空间数据库2018秋季

空间数据模型-1018

陈斌

北京大学地球与空间科学学院

gischen@pku.edu.cn

目录

- 数据抽象和数据建模
- 空间信息模型
  - 概念模型
  - 基于场的模型
  - 基于对象的模型
    - 空间数据类型、空间对象关系
    - 两种拓扑关系代数：9IM，RCC

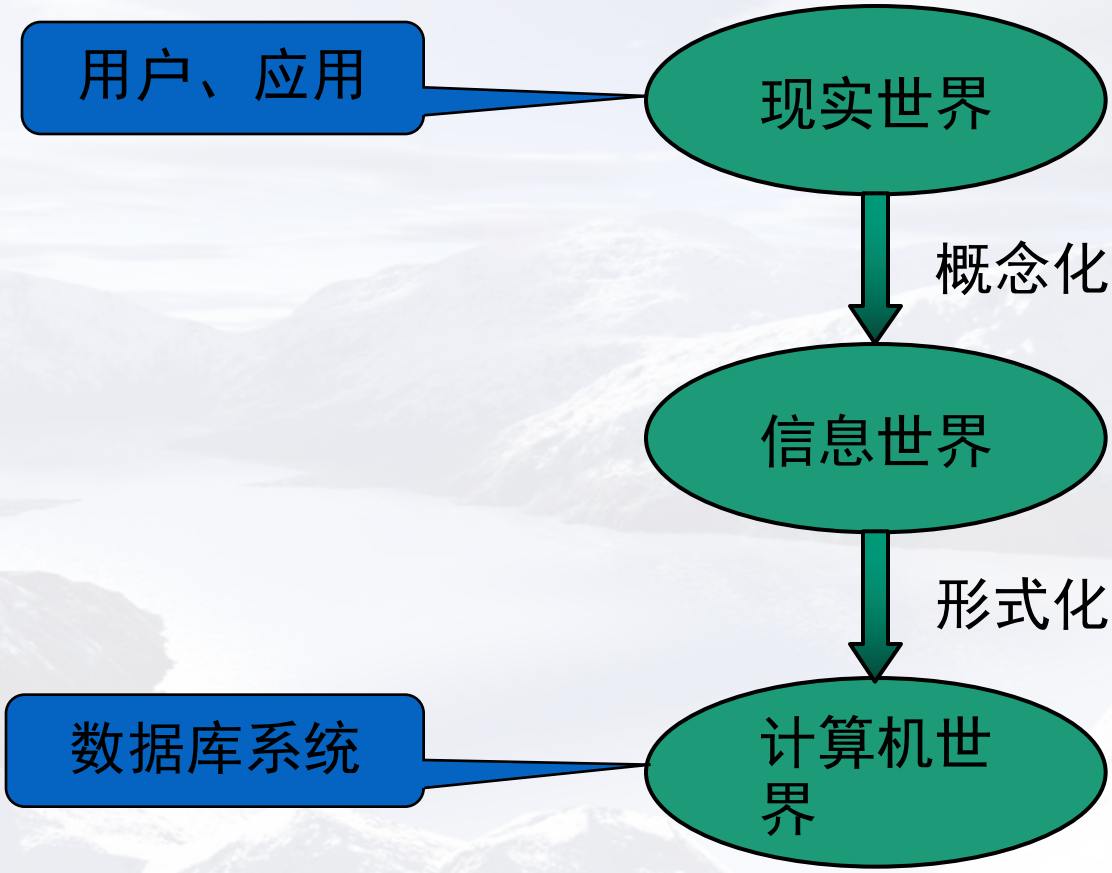
数据库设计的三个步骤

- 概念模型
  - 按用户的观点从现实应用中抽象出事物以及事物之间的联系
- 逻辑建模
  - 建立概念和联系的逻辑结构
- 物理设计建模
  - 对逻辑结构进行具体实现方面的安排和考虑
  - 存储组织、索引、内存管理……

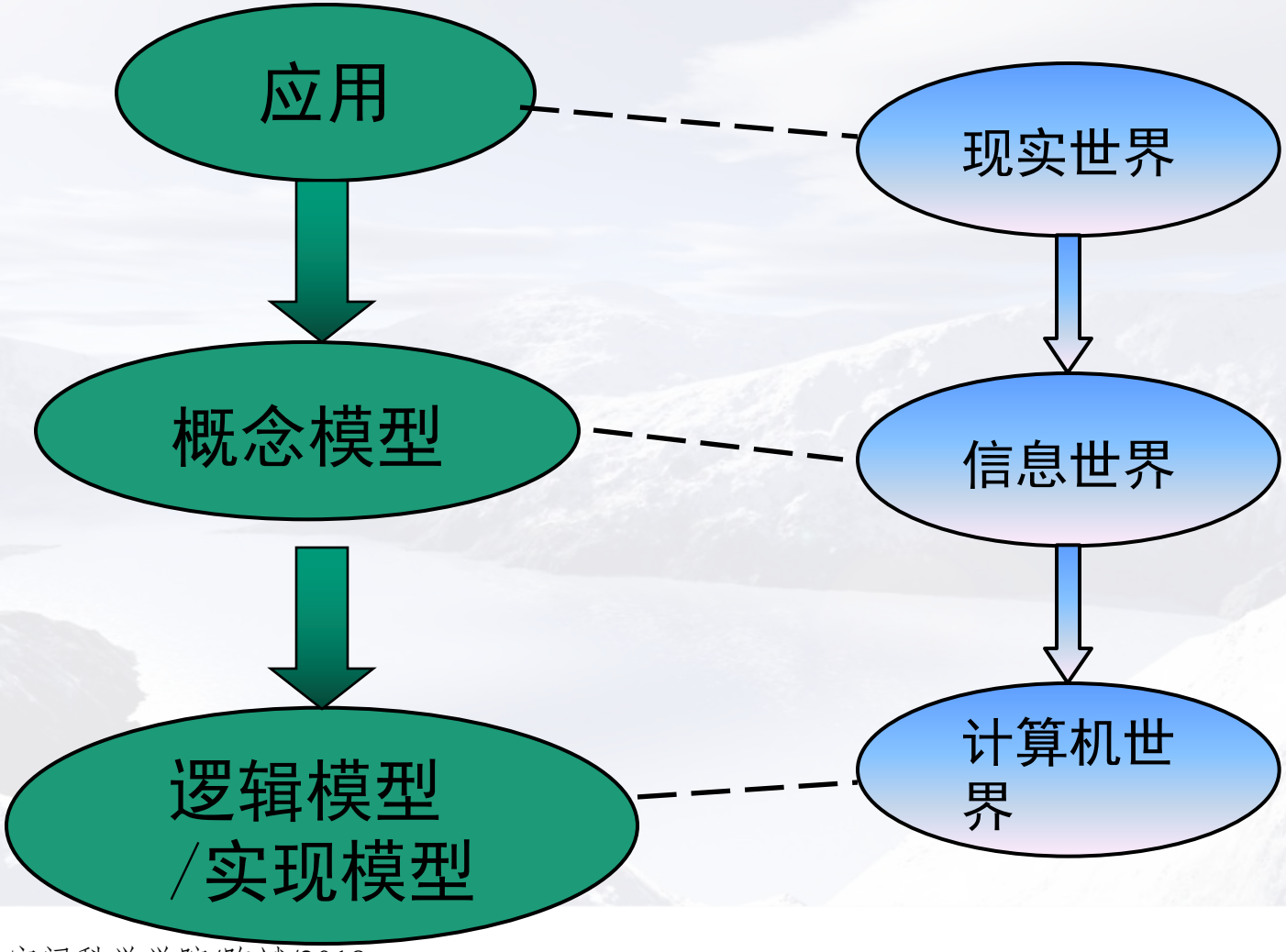
数据抽象

- 现实世界、信息世界、计算机世界
  - 数据库系统是面向计算机的，而应用是面向现实世界的
    - 两个世界存在着很大差异，要直接将现实世界中的语义映射到计算机世界是十分困难的
    - 引入一个信息世界作为现实世界通向计算机实现的桥梁
  - 信息世界是对现实世界的抽象，从纷繁的现实世界中抽取出能反映现实本质的概念和基本关系
  - 信息世界中的概念和关系，要以一定的方式映射到计算机世界中去，在计算机系统上最终实现
- 信息世界起到了承上启下的作用

