

大型 GIS 地理数据库的无缝组织*

李爱勤 龚健雅 李德仁

(武汉测绘科技大学 GIS 研究中心, 武汉市珞喻路 129 号, 430079)

摘要 提供了大型地理数据库无缝组织的实例。模型采用了纵向与横向地理数据库设计策略, 并提出了一种高效的改良 Morton 码空间索引和无缝数据组织方法。

关键词 GIS; 无缝地理数据库; 分级扩散二叉树编码

分类号 P208 TP392

地理变化无限复杂, 空间数据库中的信息编码也只能是真实世界的近似表达。“某条海岸线有多长?” 这样的问题必须依赖于具体的测量设备的分辨率或者是地图的比例尺才能有较为明确的答案^[1]。一般来说, 地理数据空间分辨率越高, 表达越详细。地理面块和曲线缺乏较好的定义, 一般采取综合简化的方法, 而并不采用连续函数来描述地理实体。

在资源管理一类典型的地理数据库中, 用户常会发现数据库不能完全满足某些方面的要求, 如空间细节、通用性、地理覆盖, 甚至于表达范围。实质上, 用户的需求总是可能会超过数据库的内容和对地理实体的知识表达, 信息总是不能完全满足空间分辨率、专题或地理层的无限细分扩展。基于这种假设, 定义“大型”为具有合适的效率, 超过当前系统规模的概念。

1 传统方法

地理信息采用分幅地图表达的形式已经很久了, 基于分幅纸质地图的大量地理数据的保存、更新、使用显然具有很大的局限性。因而, 基于这种分幅地图思想的传统 GIS 的地理数据库的组织也同样具有许多弱点, 主要表现为:

1) 查询往往涉及到多幅图或在不同的专题间进行。例如, 某条路穿越多幅图, 系统就必须在多幅相邻图和专题间靠地物标识进行空间和逻辑上的匹配、关联等操作。

2) 地理实体的完整性和一致性难以维护。同一实体分布或穿越多幅图, 在系统中分块或分段存储具有多个编码, 存储分散必然导致地物整体

一致性维护的复杂性和困难。

3) 不便于数据库的分布式管理。同一地理实体, 若存储分布在不同的网络节点上, 则分布式检索或存取效率将很难优化提高。

4) 分幅管理对于数据共享和地理实体一级的安全性、一致性管理增加了难度。

5) 专题信息的分离管理不利于各种地理数据的有效获取。如测绘部门主要生产地形图; 环保部门收集环保专题数据; 土地部门偏重土地利用等。虽专题不同, 但存在不少交叉。各行业专家观点不尽相同, 这样不利于同一地区各种地理数据的有效组织和采集。

为解决以上各种问题, 许多学者和 GIS 软件公司进行了大量的研究。例如, 基于 Morton 码的瓦片式大型地理数据库设计思想^[1], Intergraph 的 MGE, ESRI 的 ARC/INFO 等都能建立“无缝”GIS 地理数据库。然而, 它们仍然只是一种逻辑概念上的“无缝”组织, 能够完成地理数据的几何接边和逻辑接边, 但物理上仍然按照图幅的概念进行存储管理, 对同一地理实体在多个几何标识和同一地物标识间进行后台关联处理, 对用户来说是不可见的, 因而说是逻辑上的“无缝”组织。

2 一种无缝地理数据库模型

笔者为大型地理数据库提出了一种逻辑和物理概念上真正的地理数据无缝组织模型。该模型基于 Client/Server 结构, 保持地物的连续、完整和一致性表达, 地理数据分布存储, 突破了传统图幅分块的思想, 对地理数据库进行纵向和横向设计, 并设计出一种改良 Morton 码的“分级 4 象限

扩散四叉树编码”方案进行地理数据的无缝管理和空间索引。

2.1 纵向设计

由于不同比例尺数据源的精度等因素的限制,在设计无缝数据库时,首先建议对于不同比例尺的数据源建立不同层次的地理数据库,在不同库之间依据地理位置和属性特征等建立关联关系。各种比例尺数据库中还可根据实际用途和情况分别建立多个同级子库,如图 1所示(以 1:1万、1:25万、1:100万为例)。

