

基于差异化存储的时空数据库的设计与实现

王宇清 李 建 唐开山 吴 军 罗光红 (西南石油大学 计算机科学学院 四川 成都 610500)

摘 要: 首先简要分析了时空数据库的功能和特点及几种典型的时空数据模型。然后在面向对象时空数据模型的基础上引入了差异化版本存储的概念,讨论了借助版本差异信息设计一种历史数据冗余小,当前数据利用率高的时空数据模型的基本方法。最后实现了一个具有基本时态功能且能有效平衡数据库存储容量和系统执行效率的地理信息管理系统。

关键词: 时态 GIS; 差异化; 版本管理; 时空数据库; 数据模型

Design and Implementation of Difference Storage-Based Spatiotemporal Database

WANG Yu-Qing, LI Jian, TANG Kai-Shan, WU Jun, LUO Guang-Hong

(School of Computer Science, Southwest Petroleum University, Chengdu 610500, China)

Abstract: This paper has analyzed the functions and features of Spatiotemporal Database and its model in brief. The conception of difference storage of version is used on the basis of object-oriented spatiotemporal data model. And then, a basic method of designing Spatiotemporal Data model with less historical data redundancy and high usage of current data is discussed. A Spatial data management system with basic functionality and effective balance of storage and operation is also implemented at the end of the paper.

Keywords: temporal GIS; difference; version management; spatiotemporal database; data model

1 引言

随着地理信息技术和计算机应用技术的发展,有关地理信息的管理、分析和利用越来越受到人们的关注。人们发现地理信息不仅具有三维的空间属性同时也具有很强的时间属性,在许多应用领域都需要分析和处理其“过去”、“现在”,甚至“未来”的信息。如何对这些具有时态特征的地理数据进行有效组织和利用成为开发人员和相关决策者所共同面临的难题。本文从地理信息系统(Geographical Information System 简称 GIS)的时态存储需求出发,借助版本管理的思想,采用差异存储的方式对构建合理的时空数据库模型做了有益的探讨,根据实际情况的需要较好的平衡了信息量和存储空间之间的矛盾。

2 时态GIS相关概述

时态地理信息系统(Temporal GIS 简称 TGIS)是一种采集、存储、管理、分析与显示地理实体随时间

变化信息(或时空信息)的计算机系统^[1]。它的核心是时空数据库的设计和管理。与传统 GIS 相比, TGIS 的时空数据模型在空间维度和属性维度的基础上增加了时间维度而构成了三维(无高度维)或四维数据模型^[2]。时空数据模型研究的关键问题是如何有效的集成时间维和空间维,满足对空间对象历史信息的高效管理。

2.1 时空数据库的特点和功能

由于采用了诸如时间点和时间段等时间原语对空间对象的标示,使得时空数据库不仅能反映空间对象当前的存在状态,而且还能表达其发展变化过程及规律,与传统的数据库相比具有动态性和全面性的特点。传统的数据库基本上不存储旧的、过时的数据,在数据库更新时,这些数据将从数据库中删除,这就不能反映出现实世界的动态过程,这在处理实时变化的数据中将是一个极大的瓶颈^[3,4]。而时空数据库中,它能把数据对象变更过程都保存下来,使数据库中存储的信息和现实世界中信息的变化随时保持着动态交换。将时态的概

念引入 GIS,为地理信息动态实时监测和分析提供了丰富的数据资源,从而提供各种决策支持和分析服务。

2.2 时空数据模型及研究现状

当前具有代表性的时空数据模型包括:序列快照模型(Sequent Snapshots)、时空立方体模型(Space-time Cube)、基态修正模型(Base State With Amendments)、时空复合模型(Space-time Composite)。但这些模型在实际的应用中都有一定的局限性有的甚至不具备实用的开发价值。比如序列快照模型对所有时空特征进行连续存储时将产生大量的数据冗余。基态修正模型在处理给定时刻的时空信息时需要经过复杂的计算来获得,处理难度大、效率低^[5]。

龚健雅提出了一种面向对象的时空数据模型。这种模型以其规范化空间对象模型为基础,采用面向对象的方法,将对象版本信息加注在对象标识中,直接表达同一对象不同时期多个版本的属性值^[6]。这样节省了存储空间,而且给空间对象的检索和历史事件的查询带来了方便。