数据抽象



数据建模



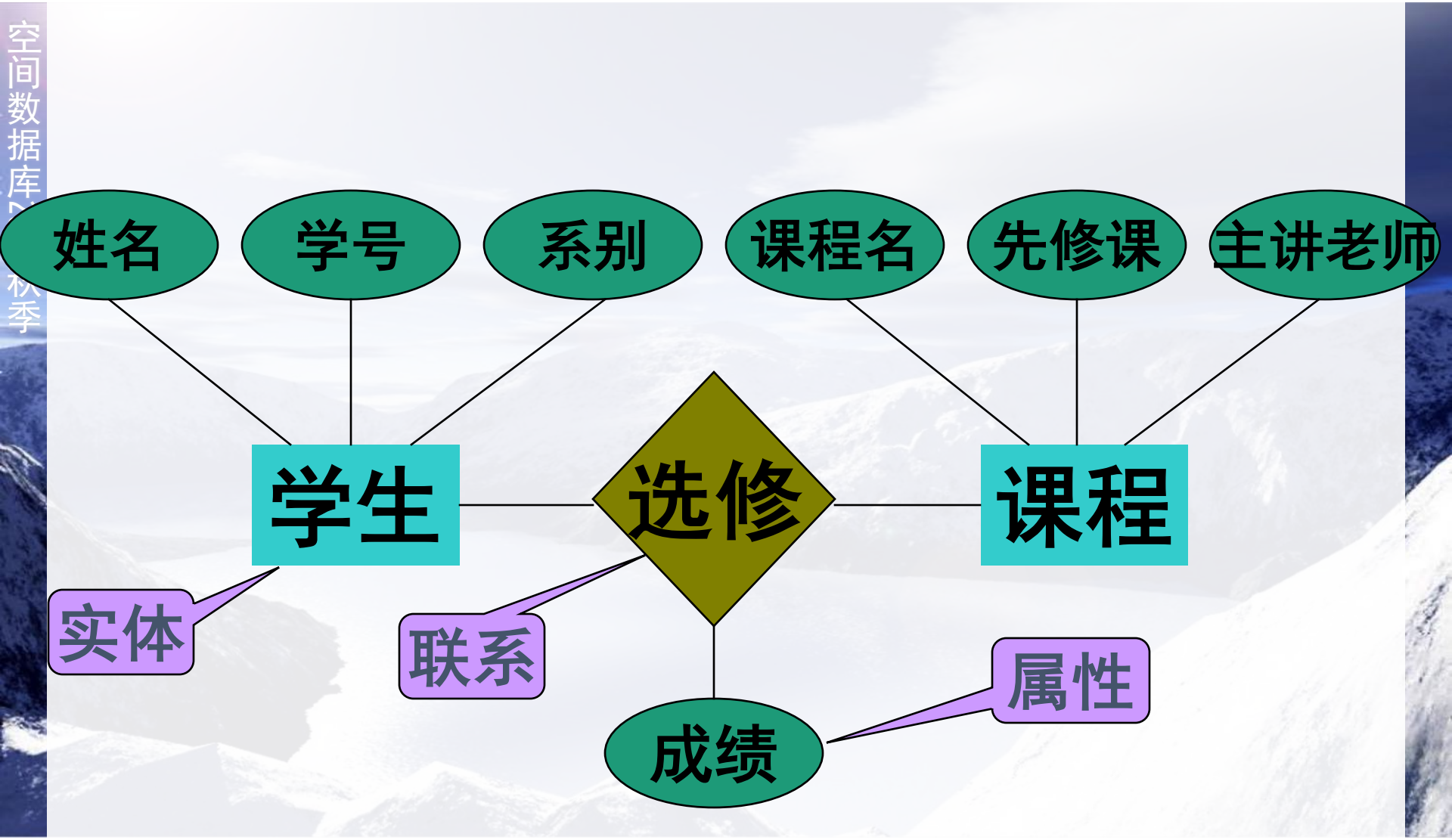
数据建模：概念数据模型

- 按用户的观点来对数据和信息建模
- 用于组织信息世界的概念，表现从现实世界中抽象出来的事物以及它们之间的联系
- 这类模型强调其语义表达能力，概念简单、清晰，易于用户理解
- 它是现实世界到信息世界的抽象，是用户与数据库设计人员之间进行交流的语言
- 如E-R模型、面向对象模型

实体-联系（E-R）模型

- 现实世界被划分为若干实体（entity），由属性（attribute）来描述性质，通过联系（relationship）互相关联
- 实体：现实中或者概念上独立存在的事物或者对象
- 属性：刻画实体性质的数值或描述
- 联系：表达实体间的关联
  - 一对一、一对多、多对多

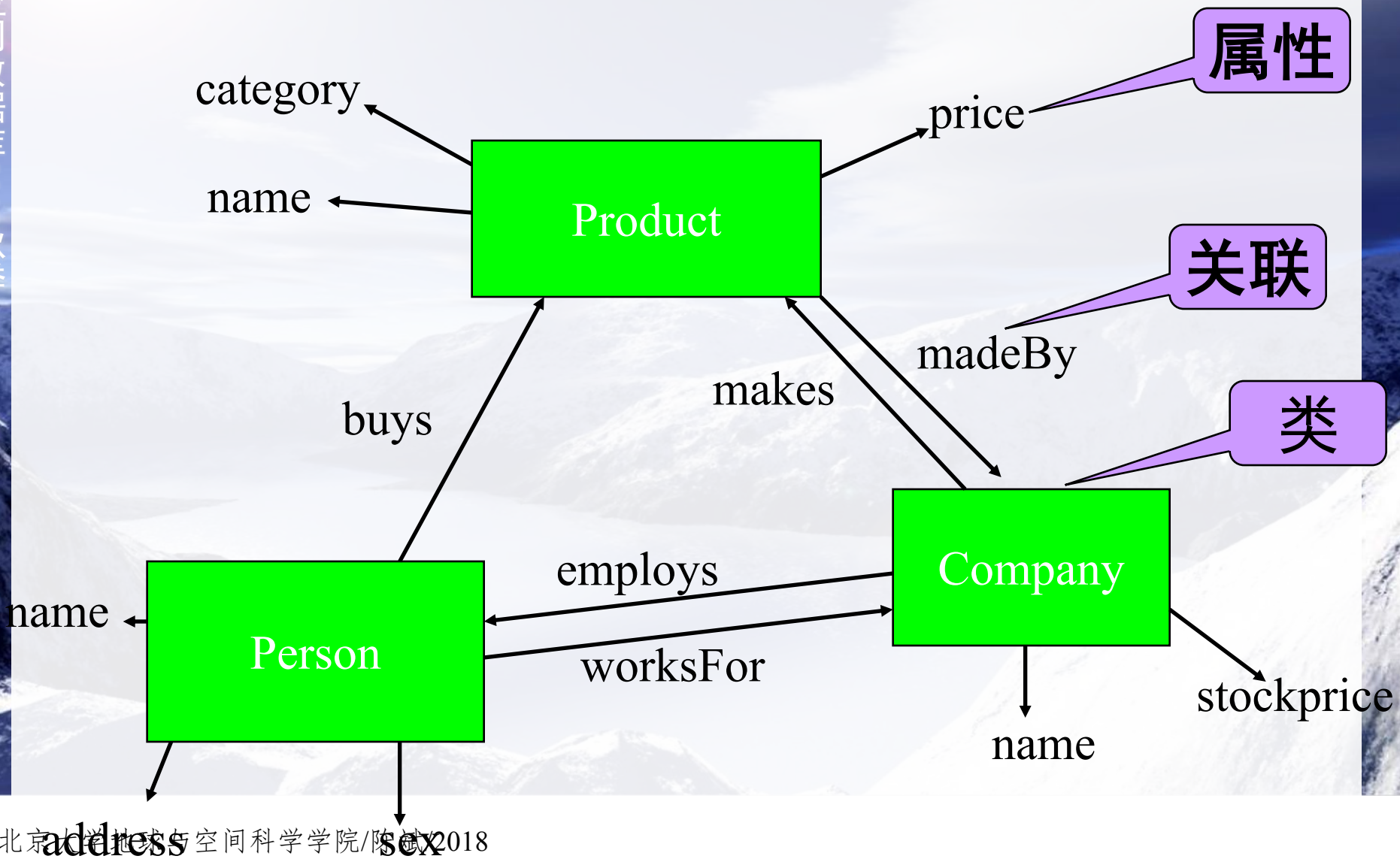
E-R图



面向对象模型

- 现实世界被看作若干对象类（class），由属性（attribute）来描述性质，方法（method）来描述行为，通过关系（relationship）互相关联
- 类：现实中具有相同性质的对象的封装
- 属性：描述对象的性质
- 方法：修改对象的状态，体现对象的功能
- 关系：类之间的关联
  - 聚合aggregation：整体-部分关系
  - 泛化generalization：一般-特殊关系
  - 关联association：其它关系

ODL模型


属性
关联
类
category
name
price
Product
madeBy
makes
buys
Company
employs
worksFor
name
Person
stockprice
name
address
sex

数据建模：逻辑数据模型

- 从计算机实现的观点来对数据建模
- 是信息世界中的概念和联系在计算机世界中的表示方法
- 一般有严格的形式化定义，以便于在计算机上实现
- 如层次模型、网状模型、关系模型、面向对象模型

逻辑模型的三要素：数据结构

- 描述系统的静态特性，即组成数据库的对象类型
- 数据本身
  - 类型、内容、性质。如网状模型中的数据项、记录，关系模型中的域、属性，关系等
- 数据之间的联系
  - 例如网状模型中的系型，关系模型中的外码
- 在数据库系统中一般按数据结构的类型来命名数据模型

逻辑模型的三要素：数据操作

- 描述系统的动态特性，即对数据库中对象的实例允许执行的操作的集合，包括操作及操作规则
- 一般有检索、更新（插入、删除、修改）操作
- 数据模型要定义操作含义、操作符号、操作规则，以及实现操作的语言

