

WebGIS 中的地理关系数据库模型研究

李 琦 杨超伟 陈爱军

(北京大学遥感与地理信息系统研究所, 北京 100871)

摘 要 目前,传统的 GIS 将属性与空间数据分离开来的组织模型已经不能满足 GIS 数据管理的需要,而面向对象的数据库 (OODB) 组织模型又未成熟. 因此,该文研究了将空间与属性数据统一存放在关系数据库中进行管理的技术,建立了用于 WebGIS 的地理关系数据库模型,并在 WebGIS 中得到了原型实现.

关键词 地理信息系统 空间数据库 数据模型

中图分类号: TP208 文献标识码: A 文章编号: 1006-8961(2000) 02-0119-05

Research on Geographical Relational Database Model in WebGIS

LI Qi, YANG Chao-wei, CHEN Ai-jun

(Institute of Remote Sensing and Geographical Information System, Peking University, Beijing 100871)

Abstract The requirement of GISs' data management demands a universal organization model instead of the conventional spatial and attribute departed data management model. The OODB organization technique is far from mature. Under this condition, a research on how to manage the attribute and spatial data in a universal way in the RDB was conducted. The organization of the spatial database is illustrated in four parts: the total database level table used to maintain the whole spatial data; the atlas table used to maintain the layers' information in a project; the map table used to maintain the layer information; the spatial index table used to store the spatial index information among a layer. The storage process of spatial data is illustrated through spatial data storage and spatial data calculation. A Geographical Relational Database Model was constructed and implemented in WebGIS.

Keywords GIS, Spatial database, Data model

0 引 言

在传统的 GIS 中,一般是将空间数据与属性数据完全分开来存放,并以属性表与空间文件一一对应的方法和根据表的关键字与文件中空间对象的标识,来定位空间对象的属性信息.目前,一方面由于 GIS 中管理的数据量逐渐增大,这种分离管理已经不能满足当前数据管理的需要;另一方面,流行的面向对象技术给予了数据管理以新的活力,已开展了大量关于 OODB 的研究^[1],因而也有人建立了基于 OODB 的空间数据管理模型^[2].在这个模型中,每一个空间对象都表示为一个对象,它以其自己的属性方法和事件来满足空间对象管理的需要,例如,一个

空间对象有一种方法可求出被它包含的空间对象,则调用这个方法就可以求出被这个对象包含的所有空间对象.然而由于目前 OODB 技术的不成熟和受计算机处理速度的限制,故这种数据管理方式尚不能实现.为满足 GIS 对数据的统一管理和运行速度这两个方面的要求,我们建立了基于关系数据库的地理关系数据库模型,从而满足了对地理数据的统一管理和运行速度这两个方面的要求,并且在 WebGIS 中得到了原型实现.

1 空间数据的逻辑组织模型

目前,地理空间数据的主要表现形式为地图,而且它们在数据库中都是以层的概念来组织与表达

的. 每一层数据有其相应的属性与空间等信息,其逻辑组织模型见图 1.

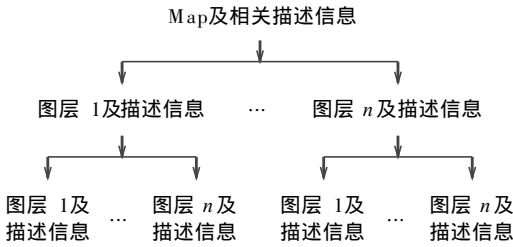


图 1 地理空间数据的逻辑组织模型

如图 1 所示, Map 的组织具有一定的层次关系,而且由若干个空间数据图层及其相关属性数据组织而成. 一个空间数据图层又是以若干个空间坐标的形式存储的. 因此,这一逻辑组织模型可概括为坐标对 空间对象 图层 地图. 一个空间对象及其属性信息在这一模型中属于最基础层次,而地图则是这个模型的最高层次.

为便于空间分析,在图层基础上可以建立相应的空间索引. 一般采用网格划分的方法,就可以建立空间索引,即将图层按一定的规律打上网格,并对每个网格的大小、位置及其中的空间对象信息进行存储记录. 这样,地理数据逻辑组织模型中的信息可进行以下的分类:

(1) 地图集 这是地理数据组织中的顶层信息,它对整个地图集进行管理,并对各个地图进行特化,以识别不同的地图,包含有以下信息:

地图引用 (表名、地图层数等)、地图坐标 (坐标系、配准信息等)、地图描述 (访问权限、地图说明等) 等信息.

