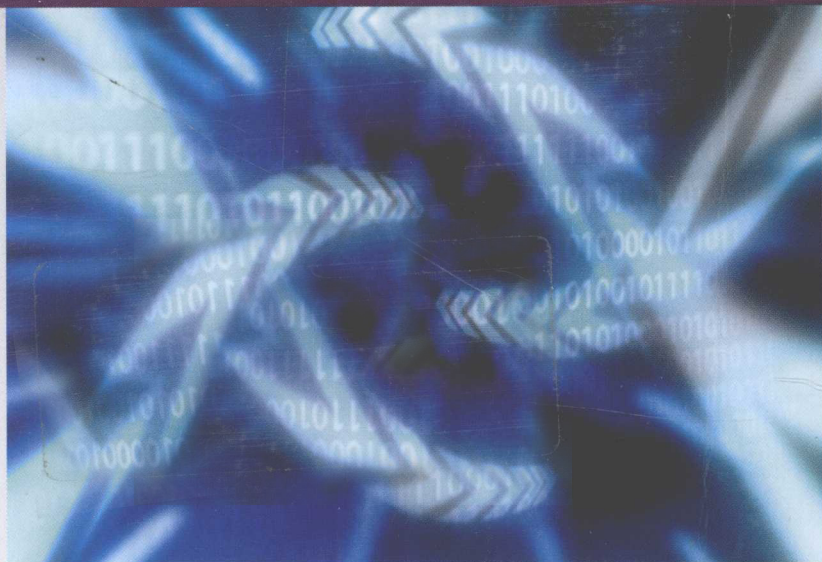


数据库应用理论系列图书

# 时空数据库 新理论

郝忠孝 著



科学出版社

数据库应用理论系列图书

# 时空数据库新理论

郝忠孝 著

科学出版社

北京

## 内 容 简 介

本书系统论述和分析了时空数据库、空间数据库、移动数据库等查询、空间数据推理新技术和新理论。全书共分8章。内容包括:基于Voronoi图的线段反向最近邻查询;空间数据间的序及数据查询;不确定性平面、区域的平面线段的近邻查询、平面动态线段的索引和查询;基于聚类的空间数据索引优化及代价分析;高维主存 $k$ NN查询, $\Delta$ -tree的非递归、递归深度优先,自底向上深度递归 $k$ NN查询,自顶向下主存 $\Delta$ -tree的高维数据相似连接;区域对象间主方向关系的反关系推理、三维空间方向关系的表达与推理、基于定性坐标的位置表达及主方向关系推理等。

本书可作为计算机科学与技术学科、地理信息系统、机器人技术、人工智能、卫星遥感、气象分析、地质灾害分析等专业的高年级本科生或硕士生选修课教材,也可供从事上述领域研究的博士生、科研人员及工程技术人员等参考。

### 图书在版编目(CIP)数据

时空数据库新理论/郝忠孝著. —北京:科学出版社,2011  
(数据库应用理论系列图书)  
ISBN 978-7-03-032313-2

I. ①时… II. ①郝… III. ①数据库系统 IV. ①TP311.13

中国版本图书馆CIP数据核字(2011)第184688号

责任编辑:耿建业 / 责任校对:陈玉凤  
责任印制:赵 博 / 封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街16号  
邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

\*

2011年9月第 一 版 开本:B5(720×1000)  
2011年9月第一次印刷 印张:14 1/2  
印数:1—2 500 字数:273 000

定价:60.00元

(如有印装质量问题,我社负责调换)



## 作者简介



郝忠孝,教授,山东蓬莱人,1940年12月生,中共党员,曾任原东北重型机械学院副校长,齐齐哈尔大学副校长,哈尔滨理工大学学术委员会主席。现任哈尔滨理工大学博士生导师、哈尔滨工业大学博士生导师(兼)。原机械电子工业部有突出贡献专家、享受国务院政府特殊津贴、全国优秀教师、黑龙江省共享人才专家、黑龙江省级学科带