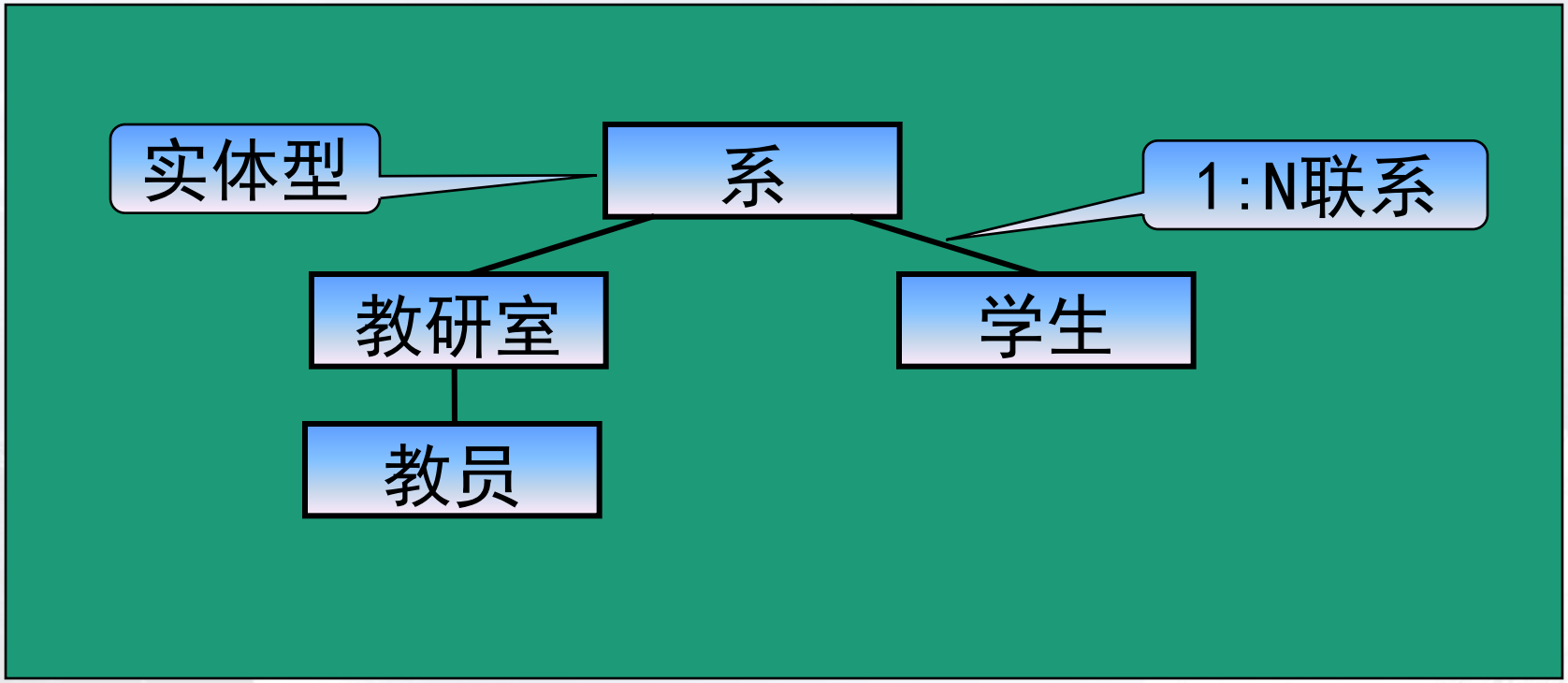
逻辑模型的三要素：约束条件

- 数据的约束条件是完整性规则的集合
- 规定数据库状态及状态变化所应满足的条件
- 以保证数据的正确、有效、相容

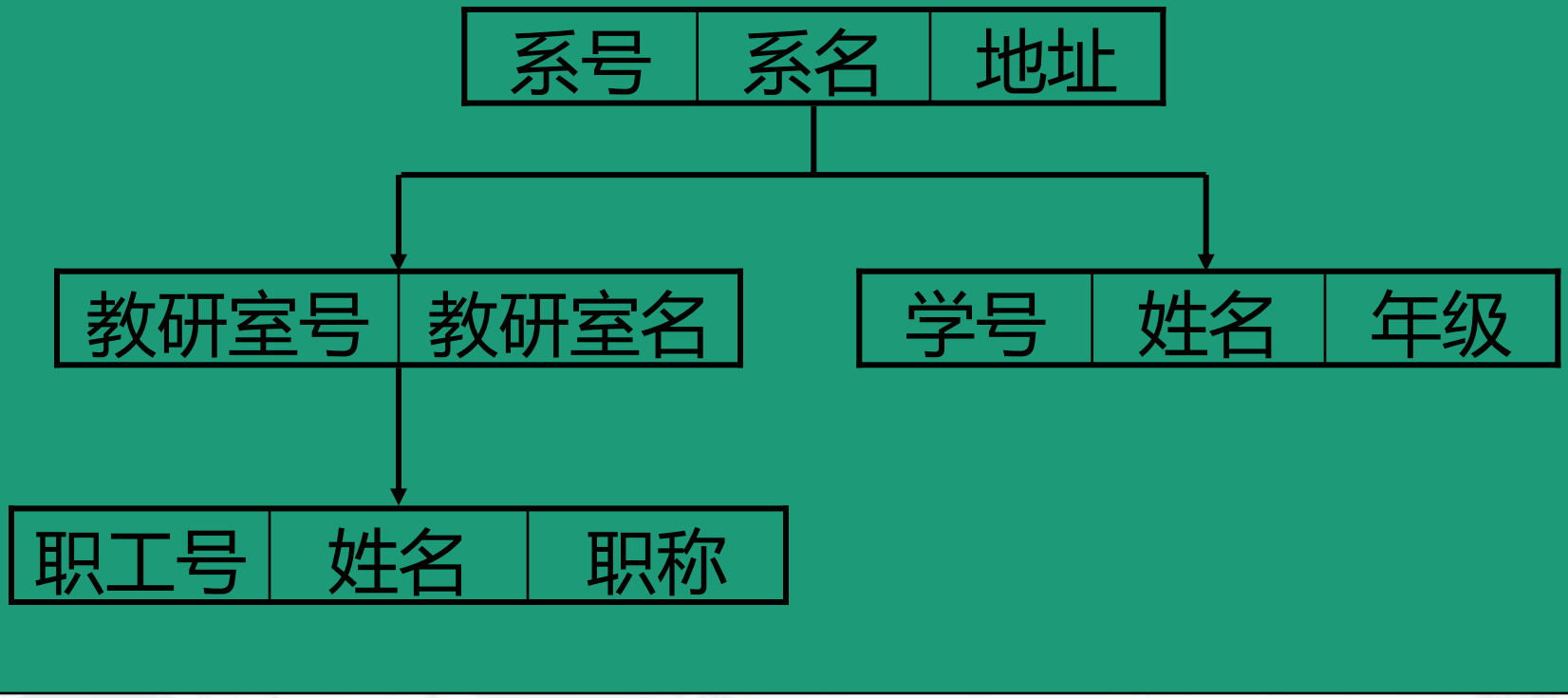
逻辑模型示例-层次模型

- 用树结构表示实体之间联系的模型叫层次模型
- 树由节点和连线组成
  - 节点代表实体型
  - 连线表示两实体型间的一对多联系
- 树的特性
  - 每棵树有且仅有一个节点无父节点，称为树的根
  - 树中的其它节点都有且仅有一个父节点

