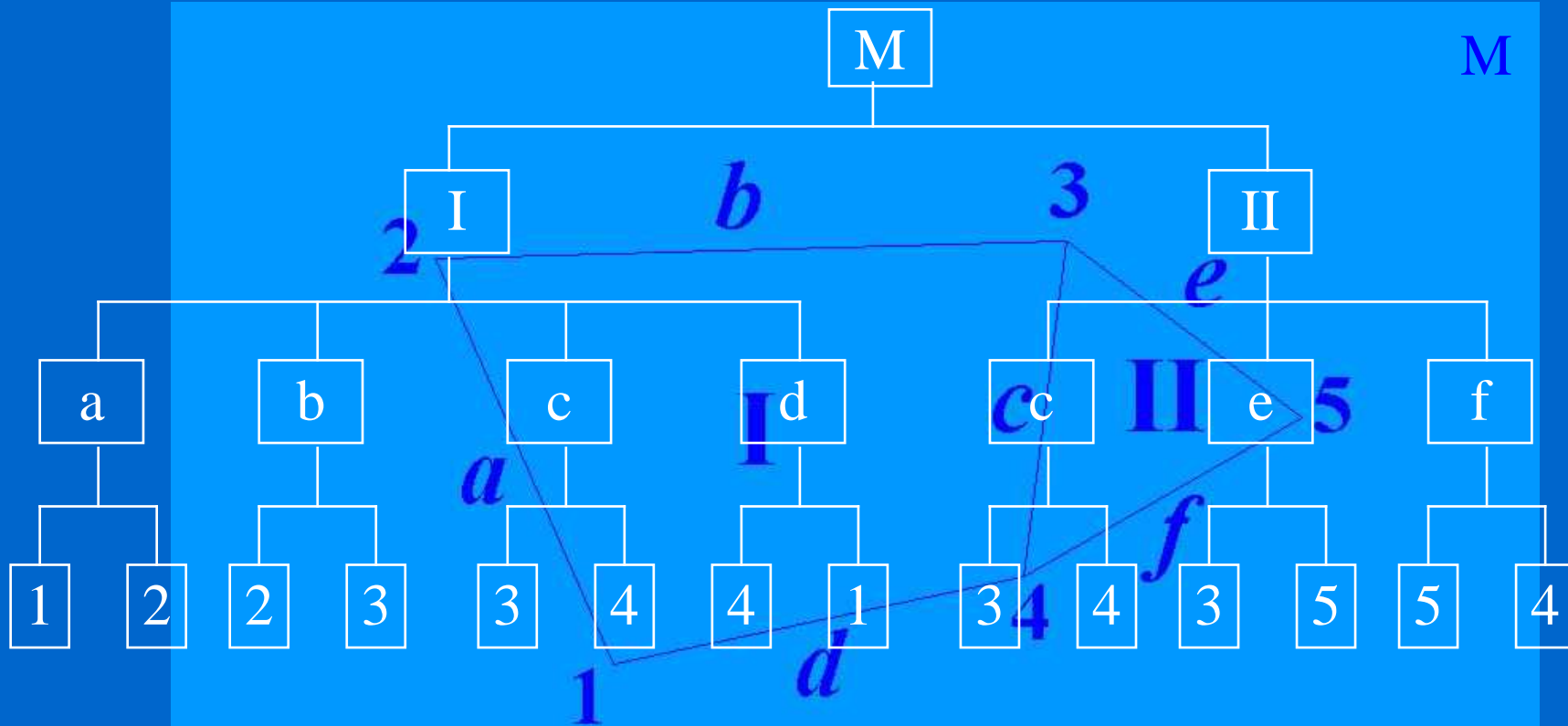
- 用表格数据表示实体之间联系的模型称之为关系模型
- 关系模型的基本特征
 - 用一系列的二维表来描述、储存复杂的客观事物，每一行代表一条记录，每一列代表一种属性(常称为字段)
 - 表中每一列属性都是不能再分的基本字段
 - 各列的字段名必须相异
 - 各行(记录)相异，不允许重复
 - 行列次序无关