3 差异化版本管理的时态数据模型

3.1 基本概念和原理

前文提到地理信息具有随时间的变化而不断的变化的特点。比如地籍的变更、城市的演化他们所描述的对象信息都是处在时空变化之中。版本管理则是记录并管理数据库在变更、演化过程中各个阶段的状态信息。一个版本就是数据库在一个阶段的逻辑快照,它并不复制数据库,但却反映数据库在那一阶段的全貌^[7]。版本与其描述的数据对象具有多对一的关系,不同的版本记录了这个对象的演变历史。因此在时空数据库中引入版本管理的概念可以较好的处理历史数据,体现时空数据库的两大基本特征。

本文中所描述的差异化版本管理的数据模型是以不改变原数据库关系模型为前提的一种处理时态地理信息的数据模型。总体思路是:首先利用程序控制的方法将不同时刻同一对象快照之间的时态属性的差异提取出来,作为一种特殊的属性存储在一个独立的版本信息库中,最大限度的保留原空间数据库的稳定性的同时实现 GIS 的时态处理功能。采用差异化版本管理的方式实现了将数据对象与它的版本信息在存储结构上实现分离,在逻辑结构上依靠提取出的差异信息

和对象标识进行关联。这样不仅可以在不改变原关系模式数据库的基础上,实现时态化信息管理,而且还降低了数据库的冗余,提高了空间数据库的查询分析效率。

3.2 数据模型的设计

基于差异化版本管理的时态 GIS 的底层数据库分为两部分,我们将其分别称为资源库和版本库。资源库中的数据对象始终和现实世界中的对象保持一致,即在资源库中我们存储的是当前现实世界中实际存在的对象状态值,如果单独看待这个库可以认为它是当前时刻数据对象的一个快照数据库。版本库中我们存放的则是数据对象的历史记录,也就是所有对象的版本信息。在处理两个库的关联关系上,我们规定每个数据对象有唯一标识,它可以对应多个版本信息编号,无论其经历多少次版本变化这个标识始终不变。这样我们就可以很方便的在版本库中找到资源库中各个对象所对应的版本差异信息。

其时态特性可以通过资源库的入库时刻和版本库的有效时段(valid-time)共同实现的。数据对象的入库时间可以视为其有效时间的起始时间,也就是数据库事务操作时间(transaction-time)。如果没有信息的更新,该对象的这些属性将一直有效,直到当其某些时态属性值发生更改。发生更改的事务时间也就是原属性的有效时间的终止时间。

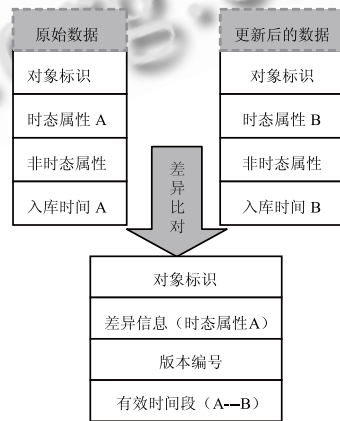


图 1 数据对象时态演变图

如上图所示,当更新后的对象与资源库中对象进行一次差异比对,发现某些时态属性发生了改变,我们就将这一差异信息存入版本库中。这样版本库中新增了一条以该对象为标识的版本差异信息。而改变后

的属性则可以直接覆盖资源库中的原属性值，因为原属性已被当作差异信息存入了版本库中作为一个历史版本。其差异信息处理遵循以下规则。

表 1 差异信息比对规则

同一对象在不同版本中的状态		差异信息	
A 版本	B 版本	差异类型	差异记录
存在	不存在	删除	A 版本属性
不存在	存在	增加	B 版本属性
存在	存在	可能更新	A 版本属性