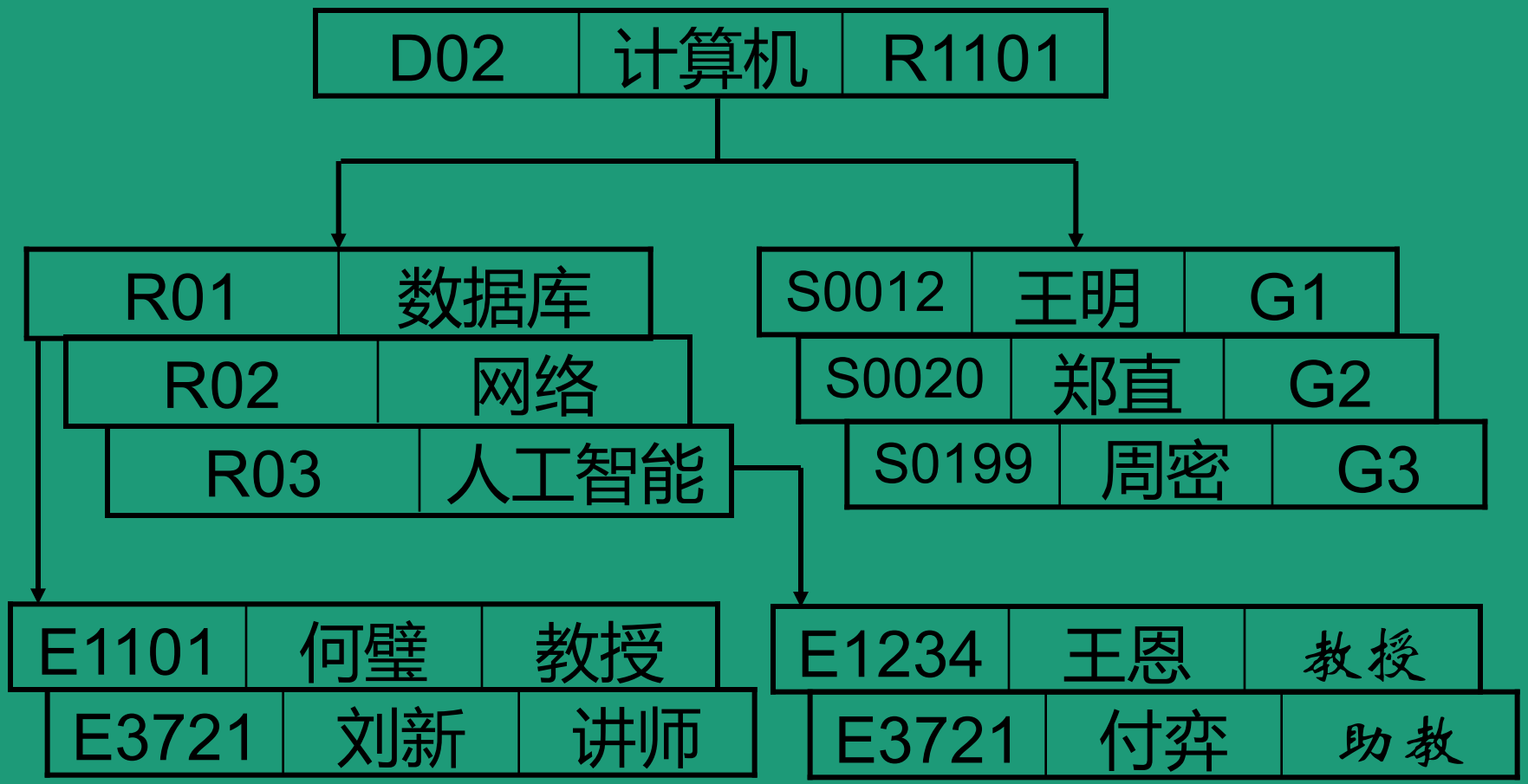
层次数据结构：实体和联系



层次数据结构：数据类型



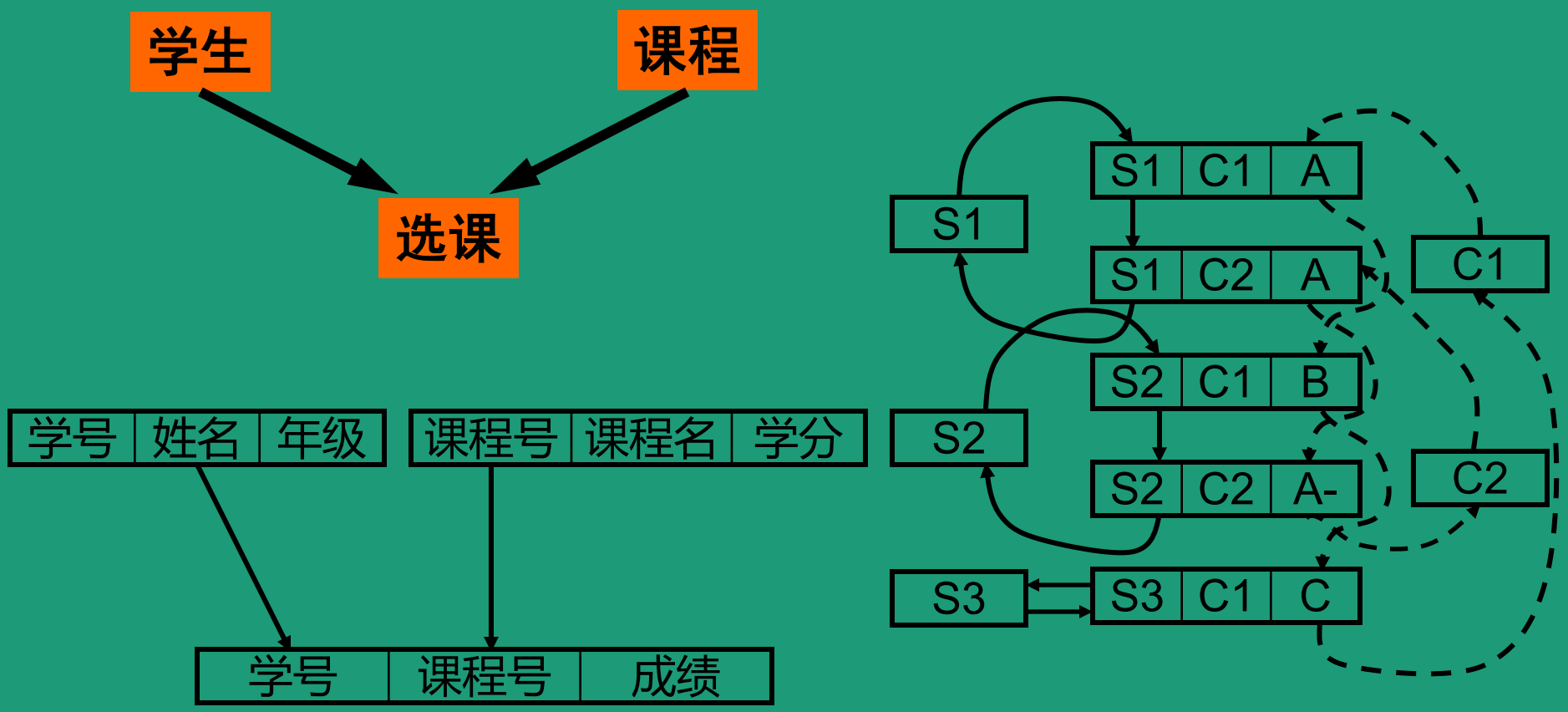
层次数据结构：数据对象


D02 计算机 R1101
R01 数据库
R02 网络
R03 人工智能
S0012 王明 G1
S0020 郑直 G2
S0199 周密 G3
E1101 何璧 教授
E3721 刘新 讲师
E1234 王恩 教授
E3721 付弈 助教

层次模型特征

- 优点：结构简单，易于实现
- 缺点：支持的联系种类太少
  - 只支持二元一对多联系
  - 只允许实体集间的一种联系，不支持实体集间的多种联系
- 缺点：数据操纵不方便
  - 子结点的存取只能通过父结点来进行
  - 插入、删除复杂，父结点的删除导致子结点的删除，丢失必要的信息
- 代表产品：
  - IBM的IMS数据库，1969年研制成功

逻辑模型示例-网状模型


学生
课程
选课
学号
姓名
年级
课程号
课程名
学分
学号
课程号
成绩
S1
S2
S3
S1 C1 A
S1 C2 A
S2 C1 B
S2 C2 A-
S3 C1 C
C1
C2

网状模型特征

- 优点
  - 表达的联系种类丰富
  - 性能良好，存取效率高
- 缺点
  - 结构复杂
  - 语言复杂
- DBTG报告
  - 1969年，由美国CODASYC（Conference On Data System Language，数据系统语言协商会）下属的DBTG（Data Base Task Group）组提出，确立了网状数据库系统的概念、方法、技术

逻辑模型示例-关系模型

- 用二维表来表示实体及其相互联系

属性

元组

学号	姓名	年龄	性别	系号
S01	张军	21	男	D01
S02	李红	22	女	D01
S03	王伟	19	男	D02

关系模型特征

- 简单，表的概念直观、单一，用户易理解
- 非过程化的数据请求，数据请求可以不指明路径
- 数据独立性，用户只需提出“做什么”，无须说明“怎么做”
- 坚实的理论基础：关系代数

空间信息模型

- 空间信息
- 几何信息、拓扑信息和属性信息
- 矢量表示与栅格表示
- 空间关系
- 空间操作
- 空间查询

空间信息

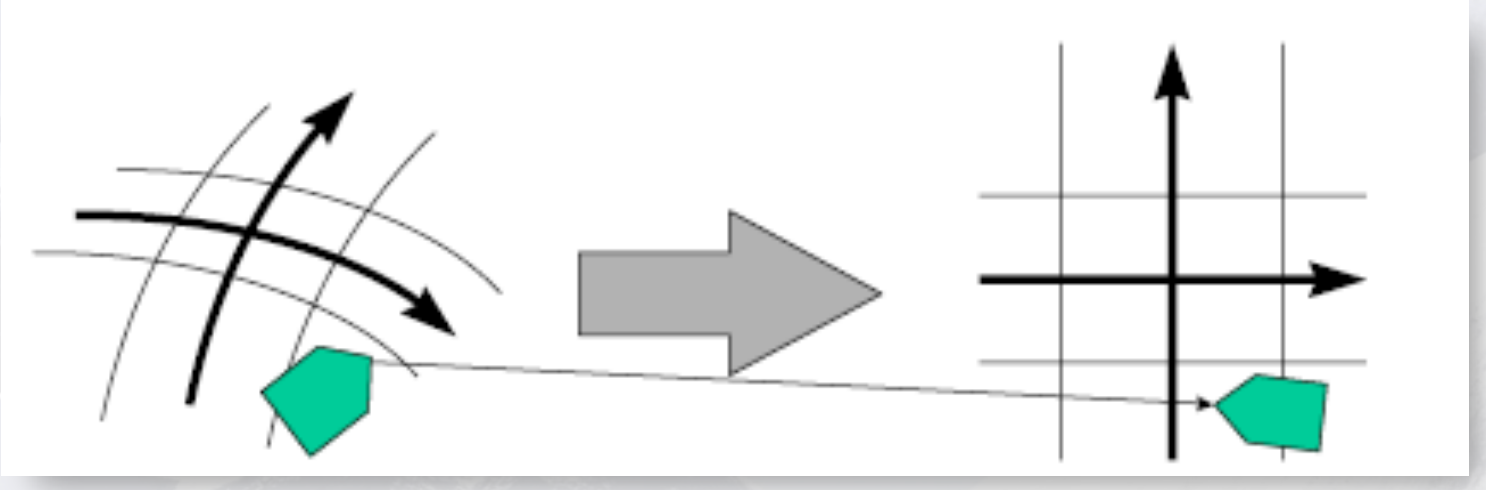
- 与地球表面有关的事物、事件
  - 行政区划
  - 交通运输网
  - 动植物分布
  - 校园建筑布局
  - 天气变化趋势
  - 污染物扩散和影响

空间信息

- 描述这些事物、事件的信息
  - 包含了空间位置和延伸信息
    - 什么地方？
    - 多大范围？
  - 因为空间位置和延伸发生关系
    - 方位？
    - 相互交叉？重叠？包含？
  - 同时包含了其它位置无关的信息
    - 名称？归属？价值？

几何、拓扑、属性

- 几何信息
  - 描述了事物在空间中的位置及所占据的范围
  - 将地球表面以投影方式转换为平面
  - 通过平面几何来抽象描述和研究事物的位置和范围
  - 用图形和符号的方式来描绘这些空间相关的事物



几何、拓扑、属性

- 拓扑信息
  - 研究空间相关的事物本身或者事物之间的在空间坐标变换下的不变性质
  - 事物本身的内外关系
  - 事物之间的相离、相接、相交
  - 事物之间相连的布局