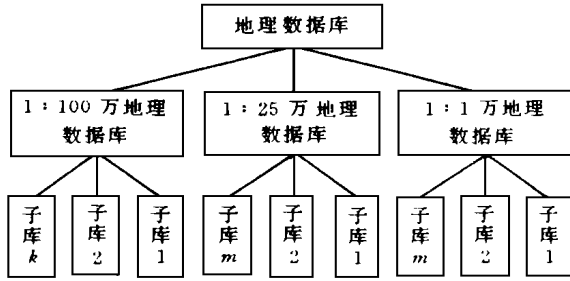


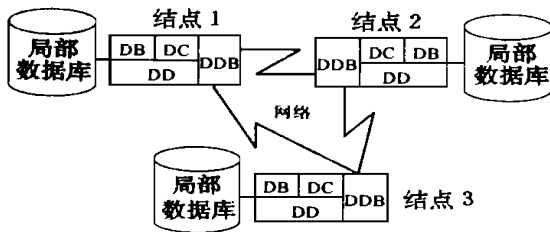
图 1 地理数据库的纵向结构

Fig. 1 Vertical Design for Geographical Databases

纵向设计一般只有在建立大型地理数据库或多比例尺地理数据库时考虑使用。

2.2 横向设计

横向设计为假定在建立同一精度数据源的大型地理数据库时的网络分布式无缝设计。由于GIS要管理地理空间中庞大、复杂的海量数据,必须充分运用先进的计算机网络和数据库技术,我们选择基于 Client/Server 网络模型的分布式数据库结构以适应数据库平稳地动态扩充和维护,如图 2所示



DB数据库; DC数据传输系统;
DD 数据字典; DDB分布式 DB

图 2 分布式数据库管理系统结构示意图

Fig. 2 Sketch of Distributed Database Management System

2.2.1 无缝地理数据库中“工程+工作区”约定

地理数据库建库的对象来源于真实世界的无缝连接的地球表面。本文所讲的“工程”是指一个完整的无缝地理数据库,因而工程中的地物要保持存储、表达的完整性和一致性。例如,一条铁路,

在工程中具有唯一的几何标识和地物标识。可以说工程是抽象表达客观世界中完整地物的数据仓库或数据服务器;而“工作区”的概念是为了方便使用地理数据库,可以说是在应用时的工程的临时子集。可以按规则大小划分,也可以相互嵌套,最大甚至可以为整个工程。工作区可以包含任何区域。任意一层或多层地物,如图 3所示

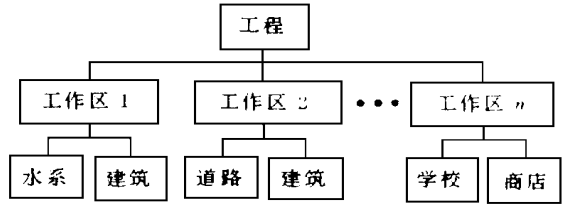


图 3 “工程+工作区”结构

Fig. 3 “Project+ Workpartition”

工作区是工程(服务器)的临时子集,作为客户端直接面对用户。而工程的数据来源于工作区,并作为服务器为客户提供服务。

2.2.2 无缝地理数据库的数据结构与组织

地理数据库中数据组织包括对空间数据、属性数据、图像数据、DEM数据、多媒体数据等的组织和管理。无缝地理数据库中数据组织的关键在于能够在空间和语义(属性)上做到完整地表达和存储