(2) 图层集 这是由多个空间图层组成的能满足一定应用需求的图层集合,包含有以下信息:

组成图层集的图层引用 (图层标号、图层表名)、图层空间索引 (大小、标号、表名)、图层显示、图层坐标范围 (坐标最大、最小值) 等信息.

(3) 图层 这是由多个具有某些相同或相似特

性的同种类型的空间对象组成的集合,包含有以下信息:

空间对象的标识 (标号、名称)、空间对象的描述 (名称、特征属性、类型)、空间对象的几何表示 (坐标的二进制大对象——BLOB 形式).

2 基于关系数据库的空间数据模型

关系数据库的层次结构具有与地理数据逻辑组织模型相似的特点^[3],而且关系数据库是建立在关系模型的基础之上的,它的基本组成是表,每个表由列 (表字段)、行 (表记录) 组成,一个数据库则由许多个表组成,这些表之间采用一定的关系组织连接.

本文构建的就是基于这一关系数据库模型的地理关系数据库. 按照层次模型及表之间的连接关系,地理关系数据库的数据组织模型见图 2.

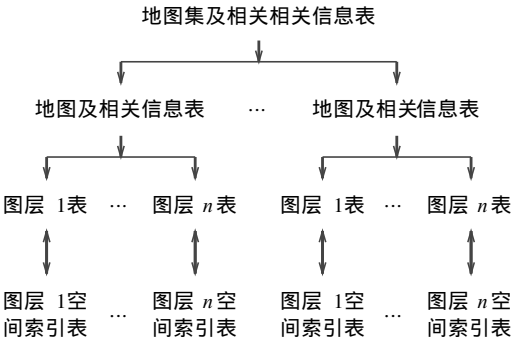


图 2 基于关系数据库的地理数据组织模型

如图 2 所示,各个表具有用于描述本层表格自身的特点的基本信息和可随意添加的多个属性值. 现分别举例说明如下 (以 CyberGIS 工程集和上海市重大工程项目为例):

(1) 地图集表 整个系统具有这样一个表,它可以作为一个地图集中的最高层次表格,在实现上可以作为多个数据库的总体规划表格. 它本身具有各个地图的引用和描述信息,如表 1 所示.

表 1 CyberGIS 工程项目的地图集表

Map ID	Map Name	Layer Numbers	Coordinate System	Map Description
1	上海市重大工程管理	5	...	...
2	云南省发展评价	4	...	...

表 1 中, Map ID 用于地图操作时的地图标识, Map Name 用于引用相应的地图集数据库, Layer Number 用于指明地图中的图层数目, Coordinate Sys-

tem 及 Map Description 用于地图相应信息的描述. (2) 地图表 每个地图都有这样一个表,它可以作为管理一个工程中地图的最高层表格,它是

一个数据库的总体规划表格的方式实现. 该表具有 各层参照、描述和引用信息 (见表 2).

表 2 上海市重大工程项目地图表

Layer ID	Layer Name	Table Name	Spatial Column	Index Name	Layer Description
1	县界	县界	Shape	县界	...
2	重大工程	重大工程	Shape	重大工程	...

表 2 中, Layer ID 用于图形空间操作时的图层标识, Layer Name 为层名, Table Name 用于引用相应的层表, Spatial Column 用于引用层表中的空间字段, Index Name 用于引用相应的索引表, Layer Description 为相应的图层描述信息.

(3) 地图层表 每个地图层都有这样一个表, 它作为管理一个图层的表格. 该表具有本图层的空间对象及各个空间对象的相应描述信息 (见表 3).

表 3 重大工程图层表

Shape ID	Shape Name	Shape Properties	Shape	Shape Type
1	金茂大厦	...	...	Point
2	黄浦江上游引水	...	...	Point

表 3 中, Shape ID 为空间对象在本图层的唯一标识, Shape Name 为相应空间对象的名称, Shape Properties 为空间对象的各种属性信息, Shape 为存储空间对象几何信息, 由坐标点对组成的二进制大对象, Shape Type 为空间对象的种类.

(4) 空间索引表 每个地图层对应一个空间索引表, 它用于协助进行空间分析, 以加快空间分析的速度 (见表 4).

表 4 空间索引表

Grid ID	minX	minY	maxX	maxY	FIDs
1	10.000	10.000	15.000	15.000	...
2	5.000	5.000	5.000	10.000	...