几何、拓扑、属性

- 属性信息
  - 与位置范围无关的其它信息
  - 描述了事物本身的内在性质和外在表现
  - 事物之间的非位置关系

供应商号	供应商名	所在城市
B01	红星	北京
S10	宇宙	上海
T20	黎明	天津
Z01	立新	重庆

矢量表示与栅格表示

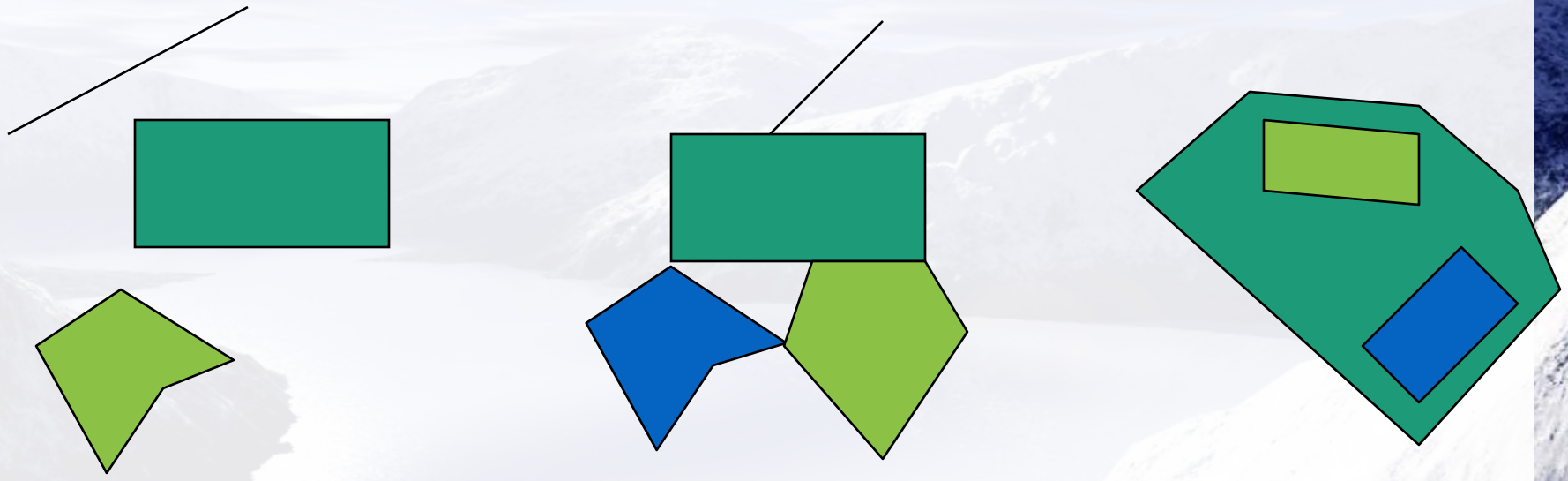
- 矢量表示
  - 采用点坐标、点之间的连接来描述地物的位置、边界

矢量表示与栅格表示

- 栅格表示
  - 采用图像方法直接描述地物

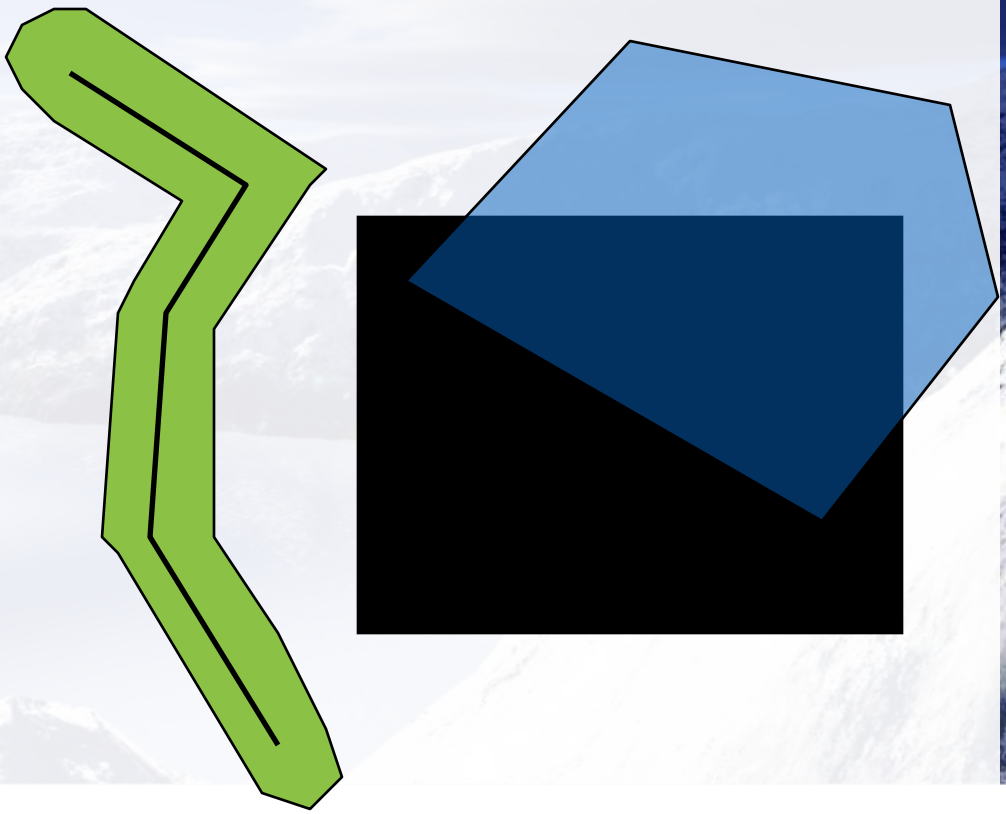
空间关系

- 描述地物之间在空间位置和范围上的相离、相接、相交、包含等关系



空间操作

- 代表地物的几何图形之间所进行的一些几何运算
  - 缓冲区生成
  - 叠加运算
  - ……



空间查询

- 获取关于地物的几何性质
  - 位置、周长、面积
- 获取地物之间的空间关系
  - 相离、相接、相交或包含
  - 可以通过叠加运算来获取空间关系
- 获取代表地物的图形的图论性质
  - 连通性分析
  - 最短/最佳路径选择
  - 最佳资源配置

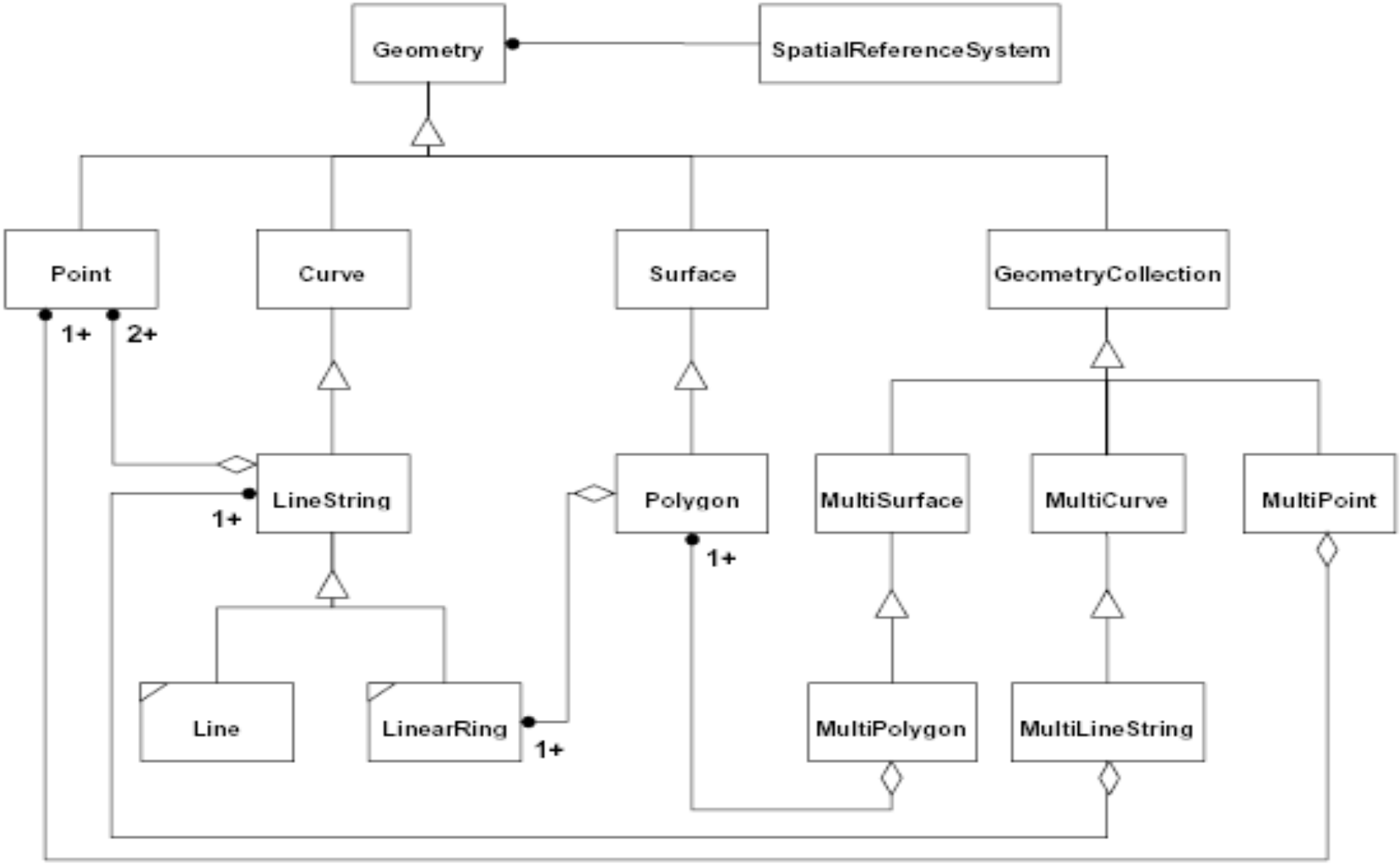
●基于场的模型

- 将空间信息看作一个平面上的函数
  - 定义域是地理空间
  - 值域是信息编码的集合
- 场模型的组成部分
  - 空间框架：构建了场的定义域
    - 连续平面：坐标定位
    - 离散网格：编码定位
  - 场函数：f:空间框架→属性域
  - 场操作：对函数的运算
    - 算术运算和函数合成
    - 微分（计算坡度）与积分（计算均值）相关

基于对象的模型

- 将空间信息看作具有属性和关系的对象的集合
  - 可识别与可区分的对象实体
  - 对象实体具有属性（空间属性/非空间属性）
  - 对象实体之间具有关联
- 对象模型的组成部分
  - 对象类型
  - 对象属性和操作
  - 对象关系