1) 空间地物划分为如下 5 大类:

a. 点对象: 包括单点(如烟筒、控制点)、有向点(如桥、涵洞)、点群(如路灯、散树)等;

b. 线对象: 相互连接的点串或弧段串,可以是网状的,但必须互相连接,如河流、河网、道路、路网等;

c. 面对象: 由一个或多个封闭不相交的多边形或环组成,可包含多个岛,如地块、房屋等;

d. 复杂对象: 由任意两个或两个以上的基本对象(点、线、面)或另外的复杂地物组成,如学校可作为一复杂地物,它包含有道路、房屋、水塔等基本地物要素;

e. 注记: 用以描述地名、地物的属性等文字信息

由于该模型要求所有地物保持完整性存放,因而每一完整地物无论大小都只能拥有唯一的几何对象标识符(OID)。在工程中,地物的空间数据必须完整存储,而地物的完整性划分完全依赖于用户或应用的要求。例如,对于长江,可以将它作为面对象(较大比例或中等比例尺建库时)、线对象(小比例尺,长江主流或主要长江水系),或

者将整个长江水系构建成为一个复杂地物来管理和描述整个与长江相关的空间和属性信息

在每一地物的空间数据中记录该地物的地物分类码 (或称为用户标记 (UserID))。

2) 属性数据

所有空间地物的属性数据在工程中统一管理,用关系型数据库管理系统 (RDBMS) 进行管理,属性数据以表的形式进行存储。用户可参照标准对工程中的地物进行统一分类 (如分成各等级的控制点、道路、居民地、水系、植被等),并指定唯一的地物分类码 (或称地物标识 (UserID))。任一完整地物只能对应一条属性数据记录,空间地物与相应的属性记录间靠唯一的对象标识 OID 进行联结,如图 4 所示。

地物空间数据 (以线对象为例)

OID	UserID	弧段 2 维/3 维 数 /DEM...	格式 化 坐 标 串	其 它 3 维 信 息 (必要时)
...
102018	W5019	14	0	...
330456	R2002	58	1	...
...

一级公路属性表 (UserID=R2002)
(用户定义属性结构)

OID	Name	宽度	路基	路面	路况	车道数	其它
...
330456	珞喻路	20	硬	砼	良好	6	...
...

河流属性表 (UserID=W5019)
(用户定义属性结构)

OID	Name	流向	水质	平均水深	源头	其它
...
102018	长江	东	3.5
...

图 4 地物空间数据与属性数据的关系

Fig. 4 Relationship Between Spatial Object and its Attributes

3) 图像数据和 DEM 数据

对于图像和 DEM 数据,由于其相似性,可以单独建库。按分辨率可提取出影像或 DEM 金字塔,各层按空间划分成格网,每块进行压缩存储,各层格网进行与空间相关的分层编码 (见本文 2.2.4 节)。

DEM 数据还可以依靠在地物的空间数据记录中的“2 维 /3 维 /DEM”附加信息类型项以自定义格式的变长字段存储于地物空间数据记录中。

4) 多媒体数据

各种多媒体数据可存储在相应地物的属性记录中并以 BLOB 形式压缩存储,与地物关联

2.2.3 无缝地理数据库的空间索引与空间表达

为了在工程中有效地组织和表达空间地理实体,我们按照地物的大小 (这里指地物覆盖的最大范围) 对其进行分级抽取,对不同大小地物的几何对象标识 OID 进行整理、分层,建立空间索引。

参考建库比例尺、内容、性质等因素将地面沿坐标轴方向分割成若干层正交格网,层与层间格网大小不同,相差倍数应仔细选择 (此处格网的划分并非对空间的实际分割,只是为空间索引和数据组织服务) 与地图分幅不同,无缝地理数据库中不存在分幅的概念,地物入库时,必须进行几何和逻辑 (指属性语义数据) 接边。同一地物在工程中合成为一完整实体 (对象),该对象记录地物的完整空间信息和属性数据,由系统分配给它一个唯一的 OID,并按其实体大小将该 OID 存入相应的某层格网中,如图 5 所示。