头人、黑龙江省计算机学会副理事长。

主要研究领域:①空值数据库理论。在国内外首次提出了空值数据库数据模型,完成一系列相关研究,形成了比较完整的理论体系,著有国内外第一部该领域的论著《空值环境下数据库导论》。②数据库 NP-完全问题的求解问题。首次基本解决了求全部候选关键字、主属性,基数为  $M$  的候选关键字,最小候选关键字等问题,著有《关系数据库数据理论新进展》一书。③数据库数据组织的无环性理论研究。在无  $\alpha$  环、无  $\beta$  环、无  $\gamma$  环的分解条件与规范化理论研究方面有了突破性进展,著有《数据库数据组织无环性理论》。④时态数据库理论研究。系统提出并完成了时态数据库中基于全序、偏序、多粒度环境下的各种时态理论问题研究,著有《时态数据库设计理论》的论著。⑤主动数据库理论研究。著有国内外第一部该方面的论著《主动数据库系统理论基础》。⑥空间、时空数据库理论研究。首次解决了空间数据库线段最近邻查询的问题,著有《时空数据库查询与推理》、《时空数据库新理论》等。⑦不完全信息下 XML、概率 XML 数据库理论研究。首次解决了不完全信息下 XML 数据库部分理论研究问题,著有《不完全信息下 XML 数据库基础》。

发表学术论文 230 余篇,其中,国家一级论文 160 余篇;在《计算机研究与发展》上发表个人学术论文专辑两部,被 SCI、EI 等检索 140 余篇。著书 8 部。

## 前 言

数据库技术和计算机网络技术已成为当今世界计算机应用中两个最重要的基础领域。针对一些特殊领域的应用,时态数据库、空间数据库已经成为现代数据库的两个重要分支。但是,单独的时态数据库和空间数据库已经无法满足需求,时空数据库的诞生成为必然。作者于2010年出版了《时空数据库查询与推理》一书,较为全面的讨论和论述了时空数据库中的空间、移动及时空数据库中的查询与空间推理。本书是对近几年最新成果的讨论与汇集,是《时空数据库查询与推理》一书的延续,将对空间、移动及时空数据库最近邻和其他类型的查询所取得的部分成果分别作出讨论和较为详细介绍。

本书共分8章。为了便于读者理解,第1章概略介绍了有关时空数据库的一些基本概念。第2章讨论了基于QR-树的受约束空间连接查询,基于Voronoi图的线段反向最近邻查询,变体-障碍反向最近邻查询;第3章讨论了空间数据间的序及数据查询;第4章讨论了基于不确定性平面、区域的平面线段的近邻查询,平面动态线段的索引和查询;第5章讨论了基于聚类的空间数据索引优化及代价分析、计算的索引结点优化、基于聚类分析的结点优化;第6章讨论了自顶向下主存 $\Delta$ -tree的高维数据相似连接;第7章讨论了 $\Delta$ -tree的非递归、递归深度优先、自底向上深度递归 $k$ NN查询;第8章讨论了基于MBR的主方向关系、区域对象间主方向关系的反关系推理,三维空间方向关系的表达与推理,基于定性坐标的位置表达及主方向关系推理等。所有这些在《时空数据库查询与推理》一书都是没有讨论的。

如果通过本书的讨论能够起到抛砖引玉的作用,正是作者所期盼的。另一个目的是作者从事数据库理论研究工作三十余年,作为一种责任想把研究的一些结果留给年轻同志,使有兴趣的同行和读者深入到这一领域。

本书以时空数据库各类查询为主线,力求用通俗易懂的语言来较为全面、系统地讨论它们,做到条理清晰、逻辑性强、易于理解。

本书的出版得到了哈尔滨理工大学领导,科学出版社林鹏、耿建业同志的大力支持和帮助,在此表示衷心感谢。

我的学生李松、刘润涛、刘艳、宋广军、王森、孙冬璞、杨泽雪等博士为本书做了富有成效的工作。本书的所有图表的绘制分别由李松、张丽平、郝晓红完成。在此书出版之际,表示诚挚的感谢。