表 4 中, Grid ID 为空间索引划分的网格标号, minX minY maxX maxY 分别为此网格的坐标范围 (经纬度最大、最小值). FIDs 为由此网格中所拥有的空间对象标号组成的二进制大对象.

3 空间数据的存储、运算模型

空间数据在关系数据库中存储时, 是按照二进制大对象 (BLOB) 的形式存储的. 因此, 存储时所要考虑的主要问题有两个^[4]: (1) 空间数据的存储量; (2) 空间数据的访问速度. 针对这两方面的问题, 应依据计算机技术的特点来设计相应的存储结构和算法.

3.1 坐标存储结构

一般来说, 矢量空间数据是按坐标进行存储的, 而且这些坐标均是浮点或双精度数据, 它们占用的存储结构大 (一般为 8byte 存储), 运算速度也比较慢. 众所周知, 在各种十进制数据类型中, 整数是一种占用较少存储空间和具有较快的运算速度的数据类型, 而且它能表示到一定的精度, 可表示 0 到 2 147 483 647 (采用 4byte 32 位整数, 对大多数数据存储和运算来说已足够了). 根据这些特点, 我们可以采取用整数来存储空间数据的策略, 这样可以减少 1 倍存储空间, 为原来的 1/2.

对由整数表示的空间数据, 由于坐标是相邻存储的, 所以各个坐标之间的差别不大, 如果以 4byte 整数表示, 这些差别大多数情况下只有几十、几百, 也就是说, 如果只存储这些差别, 那么就可以只用一个字节来表示. 这样就可以减少 3 倍存储空间, 为原来的 1/4.

这样, 由于表示的方法不同, 可以得到 8 倍的数据压缩率, 如图 3 所示.

浮点、双精度数 $\xrightarrow{\text{压缩率 } 1/2}$ 整数 $\xrightarrow{\text{压缩率 } 3/4}$ 压缩整数

图 3 坐标数据的不同表示与数据压缩率

这些数据是按照字节方式存储的, 但在处理时, 使用整数, 将每个空间对象存储在一个 BLOB 中, 并将这些空间对象的坐标序列存储下来, 对每个点采用整数压缩格式, 而且用 16 进制表示来压缩这种数据, 同时设立控制位 (见表 5).

表 5 空间对象坐标的物理表达

字节数	位数	值的意义
0	0	控制位 (0= 最后一位, 1= 不是最后一位)
0	1	符号位 (0= 正整数, 1= 负整数)
0	2~ 7	低 6 位数据
1~ 4	0	控制位 (0= 最后一位, 1= 不是最后一位)
1~ 4	1~ 7	下 7 位数据

如同各种流式文件一样, 二进制大对象 (BLOB) 坐标也有一个头, 如表 6 所示.

表 6 BLOB坐标二进制头

字节数	用 法
0~ 4	坐标流长度,采用压缩整数格式
5	坐标维数
	0 二维空间数据
	1 三维空间数据
	2 二维空间与量值数据
	3 三维空间与量值数据
6~ 7	保留未用

3.2 坐标压缩运算算法

为将空间对象的坐标转换为整数,需要将空间数据小数点后部分变为整数,同时为减少存储量,应取一个数据密集区中心为原点,进行偏移转换.再在此基础上进行点压缩(如图 4所示).

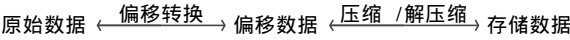


图 4 空间坐标数据转换与压缩过程图

偏移转换算法:

变整转换:

$absolute_value[m] = Original_value[m] \times multiple$
整数压缩转换:

$relative_value[0] = absolute_value[0];$

$relative_value[1] = absolute_value[1] - absolute_value[0];$

...

$relative_value[n-1] = absolute_value[n-1] - absolute_value[n-2];$

位压缩转换:

十进制数 $\leftrightarrow$ 十六进制数 $\leftrightarrow$ 按字节存储读取

其中, $relative_value[m]$ 为第 m 个坐标的存储值;
 $absolute_value[m]$ 为第 m 个坐标的实际整数;
 $Original_value[m]$ 为第 m 个坐标的实际值; $Multiple$ 为转换倍数.

3.3 处理速度分析

由于采用了整数压缩转换算法,因此在实际运算中的各种空间分析和相关运算速度加快(加、减、乘、除等运算速度加快所致),而且除这些运算外,所采用的全部是移位或比较运算,由于这两种处理是计算机运算速度最快的,所以系统的空间处理部分速度比较快.