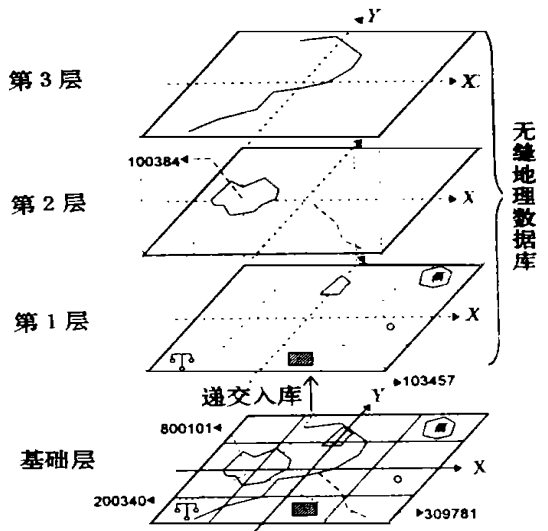


图 5 地理数据库无缝空间表达

Fig. 5 Spatial Represent for Seamless Geographical Database

现以一地块 (带有 OID 标签) 为例 (图 5),该地块在基础层上由于地图分幅的影响,被分割在 4 幅相邻的图幅中。假设这里按图幅范围建立了各个工作区,则该地块在相应工作区中的 OID 分别为 800101, 200340, 103457, 309781。OID 由系统自动产生,此时 OID 只能保证在各工作区内部保持唯一,但工作区间 OID 可能会重复出现。可见,在基础层上,该地块被分成不同的 4 块,每块在各自所属的工作区中都不完整,甚至各自关联着自己的属性记录。当将这 4 个工作区递

交到工程中时,必须进行空间接边,形成图中OID=100384的完整无缝的地物。由于该地物的大小(范围)已超出第一层格网的大小,又小于第3层,其OID则被记录在第2层,由第2层提供其OID值100384,在整个无缝数据库具有唯一性。同时,还要进行属性数据的合并和统一,即所谓的逻辑接边。

照此规则,任一地物OID只能存入唯一的某层中,而且它在该层最多只能跨越或覆盖到4个格网大小,否则就得被提到上层或更高层中去(这里分层只是基于同一基础层——全要素地理层,按地物的分布和大小将它们组织在不同层上,便于管理和表达,并没有进行任何情况的地物压缩或失真表达,因而它完全不同于图像处理中常用的“金字塔”式分层的概念,应加以区分)。这样,在每一格网中记录它所在范围内相应区域的地物OID,再对规则格网进行空间编码,就能建立起无缝地理数据库的空间索引,并优化了空间数据组织。

2.2.4 无缝地理数据库中分层格网的“分级扩散四叉树编码”

如2.2.3节所述,为了建立无缝地理数据库的有效空间索引机制,首先必须建立起分层格网与空间关系的函数表达,对每个格网进行有效的空间编码。

Morton码是一种基于十进制的四叉树编码方法 [Morton, 1966; Mark, 1985],其优点是可通过快速二进制插位运算进行Morton码与对应行列之间的转换运算^[1,2]。这种编码具有局部“聚簇(Cluster)”特点,然而,Morton码要求行列必须为正,只能朝一定方向上扩展。要想用于大型GIS地理数据库的空间索引显然不够,必须能够朝任意方向扩展,因而我们设计出如图6所示的“扩散四叉树编码”方案。

	Y																				
	←																→				
↑	II F	II E	II B	II A	I A	I B	I E	I F	↑												
	II D	II C	II 9	II 8	I 8	I 9	I C	I D													
	II 7	II 6	II 3	II 2	I 2	I 3	I 6	I 7													
	II 5	II 4	II 1	II 0	I 0	I 1	I 4	I 5													
	III 5	III 4	III 1	III 0	IV 0	IV 1	IV 4	IV 5	X												
	III 7	III 6	III 3	III 2	IV 2	IV 3	IV 6	IV 7													
	III D	III C	III 9	III 8	IV 8	IV 9	IV C	IV D													
↓	III F	III E	III B	III A	IV A	IV B	IV E	IV F	↓												
	←																→				