由于作者水平有限,书中难免会有纰漏和疏忽,敬请读者批评指正。

作 者

2011年7月于哈尔滨

# 目 录

## 前言

<b>第 1 章 时空数据库基础</b> .....	1
1.1 时空数据库概述 .....	1
1.2 空间数据库概述 .....	5
1.2.1 空间数据结构 .....	5
1.2.2 空间查询 .....	5
1.2.3 空间数据库索引技术 .....	6
1.2.4 空间索引的基本思想 .....	7
1.2.5 空间对象近似化 .....	8
1.2.6 空间查询处理步骤 .....	8
1.3 几种典型的空间索引结构 .....	9
1.3.1 B-树和 B <sup>+</sup> 树 .....	9
1.3.2 R-树 .....	11
1.3.3 R <sup>+</sup> 树 .....	12
1.3.4 四叉树 .....	13
1.4 空间网络数据库概述 .....	13
1.5 移动对象数据库概述 .....	14
1.5.1 移动对象的描述 .....	14
1.5.2 移动对象信息存储 .....	15
1.5.3 移动对象的位置表示 .....	15
1.5.4 移动对象位置不确定性的表示与处理 .....	16
1.5.5 移动对象的索引 .....	18
1.5.6 移动对象的查询 .....	19
1.6 时空数据模型和索引 .....	20
1.6.1 时空数据模型 .....	20
1.6.2 时空数据索引 .....	21
1.7 本章小结 .....	22
<b>第 2 章 空间数据库几种查询方法</b> .....	24
2.1 基于 R-树的最近邻查询理论基础 .....	24
2.1.1 最近邻查询的定义 .....	24

2.1.2	最近邻查询的测量距离 .....	25
2.2	基于 R-树的受约束空间连接查询 .....	28
2.2.1	基于 R-树的受约束空间连接查询的直接方法 .....	28
2.2.2	基于 R-树的受约束空间连接查询算法 .....	28
2.3	基于 QR-树的受约束空间连接查询 .....	30
2.3.1	QR-树 .....	31
2.3.2	基于 QR-树的受约束空间连接查询算法 .....	32
2.4	基于 Voronoi 图的线段反向最近邻查询 .....	38
2.4.1	线段 Voronoi 图的定义和性质 .....	38
2.4.2	基于线段的反向最近邻 .....	39
2.4.3	线段的查询区域 .....	40
2.4.4	判断线段与查询区域相交的方法 .....	41
2.4.5	Voronoi 图的线段反向最近邻查询算法 .....	41
2.5	空间数据库中的障碍反向最近邻查询 .....	43
2.5.1	可视性判断 .....	43
2.5.2	障碍距离的计算 .....	46
2.5.3	障碍反向最近邻查询算法 .....	48
2.6	本章小结 .....	53
<b>第 3 章</b>	<b>空间数据间的序及数据查询 .....</b>	<b>55</b>
3.1	数据空间的二分划分 .....	55
3.1.1	空间数据间的序 .....	55
3.1.2	二分划分 .....	57
3.2	数据空间的四分划分和区域查询 .....	58
3.2.1	极小化交叠的数据空间的四分划分 .....	59
3.2.2	具有相对位置关系的数据空间的四分划分 .....	60
3.2.3	RQOP 树的结点插入 .....	62
3.2.4	RQOP 树的区域查询 .....	63
3.3	数据空间的 M 分划分和 MB-树的区域查询 .....	63
3.3.1	MB-树的生成 .....	64
3.3.2	MB-树的结点插入 .....	66
3.3.3	MB-树的区域查询 .....	68
3.4	基于序的空间数据索引 MOIS-树的区域查询 .....	70
3.4.1	MOIS-树的生成算法 .....	71
3.4.2	MOIS-树的结点插入算法 .....	74
3.4.3	基于 MOIS-树区域查询的剪枝规则 .....	76