OGIS空间数据类型


Geometry
SpatialReferenceSystem
Point
Curve
Surface
GeometryCollection
1+
2+
LineString
1+
Polygon
1+
MultiSurface
MultiCurve
MultiPoint
Line
LinearRing
1+
MultiPolygon
MultiLineString

空间对象操作与关系

- 对象的集合操作和集合关系
  - 对象作为元素或者集合来处理
- 拓扑关系
  - 空间对象之间与形状、大小等度量无关的关系：相接/相离/包含/相交
  - 基于点集拓扑的表达和代数拓扑的表达
- 度量关系/性质
  - 和度量空间中距离相关的关系和性质：距离/长度/面积
- 方位关系
  - 东南西北前后左右等绝对或者相对的方位

度量空间

- 设X是一个非空集合，如果已知X中任何一对元素x,y，均给定一个实数d(x,y)与之对应，而且满足下列条件
  - d(x,y)>=0（非负）；
  - d(x,y)=0⇔x=y（到自身距离为0）；
  - d(x,y)=d(y,x)（对称）；
  - d(x,y)<=d(x,z)+d(z,y)（三角不等式）
- 则称d(x,y)是x,y间的距离，又称集合X按距离d成为赋度量空间或赋距离空间
- 简称为度量空间或距离空间，并记做(X,d)

度量空间

- 从距离定义出发，可以定义一系列空间关系的概念
  - 开集、闭集
  - 邻域
  - 边界
  - 内部、外部

拓扑空间

- 拓扑空间对度量空间进行进一步的一般化
  - 除去了度量空间中的“距离”定义，使得可以在任意集合上定义开集、邻域、内部等概念
  - 度量空间是拓扑空间最重要的一类
- 拓扑学研究与距离度量无关的空间性质，以及在同胚变换下的一些不变性质
  - 同胚是指一个图形可以一对一地双方连续地变换为另一个图形

拓扑关系代数：DE-9IM

- 维度扩展的9交叉模型DE9IM
  - Dimensionally Extended 9-Intersection Model
  - 4交叉模型只两两测试几何形的内部和边界的相交情况，不能处理复杂的几何形
  - 9交叉模型增加了对外部相交的测试，同时将测试结果扩展为相交点集的维度
  - 可以表达点线面之间的空间关系，可以处理带洞的多边形以及线和多边形集合

拓扑关系代数：DE-9IM

- 记号
  - 给定几何形a，I(a), B(a), E(a)分别代表a的内部、边界和外部
  - 将两个几何形的内部、边界、外部分别两两做相交操作，操作的结果记为x，x可能是多个不同维度几何形的集合
  - 将x中维度最高几何形的维度记为dim(x)，取值可能为0、1、2；如果x为空集，则dim(x)=-1

拓扑关系代数：DE-9IM

- 定义

	Interior	Boundary	Exterior
Interior	$dim(I(a) \cap I(b))$	$dim(I(a) \cap B(b))$	$dim(I(a) \cap E(b))$
Boundary	$dim(B(a) \cap I(b))$	$dim(B(a) \cap B(b))$	$dim(B(a) \cap E(b))$
Exterior	$dim(E(a) \cap I(b))$	$dim(E(a) \cap B(b))$	$dim(E(a) \cap E(b))$

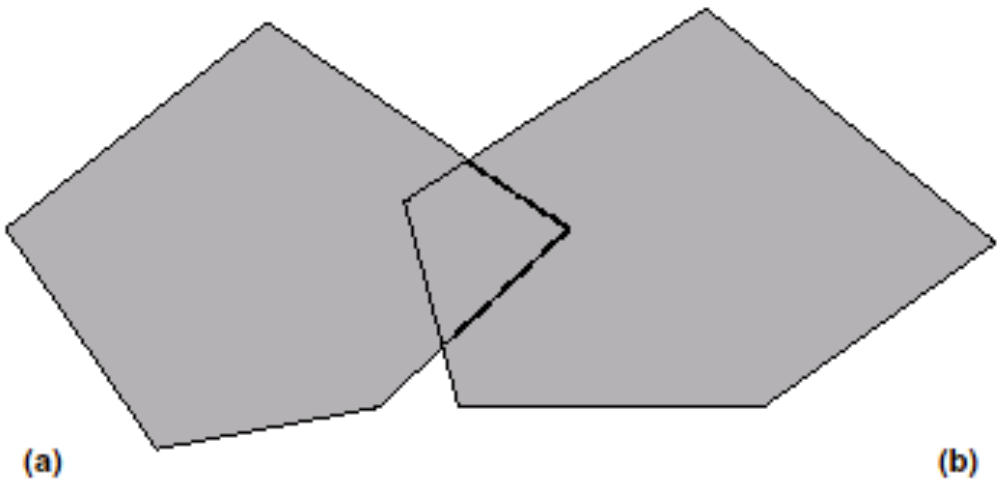
Table 2.1—The DE-9IM

b

a

拓扑关系代数：DE-9IM

- 测试示例



	Interior	Boundary	Exterior
Interior	2	1	2
Boundary	1	0	1
Exterior	2	1	2

拓扑关系代数：DE-9IM

- 模式矩阵pattern matrix
  - 两个几何形之间的空间关系判定可以通过其DE-9IM测试矩阵和模式矩阵的匹配来实现
  - 模式矩阵是3*3的矩阵，与DE-9IM矩阵一一对应，但是分量的取值范围是{T,F,*,0,1,2}
    - T匹配0,1,2，代表交集不为空
    - F匹配-1，代表交集为空
    - *匹配-1,0,1,2，代表不关心
    - 0，1，2分别匹配自身
  - 模式矩阵可以行优先的方式表示为一维数组

拓扑关系代数：DE-9IM

- 模式矩阵示例

```
char * overlapMatrix =  'T*T***T**' ;
Geometry* a, b;
Boolean b = a->Relate(b, overlapMatrix);
```

拓扑关系代数：DE-9IM

- 优势
  - 可以表达多种空间关系
  - 可以对特定的空间关系进行微调
- 缺点
  - 所表示的空间关系不能通过自然语言表达
  - 对于一般开发者形成理解障碍
- 解决方案
  - 将普遍理解的空间关系用DE-9IM重新定义

拓扑关系代数：DE-9IM

- 相离Disjoint

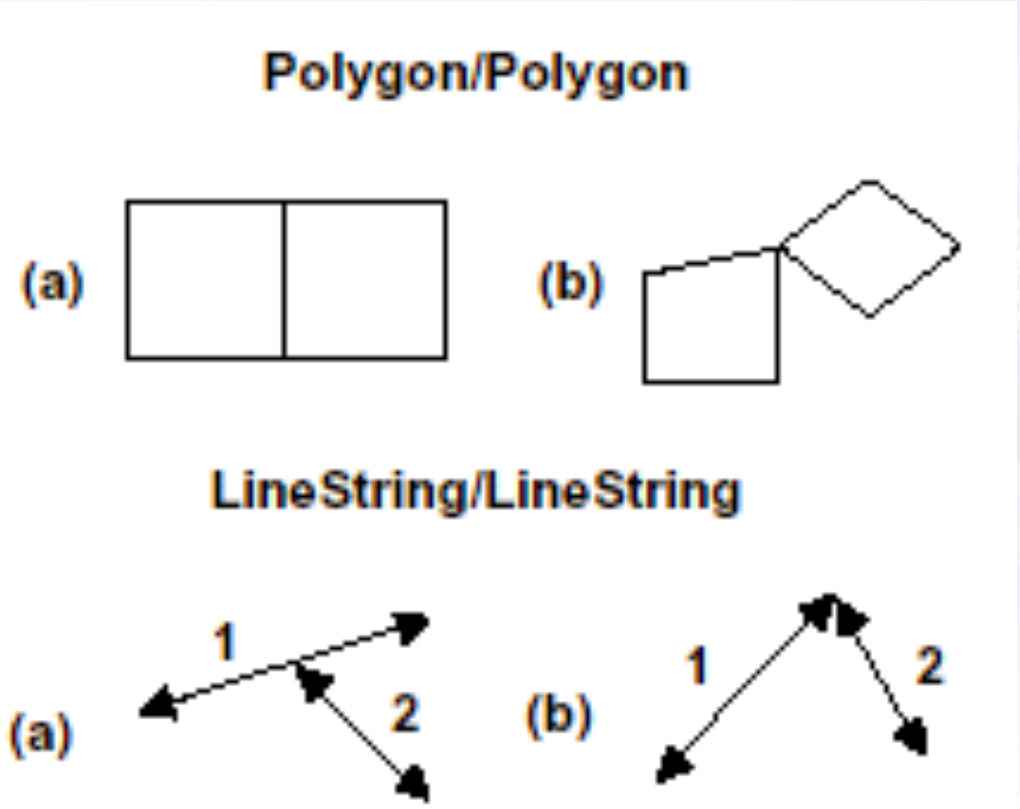
$$a.Disjoint(b) \Leftrightarrow a \cap b = \varnothing$$

