

Shapefile格式电子海图数据集成存储研究

Research on Integrated Storage of Electronic Chart Data of shapefile Format

(哈尔滨工程大学) 郝燕玲 唐文静 赵玉新 李 宁

HAO Yan-ling TANG Wen-jing ZHAO Yu-xin LI Ning

摘要: shapefile 格式的电子海图数据通常采用文件系统存储,难以保证数据的安全性、一致性、完整性,且数据访问效率低。本文在认真分析 shapefile 格式的电子海图数据组织结构基础上,研究并提出了电子海图空间数据与属性数据集成存储的方法,设计了集成存储电子海图数据的空间数据库结构,并以 Oracle 数据库为平台在 Visual C++ 环境下实现了空间数据与属性数据的集成,有效地解决了文件系统存储海图数据的弊端。

关键词: shapefile 格式; 电子海图; 空间数据库; 集成存储

中图分类号: P208

文献标识码: A

Abstract: The electronic chart data of shapefile format usually was stored in the file system which was difficult to ensure security, consistency, integrity of data, and the efficiency of accessing data was low. Based on seriously analyzing the organizing structure of electronic chart data of shapefile format, the method of storing integrated electronic chart spatial data and attribute data was discussed and put forward, and the spatial database integrated storage structure was finally designed in this paper. The integrated storage of spatial data and attribute data was implemented in Oracle database with VC++. It effectively solves the drawbacks of electronic chart data in file system.

Key words: shapefile format; electronic chart; spatial database; integrated storage

引言

空间数据的分布性和非结构化等特征导致了其存储结构的发展经历了从文件系统存储、混合数据存储、数据库存储方式的演变。近年来,随着计算机技术、图形图像处理技术的迅猛发展,以数字形式表示的电子海图具有功能丰富、灵活性高等特点,因此得到了迅速的发展和广泛的应用。电子海图数据有多种存储格式,shapefile 格式是其中的一种。它采用纯文件方式管理空间数据。这种方式将几何数据与属性数据各自独立组织、管理与检索,仅仅通过唯一标识符来建立它们之间的连接,不但使用起来非常不方便,而且无法保证数据的一致性和完整性。针对这些问题,本文提出了将空间数据和属性数据集成存储的方案,充分利用数据库管理系统处理和分析海量数据的能力,真正实现图形数据与属性数据的融为一体。

1 解析电子海图 shapefile 格式

Shapefile 格式的电子海图数据没有拓扑信息,一个 Shapefile 由一组文件组成,其中必要的基本文件包括图形文件(.shp)、索引文件(.shx)和属性文件(.dbf)三个文件。图形文件为直接存取的可变长记录文件,记录了海图要素的坐标位置数据;索引文件记录描述了对应的图形文件记录相对于图形文件开始点的偏移量;属性文件为 dBase 表文件结构,记录了海图要素的属性。图形文件记录与属性文件记录通过记录号一一对应。

1.1 图形文件的组织

图形文件用于记录空间坐标信息,由文件头和实体信息两部分构成(如图 1 所示)。

文件头是一个长度固定(100 bytes)的记录段,其中包括文件长度、shape 类型、边界盒等信息。每个图形文件只能表示一种图形类型,用不同的整数分别表示。类型值的定义如图 2 所示。实体信息负责记录坐标信息,它以记录段为基本单位,每一个记录段记录一个地理实体目标的坐标信息,每个记录段分为记录头和记录内容两部分。记录头长 8 字节,其内容包括记录号(Record Number)和坐标记录长度(Content Length)两个记录项。记录内容包括目标的几何类型(ShapeType)和具体的坐标记录(X、Y),记录内容的具体内容及格式因要素几何类型的不同而不同。

文件头 (100 字节)	
记录头 (8 字节)	记录内容 (变长度)
记录头 (8 字节)	记录内容 (变长度)
.....
.....
记录头 (8 字节)	记录内容 (变长度)

图 1 图形文件组织

文件头 (100 字节)	
记录头 (8 字节)	记录内容 (变长度)
记录头 (8 字节)	记录内容 (变长度)
.....
.....
记录头 (8 字节)	记录内容 (变长度)

图 2 图形类型

郝燕玲: 博导 教授

基金项目: 国防预研基金(06J3.8.6)

1.2 属性文件的组织

属性文件(.dbf)用于记录属性信息。它是一个标准的 DBF 文件,也是由头文件和实体信息两部分构成。

文件头部分的长度是变长的,它主要对 DBF 文件作了一些总体说明,其中最主要的是对这个 DBF 文件的记录项的信息进行了详细的描述,比如对每个记录项的名称、数据类型、长度等信息都有具体的说明。实体信息部分就是一条条属性记录,每条记录都是由若干个记录项构成。

从 shp 文件和 dbf 文件的内部组织情况可以看出,图形文件中的数据是以二进制形式直接存储,每一个图形数据记录有一个相应的属性数据记录,每一个图形数据在属性数据文件中记录的顺序与在图形数据文件中记录的顺序一致。这种将空间数据和属性数据分开存储的方式具有数据模型简单的特点,适用于数据量较少的系统,但存在诸多不便:

- (1) 不能保证数据的安全性、一致性、完整性。
- (2) 不能处理大量的数据。
- (3) 数据访问的效率低。
- (4) 难以保证多用户并发操作的实现。

2 集成空间数据结构设计

空间数据具有与一般数据不同的存储特征,这主要体现在空间数据是海量数据、数据文件类型复杂、数据使用频率不一致等几个方面。要实现空间数据的有效存储管理,必须实施数据的集中管理。在具体设计存储系统时,应着重考虑以下两个方面:

* 存储空间动态扩充

由于空间数据总是不断增加,因此存储系统应该是一个集中管理的系统,具备可扩充性的特点,这样可以动态扩展存储空间。

* 数据快速存取

数据存储的目的就在于提供应用,存储系统的设计就应实现让用户快速、正确、简便的获取数据,而不用去了解数据实际存储物理位置的功能。

对于现实中的实体可以由一个二维空间信息所表示,包括属性数据和空间数据。空间数据是和空间位置、空间关系有关的数据,而地理元素中非空间的属性信息,称为属性数据。空间数据被分层存储在图形文件中,它是一种由元素、几何实体及层所组成的层次结构,包括点、线、面三种基本要素:

- * 点,定义为没有面积和长度度量的、离散的空间地理要素,由一对 x y 坐标表示。如灯塔、码头等。
- * 线,定义为由一串有序的 x y 坐标对表示,如边界、河流等。
- * 面,由一串有序的,且首尾坐标相同的 x y 坐标对表示,如岛屿等。

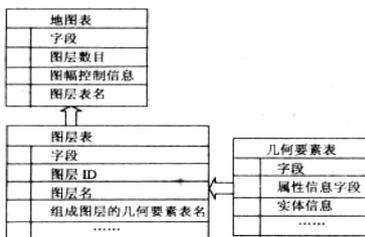


图3 电子海图数据逻辑结构

根据空间数据的层次结构来设计空间数据库,其数据结构

包括几何要素表、图层表、地图表。地图由图层组成,而图层由几何要素组成,其空间数据逻辑结构如图3所示。

(1) 几何要素表,根据组成几何实体的要素进行设计,存储该图层的所有空间实体的空间和属性信息,所以几何要素表包括两部分内容:

* 实体信息,包括几何要素类型、空间坐标系、空间坐标信息等。

* 属性信息,不同图层表示不同地理现象,海图基础地理信息要素共分18类,包括水域/陆地、礁石沉船与障碍物等。不同几何类型的属性数据项是不同的,如点类型的属性数据文件包括3个标准数据项,分别是 ARC、COVER、COVER_ID,线类型的属性数据文件包括7个标准数据项,分别是 FNODE、TNODE、LPOLY、RPOLY、LENGTH、COVER、COVER_ID,面类型的属性数据文件包括4个标准数据项,分别是 AREA、PERIMETER、COVER、COVER_ID。这部分域段在创建几何要素表时指定。另外,根据图层表示的地理现象的不同,属性数据文件还增加了对它们特有的属性进行描述的扩展字段。

(2) 图层表,它记录了数据库中的图层信息,包括图层ID、图层名、图层包括的几何要素表名等。图层表体现了图层与几何要素表的对应关系,根据图层所表示的要素的几何类型,每个图层可以由1~3个几何要素表组成。

(3) 地图表,主要记录了该幅地图的图层数目、图层表名、图幅控制信息,包括图号、图名、投影方式、比例尺、图幅范围等信息。

这样一种组织形式基于对象关系模型,将不同几何类别的空间要素对象化,这样空间数据和属性数据一同存储在数据库中,使得对空间数据的并发操作更加方便和迅速。该数据库按照图层来对电子海图数据实现存储,每个图层实现一个专题空间图形,不同几何类别的要素数据存放于不同的表中。这种管理机制保证了数据的安全性、完整性、一致性。

3 电子海图数据集成存储的实现

线状要素属性的标准数据项

字段名	FNODE	TNODE	LPOLY	RPOLY	LENGTH	OCLINT	OCLINT_ID
字段类型	N	N	N	N	F	N	N
长度	11	11	11	11	20	11	11
小数位数	0	0	0	0	5	0	0
能否为空	Not null	Not null					
备注	左节点的内部码	右节点的内部码	左多边形的内部码	右多边形的内部码	长度	内部标识码	用户标识码
字段名	名称	名称	名称	名称	名称	名称	名称
字段类型	N	C	N	C	N	MDSYS.SDO_GEOMETRY	MDSYS.SDO_GEOMETRY
长度	11	40	5	40	变长		
小数位数	0	0	0	0			
能否为空	NULL	NULL	NULL	NULL	Not null		
备注	编码	名称	深度	备注			几何实体信息

图4 线状要素属性的标准数据项

其中, N 代表数值型, F 代表浮点型, C 代表字符型, MDSYS.SDO_GEOMETRY 代表对象类型

图4 线状要素属性的扩展数据项

图4 线状要素实体信息数据项



图5 实现界面

按照面向对象思想,每种地理现象都可以被抽象为某一类

技术创新