地理数据模型应用分析

M

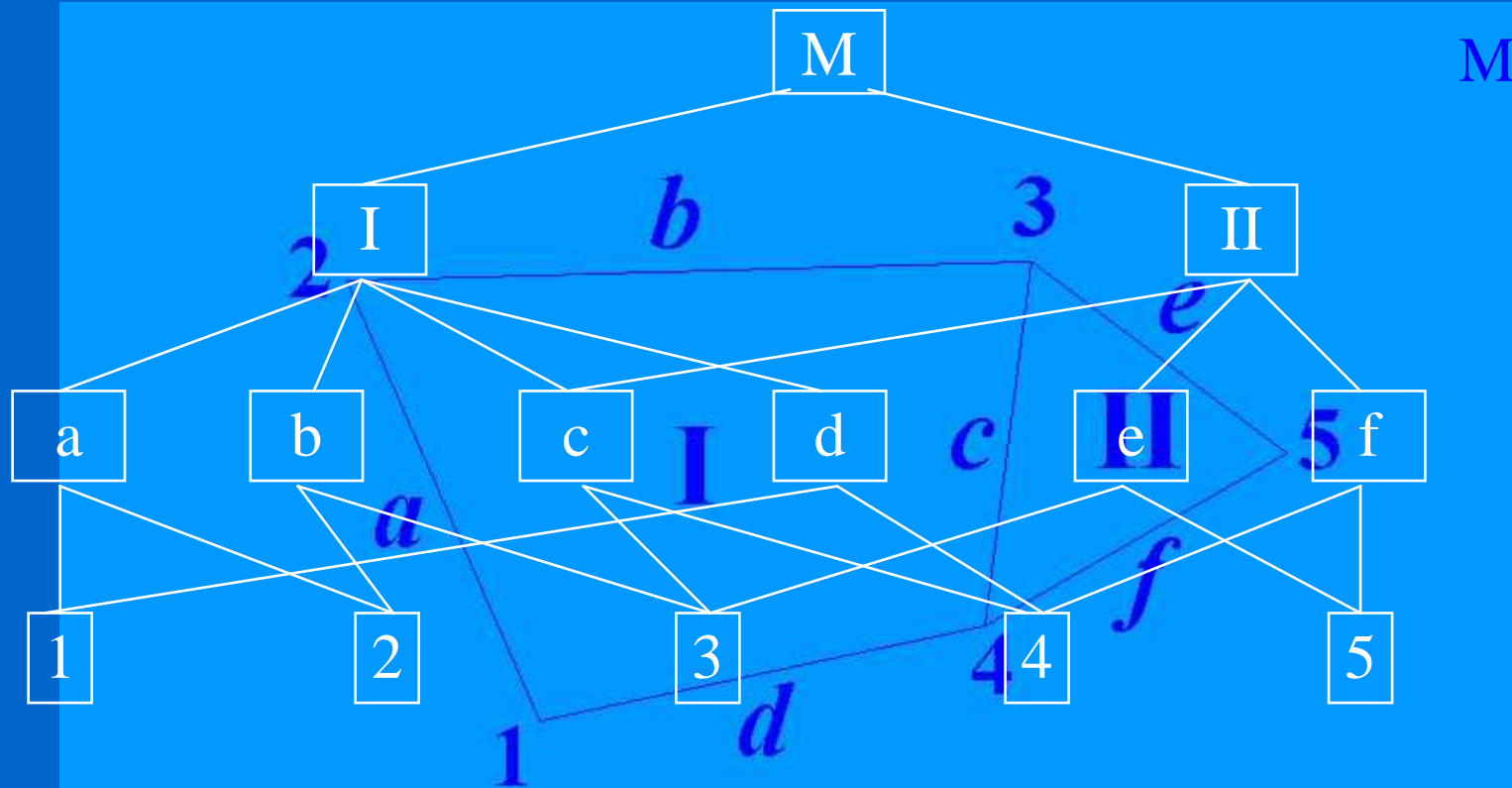


地理数据模型应用分析



层次模型不能表示多对多的联系，难以顾及公共点、线数据共享和实体元素之间的拓扑关系，使数据冗余度增加。

地理数据模型应用分析



网络模型能反映多对多的关系，但结构复杂，增加了查询和定位的困难。

地理数据模型应用分析

编号	边界	边长	边界	起点号	终点号	点号	x坐标	y坐标
I	a	2	a	1	2	1		
I	b		b	2	3	2		
I	c		c	3	4	3		
I	d		d	4	1	4		
II	c		e	3	5	5		
II	e		f	4	5	4		
II	f					5		

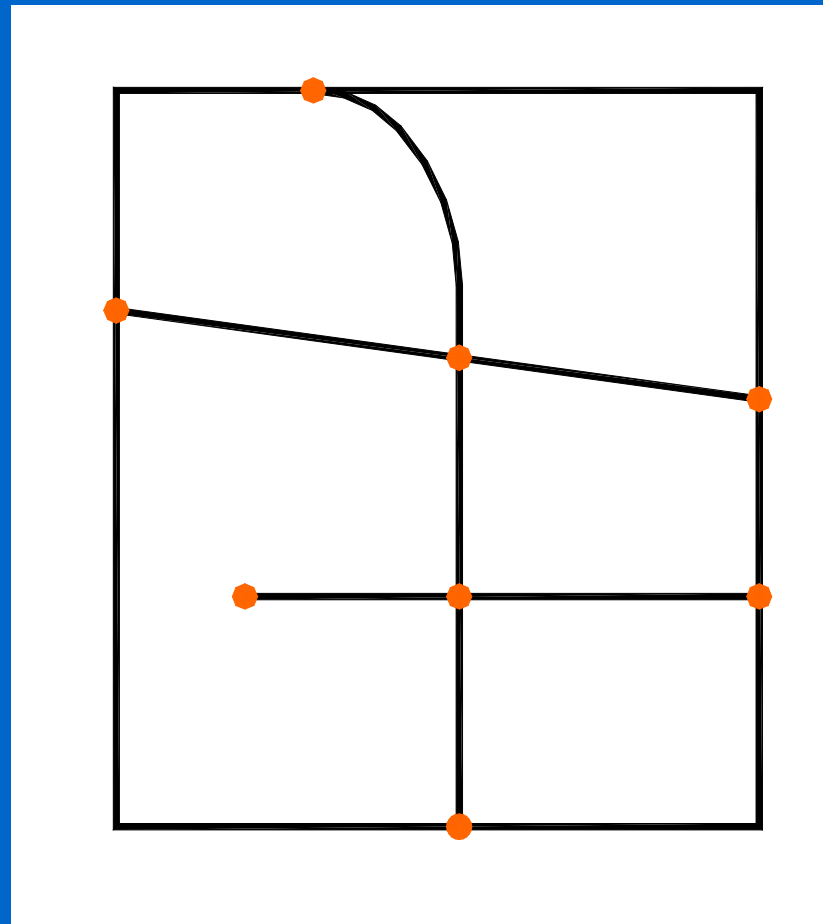
关系模型最大特色是描述的一致性，数据通过公共值表达它们的联系，结构简单、灵活、方便修改、容易维护等。