4 空间数据访问模型

通过地理关系数据库的数据模型,可以将属性

数据与空间数据统一存放在关系数据库中,从而实现了数据的统一管理,实现了如图 5所示的空间数据管理变化.

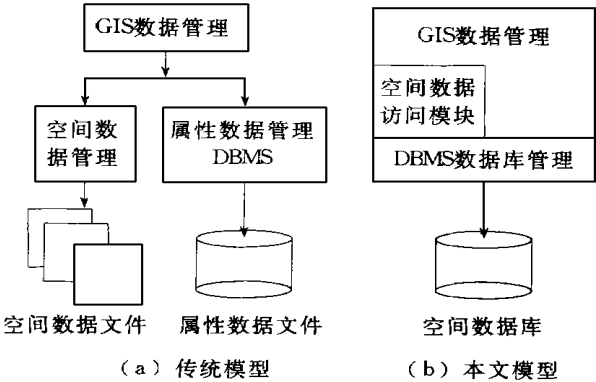


图5 地理空间数据在属性、图形分离和统一于关系数据库管理时的访问对比

由图 5可见,由于将数据统一于关系数据库中进行管理,因而可以直接利用 DBMS的管理功能以充分利用关系数据库中的许多查询优化,提高数据的访问速度.

其中的空间数据访问模块是进行空间数据的压缩、解压缩(实现空间对象的正常表示与空间对象的二进制大对象表示之间的转换)、空间分析、空间数据提取等的模块,它向上提供一般地理信息系统的空间分析、数据提取功能,向下则具有调用数据库管理系统(DBMS)的功能,是整个地理关系数据库模型中数据访问的核心.如图 6所示.

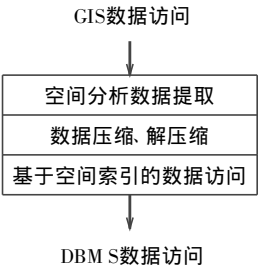


图 6 空间数据访问模块的组成及调用关系

5 基于 WebGIS的具体实现

在 WebGIS中用 ODBC实现了地理关系数据库模型^[5],最终的数据访问界面如图 7和图 8所示,其中图 7为属性数据访问界面,图 8为空间数据访问界面.本系统可以通过该界面来访问基于地理关系数据库模型的数据库中的数据.



图 7 属性数据访问界面



图 8 空间数据访问界面

6 结 论

鉴于当前对将空间数据与属性数据统一组织管理的广泛需求,因此在 OODB 技术不成熟的情况下,作者研究了基于关系数据库的地理关系数据库模型,并在 Web GIS 中用 ODBC 原型实现了这一模型.但限于当前数据库技术的发展水平,关于空间数据与属性数据统一管理的空间数据库的研究还不完善,而且随着空间信息系统的进一步发展,还可能提出对多种信息集成的要求.因此,空间数据模型方面的研究还有待进一步深入,以满足不断增长的空间信息与超媒体信息集成的需要.

杨超伟 1973年生,北京大学遥感与地理信息系统研究所博士生,美国科学促进会专家会员.已在国际、国内刊物上发表论文 10 余篇,其中数篇被 SCI、EI 收录.研究领域为网络化空间信息系统、空间数据仓库、空间信息系统集成、空间信息科学.

参 考 文 献

- 1 Date C J. An Introduction to Database, Mc-Graw Hall, 1997, 1, 2
- 2 Kang S B. Object-relational data modeling for GIS. In: GIS/LIS Annual Conferencing Proceedings, 1995, 1: 535~ 544.
- 3 Mark Ashworth. Spatial database standards. A MapInfo White Paper, 1997.
- 4 ESRI Corporation. Getting started with SDE. An ESRI White Paper, 1997.
- 5 Microsoft Corporation. Microsoft ODBC3.0 Software Development Kit and Programmer's Reference. Microsoft Press, 1997.

李 琦 1955年生,北京大学遥感与 GIS 研究所教授,博士生导师.中国图象图形学会技术委员会主任委员.已在国际、国内刊物上发表论文 50 余篇.研究领域为空间信息科学.

陈爱军 1972年生,北京大学遥感与地理信息系统研究所博士生.已在国际、国内刊物上发表论文数篇.研究领域为地理空间构模语言、空间信息科学.