---

3.4.4	基于 MOIS-树区域查询算法	78
3.5	基于 MOIS-树的最近邻和 $k$ 最近邻查询	81
3.5.1	点与物体的空间关系定理	81
3.5.2	基于 MOIS-树的最近邻查询算法	82
3.5.3	$k$ 最近邻查询算法	85
3.6	本章小结	89
<b>第 4 章</b>	<b>基于不确定性平面线段的查询</b>	<b>91</b>
4.1	线段与线段的空间关系描述	92
4.1.1	带有不确定影响区域线段的描述	92
4.1.2	线段与线段之间的空间关系	93
4.2	不确定性区域的平面线段的近邻查询	95
4.3	基于 Vague 集的平面线段不确定性区域	98
4.3.1	线段的模糊划分描述	98
4.3.2	平面线段的 Vague 区域描述	100
4.3.3	平面线段的 Vague 区域表示	100
4.3.4	平面线段的动态规律描述	101
4.4	平面动态线段的索引和查询	105
4.4.1	平面动态线段的索引	105
4.4.2	线段的近邻查询过程	106
4.5	本章小结	109
<b>第 5 章</b>	<b>基于聚类的空间数据索引优化及代价分析</b>	<b>111</b>
5.1	基于计算的索引结点优化	111
5.1.1	$R^*$ 树索引简介	111
5.1.2	正交区域扫描	112
5.1.3	创建二叉线段树	114
5.1.4	MBR 重叠区域计算	115
5.2	基于聚类分析的结点优化	118
5.2.1	结点的紧致结构	118
5.2.2	聚类结点 MBR 重叠的判定	119
5.2.3	DLSP 判定算法实例分析	124
5.3	改进的空间划分类索引	125
5.3.1	基本知识	125
5.3.2	索引的动态更新代价模型	128
5.4	本章小结	131

<b>第 6 章 基于主存 <math>\Delta</math>-tree 的高维数据查询</b> .....	133
6.1 理论基础 .....	133
6.1.1 主成分分析 .....	133
6.1.2 $\Delta$ -tree .....	134
6.2 高维主存 $k$ NN 连接索引结构的基础算法 .....	135
6.2.1 $\Delta$ -tree- $R$ 的基础算法 $R\_insertR$ .....	135
6.2.2 $\Delta$ -tree- $S$ 的基础算法 $R\_insertS$ .....	138
6.2.3 相关性质及定义 .....	140
6.3 非递归深度优先 $k$ NN 查询 .....	141
6.4 递归深度优先 $k$ NN 查询 .....	147
6.5 自底向上深度递归 $k$ NN 查询 .....	152
6.5.1 相关子算法 .....	152
6.5.2 BU_DF_knn_Search 算法 .....	156
6.5.3 三种主存 $k$ NN 查询算法比较 .....	157
6.6 本章小结 .....	158
<b>第 7 章 基于主存 <math>\Delta</math>-tree 的高维数据连接</b> .....	159
7.1 自顶向下主存 $\Delta$ -tree 的高维数据相似连接 .....	159
7.2 改进的基于 $\Delta$ -tree- $R$ 的 $k$ NN 连接 .....	165
7.2.1 基于 $\Delta$ -tree- $R$ 的 $k$ NN 连接算法子算法 .....	165
7.2.2 改进的基于 $\Delta$ -tree- $R$ 的 $k$ NN 连接算法 .....	171
7.3 基于 $\Delta$ -Rdnn-tree 的自连接 .....	172
7.3.1 反向 $k$ 最近邻索引结构 $\Delta$ -Rdnn-tree .....	172
7.3.2 基于 $\Delta$ -Rdenn-tree 的 $k$ NN 自连接算法 .....	174
7.4 基于 $\Delta$ -Rdnn-tree 的反向 $k$ 最近邻连接 .....	176
7.5 基于 $\Delta$ -Rdnn-tree 的反向 $k$ 最近邻查询 .....	178
7.6 本章小结 .....	180
<b>第 8 章 空间方向关系的关系推理</b> .....	181
8.1 基于 MBR 的主方向关系的反关系推理 .....	181
8.1.1 二维空间主方向关系 .....	181
8.1.2 基于 MBR 的主方向关系的反关系推理算法 .....	183
8.2 区域对象间主方向关系的反关系推理 .....	185
8.2.1 矩形主方向关系的原关系 .....	185
8.2.2 主方向关系的反关系推理算法 .....	187
8.2.3 算法验证 .....	189
8.3 三维空间方向关系的表达与推理 .....	190