4 应用实例

为验证数据模型的正确性和实用性，笔者实现了一个基于差异化版本管理的时空数据库系统。系统的逻辑结构示意图如图 2：

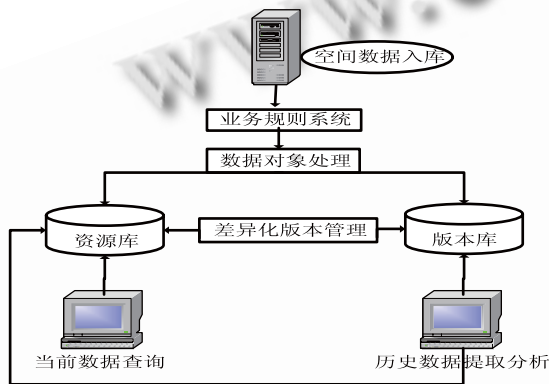


图 2 系统逻辑结构

空间数据库由 Oracle 进行管理。其中资源库中存放的是最新版本数据，版本库中则存放历史版本间的差异信息。数据对象处理模块用来完成新版数据入库时版本比对和差异信息的生成。差异信息以二进制流的形式存储在版本库中。在实际程序处理中我们将差异信息看作一种数据类型，通过序列化和反序列化的方式在二进制流和差异信息对象数组之间进行转换。差异化版本管理模块则是在获得差异信息对象数组的基础上封装了与资源库中数据对象进行组合的方法，以实现历史版本回滚和版本间差异比对等功能。

为了清楚表达何为时空数据对象的概念，我们采用了目前在测绘导航领域广为使用的网格划分方式，我们将以广州市番禺区所在的网格区域作为一个空间对象来验证系统功能。

版本表编号	MESH编号	版本编号	差异信息	起始有效时间	终止有效时间	版本状态标记
932	F49H019176	2	0x0001001	2009-4-23 15:27	2009-7-19 13:46	0
875	F49H019176	1	0x0001001	2009-3-12 09:36	2009-4-23 15:27	0

图 3 版本信息截图

上图中记录的是该对象 7 月份更新版本库后的情况。执行了新版本入库操作，版本 2 中以二进制流形式保存的信息是 7 月份与 4 月份两个版本间的差异。该对象在版本库中拥有两条版本差异信息。说明这个对象已经过了三次数据更新(进行了两次差异比对)如下面截图所示它具有三个版本的信息。



图 4 对象的三个版本信息截图

选取 4 月份的版本和 7 月份版本进行比对，系统将版本库中第二个版本的差异信息反序列化为差异信息对象数组，并找出了新增 POI(地理位置兴趣点)6 个。



图 5 版本差异详情截图

结合这些差异信息以及 7 月份资源库中该空间对象信息，我们可以很快得到 4 月份的历史记录。通过本例我们可以看出，利用差异信息可以找回历史版本，同时还能提供分析变化趋势的依据。由于在资源库中我们只存放当前最新的数据，历史数据全部转换成为了差异信息存放在版本库中，这样不仅大大减少了存

(下转第 126 页)

(上接第 139 页)

储空间还更加便于资源的管理和对当前信息的查询统计效率。此外系统还可以对多个历史版本进行查询和回滚操作,限于篇幅,不再举例说明。

5 小结

基于差异化版本管理的时态数据库采用了将当前空间数据和其历史数据分离的方法,增加适当的版本差异控制机制,实现了具有基本功能的时态 GIS。本文所论述的方法满足了现实需要,具有简单易扩展的特点,为空间数据库的改建降低了难度和开发成本,在保持原数据库稳定,减少历史信息数据冗余,提高对当前地理数据快照(资源库)查询分析效率上是一次有益的尝试。

参考文献

1 王家耀,魏海平,成毅,熊自明.时空 GIS 的研究与进

展.海洋测绘,2004,24(5):1 - 4.

2 刘茂华,杨伦,王峰.时空数据库及其版本化管理.辽宁工程技术大学学报,2005 增刊:62 - 64.

3 汤庸.时态数据库导论.北京:北京大学出版社,2004. 15 - 19

4 DonghoKim, KeunHoRyu, HongSoonKim. A Spatio-temporal database model study and query language. Journal of systems and software, 2000.

5 刘仁义.地理信息系统.北京:高等教育出版社,2002.

6 龚健雅.当代 GIS 的若干理论与技术.武汉:武汉测绘科技大学出版社,1999.73 - 7.

7 任娟,杨平.Geodatabase 版本控制原理剖析.国土资源化,2005,1(3):41 - 44.

8 Christian S. Jensen Temporal Database Management. <http://www.cs.aau.dk/~csj/Thesis/>