图6 扩散四叉树编码

Fig. 6 Spread Quadtree Coding

在图6中,I、II、III、IV分别代表4个象限,用二进制表示即为00 01 10 11,需要2 bits。I、II、III、IV后所带的十六进制数为该格网在该

象限中的实际Morton码 M_D ,如III B。假设H存放在一个32 bits的长整数的高28-29位上,对于扩散四叉树编码M有:

$$M = (H \ll 28) \& M_D \quad (1)$$

对于2.2.3节中描述的分级格网的层次信息(如图7),为在无缝地理数据库中保证各层格网编码的唯一性,同样我们用4个二进制数00 01 10 11来表达分层信息 β ,也占用2 bits。假设 β 被存放在一个32 bits的长整数的最高两位上,就形成了图7所示的“分级扩散四叉树编码”。所以,对于分级扩散四叉树编码 M_{sd} 有:

$$M_{sd} = (U \ll 30) \& (H \ll 28) \& M_D \quad (2)$$

该编码既能表达空间位置,又能区分层次关系并具有唯一性。它不但可用于矢量地理数据库的空间索引编码,而且同样适用于各种遥感、航空影像的图像数据的分块编码及“金字塔”分层建库。

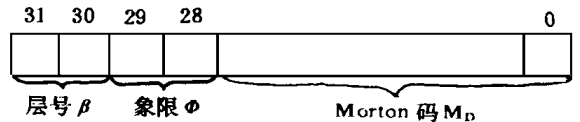


图7 分级扩散四叉树编码 M_{sd}

Fig. 7 Structure of Layered Spread Quadtree

3 结 论

本文提出的大型地理数据库的无缝数据组织模型,已经在武汉测绘科技大学GIS研究中心研制的GIS基础软件GeoStar NT 1.0中加以实现,基于Windows NT 3.51操作系统和MS SQL Server 6.5关系型数据库,试验建立了我国1:100万地形数据库,结果表明该模型是可行的。

大型无缝地理数据库领域中,很多方面如二维、三维甚至时态数据库的建立需要更深层次的研究。本文提出的模型中图形数据和属性数据的关系表的划分、组织,空间索引中格网大小的合理选择,不同地物的空间索引的建立,图形、DEM、影像数据库的一体化融合,Internet/Intranet网上数据库的数据共享、发布、安全性和地理数据的查询、更新等问题都需要更细致、深入地优化和研究。

参 考 文 献

1 Goodchild M. F. Tiling Large Geographical Databases. In: Buchmann A., Githler O., Smith T. R., Wang

Y F, eds. Design and Implementation of Large Spatial Databases. New York: Springer-Verlag, 1990. 137~ 146

2 龚健雅. 整体 GIS 的数据组织与处理方法. 武汉: 武汉

测绘科技大学出版社, 1993.

3 李爱勤, 张 巍. 面向对象的 GIS 数据结构与空间对象索引新机制. 武汉测绘科技大学学报, 1995, 20 (增刊): 43~ 49

Organizing for Large Seamless Geographical Databases

Li Aiqin Gong Jianya Li Deren

(Research Center for GIS, WTUSM, 129 Luoyu Road, Wuhan, China, 430079)

Abstract This paper provides a general model of a large seamless geographical database. A modified Morton key as a means of spatial index and seamless geographic data organizing is presented.

Key words GIS; seamless geographical databases; classified spread quadtree coding

(上接第 17 页)

The Variance Estimation of Corresponding Lines in GIS Overlay Operation

Dai Honglei Xu Panlin Lu Xiushan

(Dept. of Geo-science, Shandong Mining Institute, 223 Daizong Street, Taian, China, 271019)

Abstract This paper extends the variance estimation results of equal accuracy corresponding points to unequal accuracy, derives a general formula to estimate the variance of corresponding lines on any point and any direction, gives the variance estimation formula of the line's length.

Key words corresponding lines; overlay operation; variance estimation