---

8.3.1 三维空间主方向关系模型 .....	191
8.3.2 三维空间方向关系推理 .....	194
8.4 基于定性坐标的位置表达及主方向关系推理 .....	204
8.4.1 空间距离关系 .....	204
8.4.2 定性位置表达 .....	205
8.4.3 基于定性位置的主方向关系推理 .....	209
8.5 本章小结 .....	214
参考文献 .....	216

# 第 1 章 时空数据库基础

## 1.1 时空数据库概述

概略地说,时空数据库是空间数据库(Spatial Database, SDB)与时态数据库(Temporal DataBase, TDB)系统的、有机的结合体。

空间数据库、时态数据库、移动数据库是时空数据库的特例。

空间数据库与传统关系数据库相比,具有更广泛的应用前景,它能够支持多种空间数据模型、空间抽象数据类型以及一种能够调用这些抽象数据类型的查询语言,并支持索引、查询、更新操作以及用于查询优化的特定领域规则。

空间数据指的是用于表示空间物体的位置、形状、大小和分布特征等诸方面信息的数据,适用于描述所有二维、三维和多维分布的关于区域的物体。空间数据的特点是不仅包括物体本身的空间位置及状态信息,还包括表示物体的空间关系(即拓扑关系)信息。属性数据指的是非空间数据,用于描述空间物体的性质,对空间物体进行语义定义。

时空数据库理论是伴随着现代生产、生活的需要而产生的。任何实体都有其时间(时态)和空间特性。在许多现实应用中,实体本身及实体之间的时态、空间、时空关系往往成为人们需要对实体进行处理的主要方面。这些应用的共同点是:要求能够比较真实的记录并处理实体的时态信息、空间信息、空间信息随时态的变化而发展变化的信息等。这就要求面向这类应用的数据库系统除了能处理基本类型的数据外,还必须能处理空间类型(描述静态空间对象的空间信息),时态类型(描述对象的时态信息)和时空复合类型(Spatial-Temporal Data Types)(描述动态空间对象的空间信息随时态的变化而发展变化的信息)。但是现有的数据库或者只能有效处理时态对象(时态数据库),或者只能有效处理静态的空间对象(空间数据库)。

空间数据库一般不保存历史变化或保留若干典型时间点的全局状态快照序列,具有较弱的时空语义建模能力,无法提供时态分析功能,常常被称为静态空间数据库。时空数据模型是一种有效组织和管理时态数据、属性、空间和时间语义更完整的数据模型。

实际上,时空数据库的研究是对位置由于时间变化而变化的对象的相关信息数据进行时空对象表达、时空数据建模、时空数据索引、时空数据查询和时空数据

库体系结构等。同时,时空数据库原型系统、时空推理、时空查询代价模型也为时空数据库的研究带来了一定的挑战。

时空数据库是能够同时处理时态数据和空间数据的数据库系统。在时空数据库中通常管理着两类空间对象,一类是静态的空间对象,另一类则是移动对象。所谓移动对象就是指随时间的变化位置也在不断变化的物体,移动对象的特点是在任意时刻都同时具有时间和空间特性。尽管传统的数据库技术为移动对象的管理提供了基础,但要在数据库中表示移动对象的信息,还需要考虑移动对象所独有的特性,即移动性。