地理数据模型应用分析

- 关系模型是当前数据库中最常用的数据模型，大部分GIS中的属性数据也采用关系模型
- 关系模型也存在描述对象语义能力较弱、可扩充性较差等问题
- 新一代的数据模型是面向对象模型

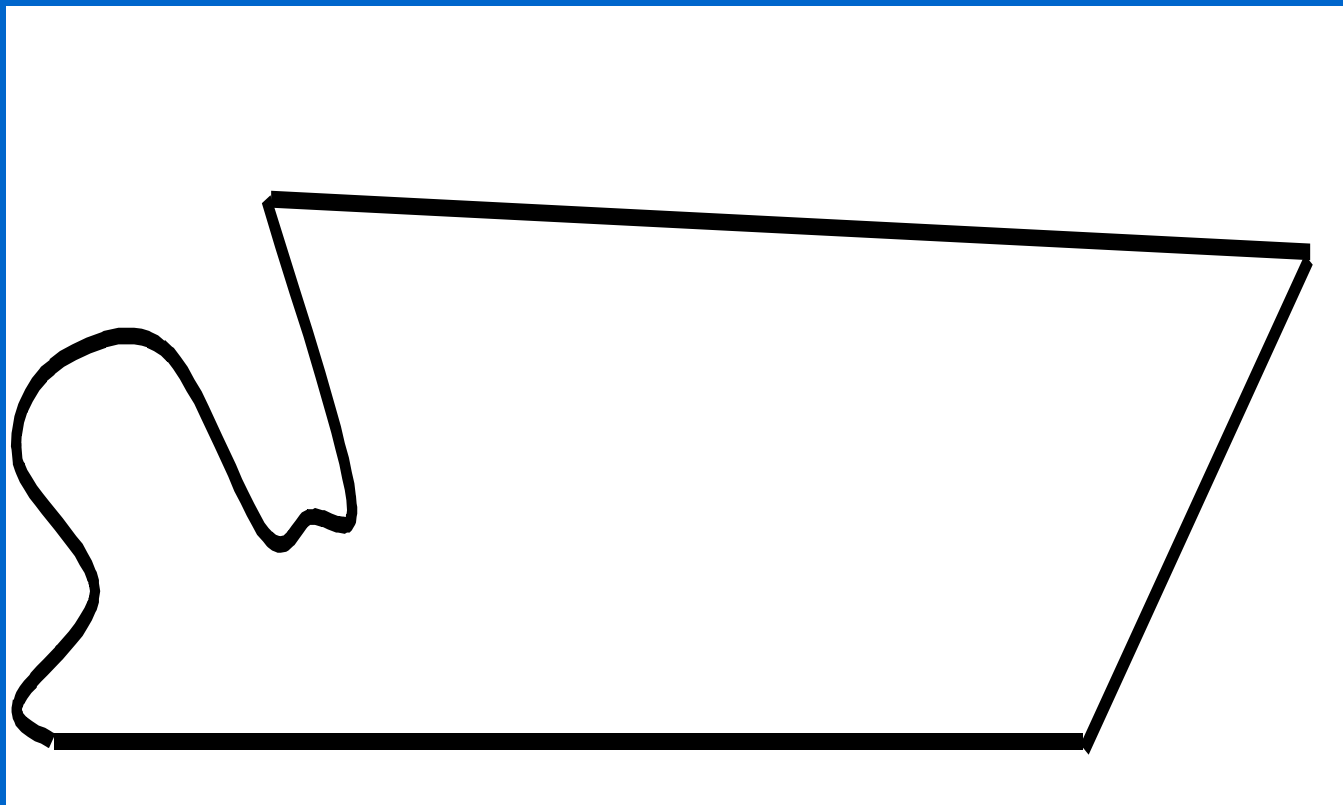


矢量模型的拓扑结构



连接性的表示

矢量模型的拓扑结构



多边形区域定义

矢量模型的拓扑结构