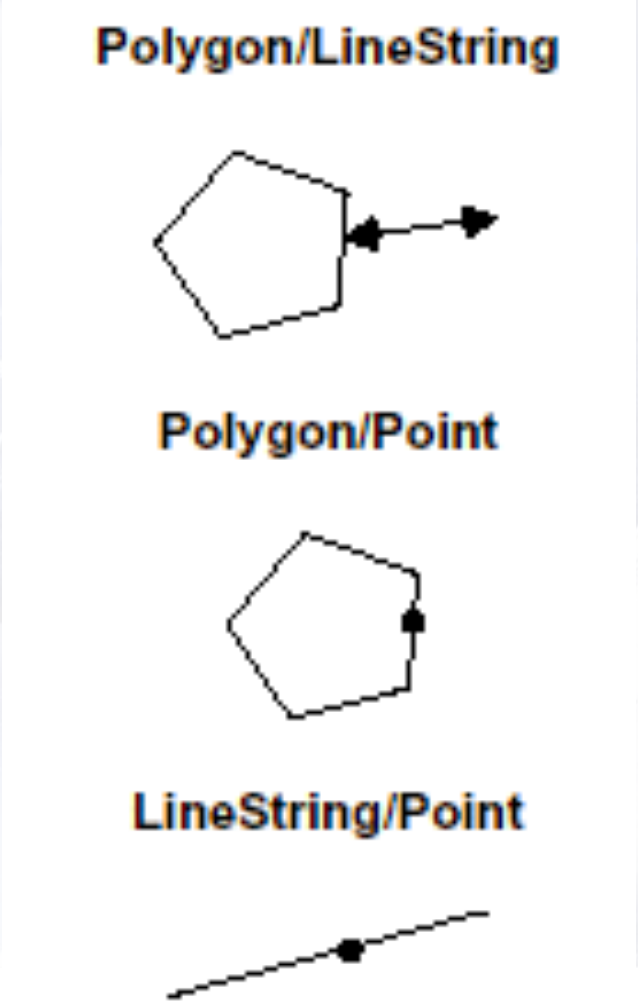
$$a.Disjoint(b) \Leftrightarrow (I(a) \cap I(b) = \varnothing) \wedge (I(a) \cap B(b) = \varnothing) \wedge (B(a) \cap I(b) = \varnothing) \wedge (B(a) \cap B(b) = \varnothing)$$
$$\Leftrightarrow a.Relate(b, \text{'FF*FF****'})$$

	内部	边界	外部
内部	F	F	*
边界	F	F	*
外部	*	*	*

拓扑关系代数：DE-9IM

- 相接Touches图示





拓扑关系代数：DE-9IM

- 相接Touches
  - 适于A/A, L/L, L/A, P/A 和 P/L

$$a.Touches(b) \Leftrightarrow (I(a) \cap I(b) = \varnothing) \wedge (a \cap b) \neq \varnothing$$

$$a.Touches(b) \Leftrightarrow (I(a) \cap I(b) = \varnothing) \wedge ((B(a) \cap I(b) \neq \varnothing) \vee (I(a) \cap B(b) \neq \varnothing) \vee (B(a) \cap B(b) \neq \varnothing))$$
$$\Leftrightarrow a.Relate(b, \text{'FT*******'}) \vee a.Relate(b, \text{'F**T*****'}) \vee a.Relate(b, \text{'F***T****'})$$

	内部	边界	外部
内部	F	T/*/*	*
边界	*/T/*	*/*/T	*
外部	*	*	*

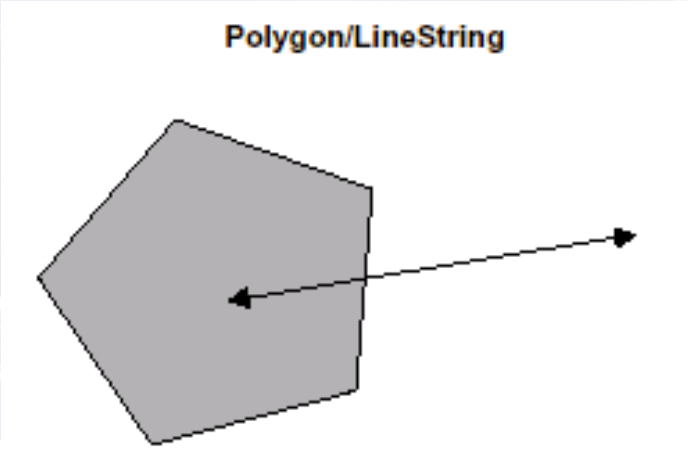
拓扑关系代数：DE-9IM

- 交叉Crosses
  - 适于P/L, P/A, L/L 和 L/A

$$a.Crosses(b) \Leftrightarrow (dim(I(a) \cap I(b)) < max(dim(I(a)), dim(I(b)))) \wedge (a \cap b \neq a) \wedge (a \cap b \neq b)$$

P/L, P/A, L/A

$$a.Crosses(b) \Leftrightarrow (I(a) \cap I(b) \neq \emptyset) \wedge (I(a) \cap E(b) \neq \emptyset) \Leftrightarrow a.Relate(b, \text{'T*T******'})$$



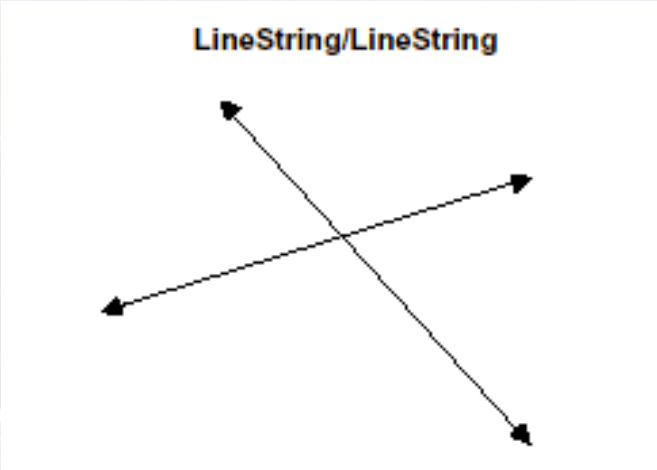
	内部	边界	外部
内部	T	*	T
边界	*	*	*
外部	*	*	*

拓扑关系代数：DE-9IM

- 交叉Crosses

L/L

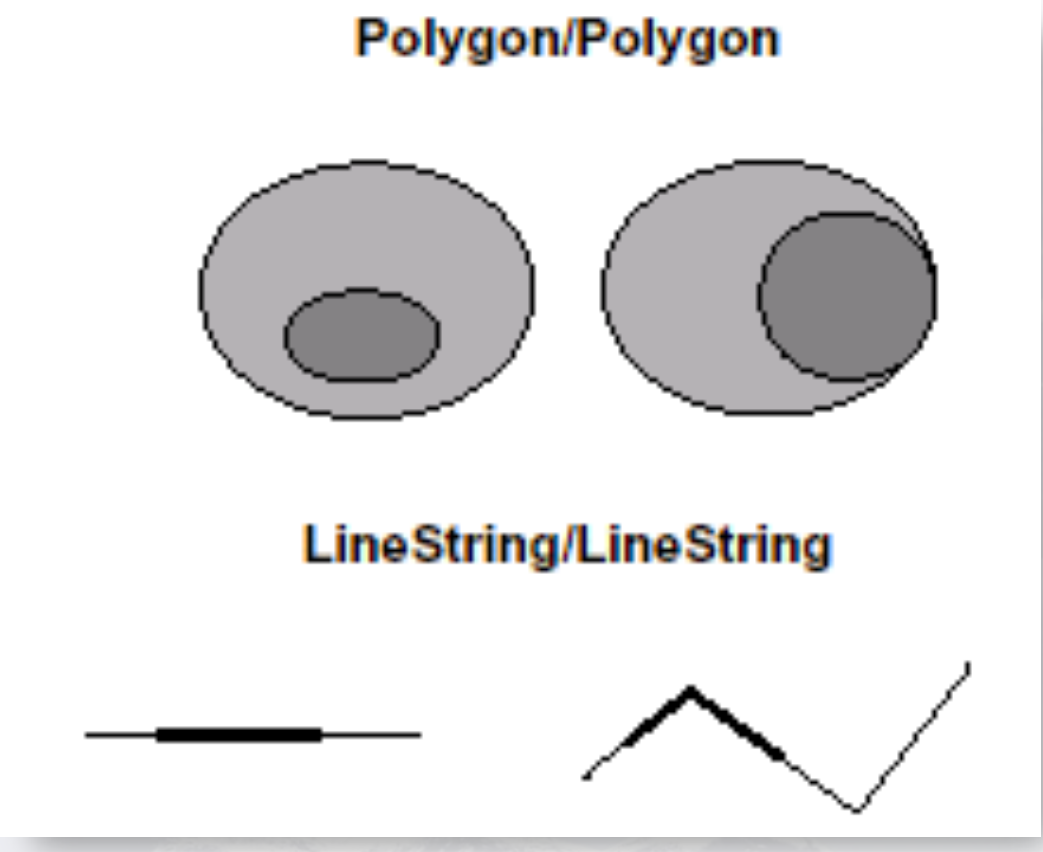
$$a.Crosses(b) \Leftrightarrow dim(I(a) \cap I(b)) = 0 \Leftrightarrow a.Relate(b, \text{'0********'});$$

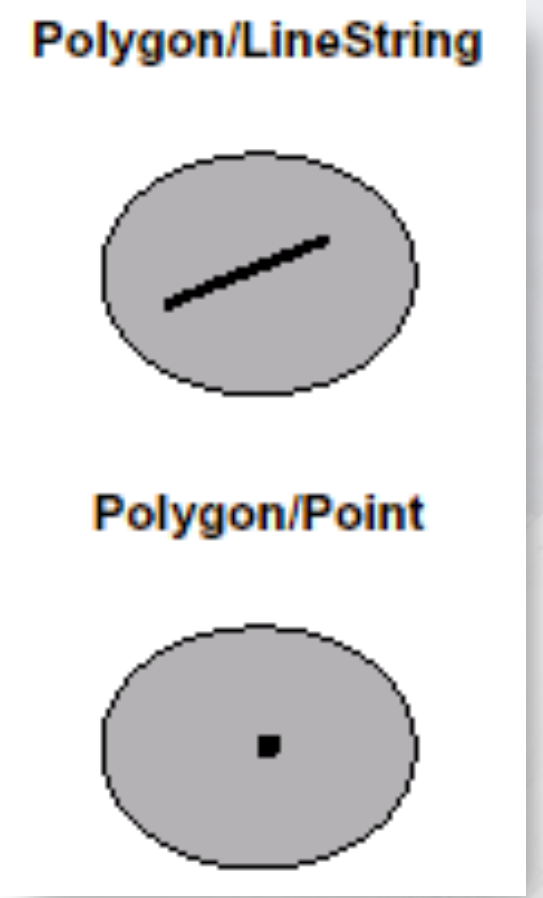


	内部	边界	外部
内部	0	*	*
边界	*	*	*
外部	*	*	*

拓扑关系代数：DE-9IM

- 被包含Within图示





拓扑关系代数：DE-9IM

- 被包含Within

$$a.Within(b) \Leftrightarrow (a \cap b = a) \wedge (I(a) \cap I(b) \neq \varnothing)$$

$$a.Within(b) \Leftrightarrow (I(a) \cap I(b) \neq \varnothing) \wedge (I(a) \cap E(b) = \varnothing) \wedge (B(a) \cap E(b) = \varnothing)\ ) \Leftrightarrow a.Relate(b,\ 'T*F**F***')$$