随着无线通信和全球定位技术的发展,移动对象轨迹的概念变得越来越重要。

移动对象数据库是对移动对象的位置和其他相关信息的描述、存储和处理的时空数据库系统。

而对用户来说,时空数据查询是对时空数据库的最直接应用,一般通过时空查询语言来表述用户的空间查询和空间分析推理的请求,从而使用户能够与系统进行交互。因此,设计一种简单易用、功能强大的时空查询语言一直是时空数据库技术研究的一个重要内容。时空查询与分析推理是时空数据库应用的核心,是其最重要和最具有魅力的功能,也是时空数据库有别于其他数据库的本质特征。但是,目前时空数据库中空间分析推理功能的发展显得比较落后。因此,研究和空间分析推理的理论成为研究重点。空间分析推理是建立在空间对象位置和属性表达以及对象间复杂空间关系表达的基础上,若要提高空间分析推理能力,必须解决空间关系描述与表达。因此,研究拓扑关系成为促进时空数据库更大发展的迫切要求,对推动时空数据库的发展有着重要的意义。

时空数据库中的数据类型包括三类:常规数据类型(基本数据类型)、时态数据类型和空间数据类型。

时态数据类型(Temporal Data Types)根据时间的功能和结构有两种分类方式,有效时间类型和事物时间类型是时态数据类型的主要组成部分。有效时间类型可以是时间点和时间区间,事物时间类型只能是时间区间。时间间隔是指用一系列时间点组成的时间片。时空数据库的主要研究对象是移动对象及其间的空间关系。所处理的是随时间变化的空间数据,移动对象是其位置及其他相关信息不断变化的空间对象,这种变化可能是离散型也可能是连续型的。

作为时空对象特征重要组成部分的时间和空间属性是紧密结合的。为了准确地描述客观世界,必须建立能够同时表达时间和空间属性的时空信息系统,将空间数据、时间特征及其他相关属性紧密结合,并解决它们之间的关系。因此,无论是在理论上还是实际应用中,人们都提出了将时间和空间结合起来进行研究并运用到实践中的要求。传统的空间数据库和时态数据库的研究基本上处于相互独立的状态。空间数据库的研究侧重于对数据库中对象的几何模型和空间查询方面的

支持;时态数据库的研究则主要关注对于当前信息和历史信息处理和扩展,即有效时间和事务时间的表达。关于时态数据库的详细讨论见作者所著《时态数据库设计理论》。但实际的应用往往要求同时处理对象的时间和空间属性。目前为止还没有能够有效处理动态空间对象的数据库管理系统出现。因此有必要将时态和空间相结合,当然,时间和空间各自的概念都是应该考虑的。然而,重点应该放在这两个概念相交的领域,只有这样才能体现时空固有的概念,即将时空作为整体来研究一种新的时空数据库(Spatial-Temporal Database, STDB)。

“时空数据库是一个包含了时态数据、空间数据和时空数据,并能同时处理数据对象的时间和空间属性的数据库”是由 Yannis Theodoridi 等定义的。

随着个人移动设备的快速应用,如何实施对移动对象/用户的有效管理成为了时空数据库领域的一个重要分支,提出了专门针对移动对象及其位置进行管理的数据库技术,即移动对象数据库(Moving Objects Databases, MOD)。

移动对象数据库(简称移动数据库)是对移动对象的位置和其他相关信息的表示和管理。移动数据对象和空间数据对象有其根本的不同点是其每一数据项的时点的值是单调的且是递增的,随着时间的不断递增,由于每一个时间点至少对应着一个可能变化的数据,数据量是海量的,同时数据又处于不断地更新和变化中。

更简单地讲,时空数据库就是支持空间对象随时间而发生变化的数据库管理系统。空间对象随时间而发生变化在时空数据库里称为时空变化。

连续的时空变化指空间对象随时间连续变化,在时空数据库中称为运动;离散时空变化指空间对象的时空变化是间隔的。具有时空变化的空间对象称为时空对象。我们把时空对象随时间而变化的空间数据称为时空数据。对时空数据的管理能力是时空数据库与时态数据库、空间数据库的主要区别。