	内部	边界	外部
内部	T	*	F
边界	*	*	F
外部	*	*	*

拓扑关系代数：DE-9IM

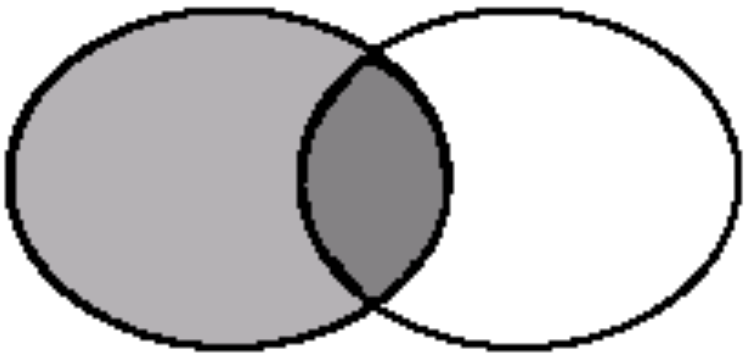
- 重叠Overlaps
  - 适于A/A, L/L, P/P

$$a.Overlaps(b) \Leftrightarrow (dim(I(a)) = dim(I(b)) = dim(I(a) \cap I(b))) \wedge (a \cap b \neq a) \wedge (a \cap b \neq b)$$

A/A, P/P

$$a.Overlaps(b) \Leftrightarrow (I(a) \cap I(b) \neq \varnothing) \wedge (I(a) \cap E(b) \neq \varnothing) \wedge (E(a) \cap I(b) \neq \varnothing) \Leftrightarrow a.Relate(b, \text{'T*T***T**'})$$

Polygon/Polygon



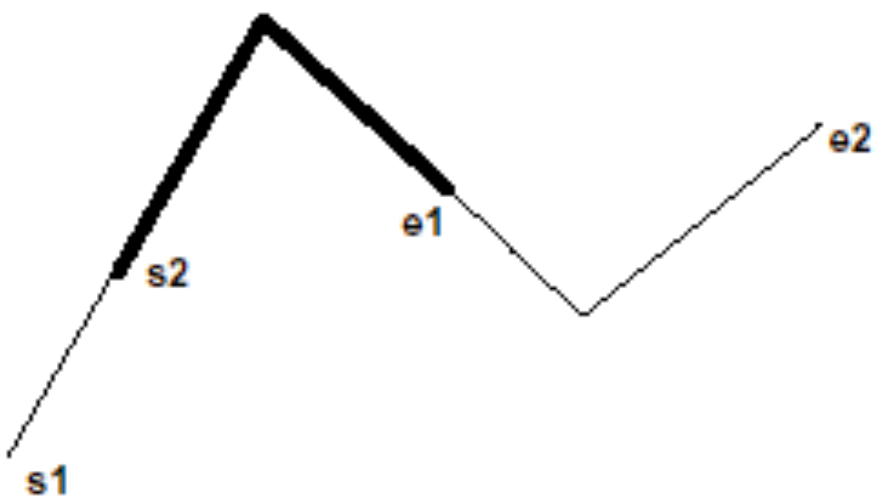
	内部	边界	外部
内部	T	*	T
边界	*	*	*
外部	T	*	*

拓扑关系代数：DE-9IM

- 重叠Overlaps

L/L

$a.Overlaps(b) \Leftrightarrow (dim(I(a) \cap I(b) = 1) \wedge (I(a) \cap E(b) \neq \varnothing) \wedge (E(a) \cap I(b) \neq \varnothing) \Leftrightarrow a.Relate(b, \text{'1*T***T**'})$



	内部	边界	外部
内部	1	*	T
边界	*	*	*
外部	T	*	*

拓扑关系代数：DE-9IM

- 包含Contains

$$a.Contains(b) \Leftrightarrow b.Within(a)$$

- 相交Intersects

$$a.Intersects(b) \Leftrightarrow !\,a.Disjoint(b)$$

拓扑关系代数：RCC

- 区域连接算子RCC：Region Connection Calculus
- RCC是一个代数系统，从定义的“区域”、“连接关系”以及两条公理出发，采用一阶谓词逻辑定义出所有的拓扑关系
- “区域”是拓扑空间中的非空集合，也就是我们需要讨论拓扑关系的对象
- “连接关系”是一个区域间的二元关系，两条公理阐明了其自反和对称的性质

拓扑关系代数：RCC

$$
\begin{aligned}
\mathsf{DC}(x,y) &\equiv_{def} \neg\mathsf{C}(x,y) \\
\mathsf{P}(x,y) &\equiv_{def} \forall z[\mathsf{C}(z,x) \rightarrow \mathsf{C}(z,y)] \\
\mathsf{PP}(x,y) &\equiv_{def} \mathsf{P}(x,y) \wedge \neg\mathsf{P}(y,x) \\
\mathsf{EQ}(x,y) &\equiv_{def} \mathsf{P}(x,y) \wedge \mathsf{P}(y,x) \\
\mathsf{O}(x,y) &\equiv_{def} \exists z[\mathsf{P}(z,x) \wedge \mathsf{P}(z,y)] \\
\mathsf{PO}(x,y) &\equiv_{def} \mathsf{O}(x,y) \wedge \neg\mathsf{P}(x,y) \wedge \neg\mathsf{P}(y,x) \\
\mathsf{DR}(x,y) &\equiv_{def} \neg\mathsf{O}(x,y) \\
\mathsf{EC}(x,y) &\equiv_{def} \mathsf{C}(x,y) \wedge \neg\mathsf{O}(x,y) \\
\mathsf{TPP}(x,y) &\equiv_{def} \mathsf{PP}(x,y) \wedge \exists z[\mathsf{EC}(z,x) \wedge \mathsf{EC}(z,y)] \\
\mathsf{NTPP}(x,y) &\equiv_{def} \mathsf{PP}(x,y) \wedge \neg\exists z[\mathsf{EC}(z,x) \wedge \mathsf{EC}(z,y)] \\
\mathsf{P}^{-1}(x,y) &\equiv_{def} \mathsf{P}(y,x) \\
\mathsf{PP}^{-1}(x,y) &\equiv_{def} \mathsf{PP}(y,x) \\
\mathsf{TPP}^{-1}(x,y) &\equiv_{def} \mathsf{TPP}(y,x) \\
\mathsf{NTPP}^{-1}(x,y) &\equiv_{def} \mathsf{NTPP}(y,x)
\end{aligned}
$$

C: connects
DC: disconnects
P: is part of
PP: is proper part of
EQ: is equal to
O: overlaps
PO: partially overlaps
DR: discrete from
EC: externally connected
TPP: tangential proper part
NTPP: non-tangential …

拓扑关系代数：RCC

- 通过上述定义的关系，可以进一步定义作用于区域的函数
  - sum(x,y): union of x and y
  - compl(x): complement of x
  - prod(x,y): intersection of x and y
  - diff(x,y): difference of x and y
- 这些函数又可以进一步定义更为复杂的关系

$$\mathsf{CON}(x) \quad \equiv_{def} \quad \forall y \forall z [\mathsf{EQ}(\mathsf{sum}(y,z),x) \rightarrow \mathsf{C}(y,z)]$$

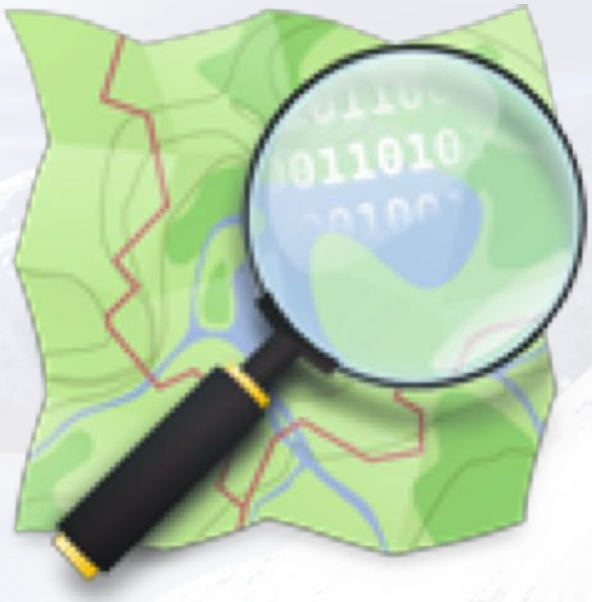
拓扑关系代数：RCC

- 还可以通过增加公理来附加区域的一些性质
  - 如表达区域都可以分出真内部区域
  - 可以表达更为复杂或者特定的拓扑关系

$$\forall x \exists y \mathsf{NTPP}(y, x)$$

课堂讨论

- 理解Open Street Map的开放数据格式
- https://planet.openstreetmap.org/
- 下载完整数据
- 并用Python库解析
  - https://github.com/omniscale/imposm-parser
  - 或其它OSM数据解析库
- 统计各种地图数据
  - 图层
  - 要素数量
  - 等



参考文献

- [TP311.13/261]空间数据库 = Spatial databases a tour (美) Shashi Shekhar, Sanjay Chawla著 谢昆青 ... 等译 北京 机械工业出版社 2004
- OGC Open Geodata Model
- J. Renz: Qualitative Spatial Reasoning with Topological Information(The Region Connection Calculus), LNAI 2293, pp. 41–50, 2002.