时空数据的管理不是空间数据管理和时态数据管理的简单组合。如果给每个空间数据附上一个时态数据,相当于给空间数据做版本。这样只能获得每次版本时间里的时空对象快照,从而只能表示离散时空变化。而且这种方法也会导致大量的冗余存储。因此真正实现时空数据库管理还要探索新的技术。

时空数据库应用是以支持时空对象以及时空对象间联系为核心的复杂应用。根据所涉及的时空对象类型,时空数据库应用可分为以下三类:

(1) 涉及连续移动的时空对象的应用:在这种应用中,时空对象的位置随时间而连续变化,但形状不变。与交通相关的时空数据库应用可归为这类应用,包括车辆交通管理、轮船航运管理、飞机航线管理以及一些军事应用等。

(2) 涉及离散变化的时空对象的应用:这样的应用所涉及的时空对象有一个空间位置,并且它们的空间属性,如形状和位置都可能随时间离散变化。这类应用包括地籍管理、城市区划管理、地表植被变化检测、疾病传染区域的检测等。

(3) 涉及连续移动并且形状也同时变化的时空对象的应用:这类应用一般与

环境相关,例如风暴监视与预测、森林火灾监控、海上石油污染监控以及种群的迁徙等;另外,生物信息处理应用也属于这类应用,例如一个细胞在运动过程中可能会同时改变形状。

时空数据库中通常管理着两类对象:一类是静态的空间对象,另一类是移动对象。移动对象指空间中一切运动变化的对象,有狭义和广义之分。狭义移动对象是指对象的位置属性随时间不断地发生改变,也就是指运动的对象,如车辆、飞机、轮船、洪水、飓风等。广义的移动对象是指含有不断变化属性的数据对象,不特指对象的运动,也就是说,静止的对象如果含有不断变化的属性,也可以用移动对象的方法进行管理。比如,房间的温度随着时间而发生变化,房间为静止对象,而温度为静止对象不断变化的属性。狭义移动对象按照对象的运动速率可以进一步分为连续变化的移动对象和离散变化的移动对象。连续变化的移动对象,如行驶中的汽车、轮船、飞机等。离散变化的移动对象,如在一座办公楼中办公的人。本书的讨论主要是针对狭义移动对象进行。

一般来说,时空数据库要处理非常大量的复杂对象,这些对象同时具有空间和时态两种特性,尤其是要处理时空移动对象。需要注意的是,时空数据库并不一定是移动数据库,其对象可以是移动的物体,但不是必须具有移动的特性,可以是静态的。因此时空数据库查询可以分为静态对象查询和移动对象轨迹查询两种。

时空数据库必须支持广泛的查询,支持查询的种类越广泛,存取方法就越有效。时空查询主要分为以下几类:

- (1) 简单时态查询:查询某对象在某时刻的状态。
- (2) 时态范围查询:查询在某一时期内某对象发生什么变化。
- (3) 简单时空查询:查询某区域在某时刻的状态。
- (4) 时空范围查询:查询在某时期内某区域发生的情况。

(5) 选择查询:查询在指定时间间隔内(或指定的时间戳)通过特定的空间区域(或空间点)的所有时空对象。它是最基本的时空查询类型,包括时间选择查询、空间选择查询和时空综合选择查询。

- (6) 连接查询:查询在特定时间间隔内,相距对象一定距离的对象。

(7) 时空最近邻查询:查询在指定时间间隔内(或指定的时间戳),离给定点(或区域)最近的一个或多个时空对象。

空间数据库的研究侧重于对数据库中对象的几何模型和空间查询方面的支持,仅能存储空间信息的当前状态;时态数据库的研究则主要关注数据的当前信息和历史信息处理和扩展,即有效时间和事务时间的表达,不能处理空间信息,这些不足限制了对时空对象的有效管理和处理。

通常所指的典型的空间数据类型分为三类:

- (1) 点类型。
- (2) 线类型。
- (3) 区域(面)类型。

空间数据类型又可分为矢量和栅格类型,通常所指的典型的空间数据类型点、线和面属于矢量类型。

空间数据如果从描述空间对象的特征可分成两类,空间特征数据和属性特征数据。空间特征数据表示空间实体的位置或现在所处的地理位置以及拓扑关系和几何特征。几何特征又称为定位特征,一般以坐标加以表示。属性特征数据主要指的是专有属性,也是非定位数据。专有属性是指实体所具有的各种性质,如房屋的结构、高度、层数、使用的主要建筑材料、功能等。专有属性通常以数字、符号、文本和图像等方式表达。属性特征数据一般简称属性数据。

## 1.2 空间数据库概述

### 1.2.1 空间数据结构

空间数据结构是对空间数据元素之间结构关系的描述。空间数据模型存储方式分为两大类,矢量数据和栅格数据。矢量数据又按其是否明确表示地理实体的空间相互关系,分为拓扑型数据和实体型数据两种类型。栅格数据按其压缩编码的方式不同,分游程编码、链式编码和四叉树编码等多种形式。

(1) 矢量数据结构:矢量数据结构通过记录坐标的方式,尽可能地将点、线、面地理实体表现得精确无误。矢量数据结构按其是否明确表示地理实体间空间相关的关系可分为实体型数据结构和拓扑型数据结构两大类。

实体型数据结构是最简单的数据结构。在实体型数据结构中,点用一对 $(x, y)$ 坐标对表示。线用两个或两个以上的有序坐标对来表示,面是由一条或一条以上的首尾相连的线段组成。

拓扑型数据结构由结点文件、线段文件和多边形文件等一系列包含拓扑关系的数据文件组成。

(2) 栅格数据结构:栅格结构是最简单最直接的空间数据结构,它是指将地球表面划分为大小均匀紧密相邻的网格阵列。每个网格作为一个像元或像素由行、列定义,并包含一个代码表示该像素的属性类型或量值,或仅仅包括指向其属性记录的指针。因此,栅格结构是以规则的阵列来表示空间地物或现象分布的数据组织,组织中的每个数据表示地物或现象的非几何属性特征。

### 1.2.2 空间查询

现在空间数据没有标准的空间代数,也没有标准的空间查询语言,加上空间算



子非常依赖于应用程序范围,目前的空间查询操作,大多采用扩展的 SQL 语句。扩展的 SQL 允许使用抽象数据类型表达空间对象以及对象的联合运算。查询的结果通常是满足查询要求的空间数据对象集合。一些常用的基本空间数据查询类型请见作者所著一书《时空数据库查询与推理》。

下面只介绍与本书讨论的相关的查询类型:

(1) 简单最近邻查询:给出一个查询点  $q$  和一个点集  $P$ ,在点集  $P$  中,找出距离查询点  $q$  最近邻  $NNS(q)$ 。

(2)  $k$  最近邻查询:给出一个查询点  $q$  和一个点集  $P$ ,在点集  $P$  中,找出距离查询点  $q$  最近的  $k$  个点。

(3) 反向最近邻查询:给出一个点  $q$  和一个点集  $P$ ,在点集  $P$  中,找出所有以点  $q$  为最近邻的点。

(4) 受约束最近邻查询:带有一定范围约束的最近邻查询。给出一个查询点  $q$  和一个点集  $P$ ,在点集  $P$  中,找出距离查询点  $q$  最近且满足一定限制条件的最近邻  $cNN(q)$ 。

空间点与点的最近邻查询除了上述所列的种类外,与具体应用有关的也不少,不再一一列举。

(5) 连续最近邻查询:给定一条线段  $L$  和一个点集  $P$ ,为线段  $L$  的每一个点在点集  $P$  中,找一个最近邻。

(6) 范围最近邻查询:给定一个范围  $R$  和一个点集  $P$ ,要求为范围  $R$  中的所有点,在点集  $P$  中找出最近邻。

### 1.2.3 空间数据库索引技术

空间数据库的三个基本功能中,其中最重要的就是对空间数据查询的支持和对空间数据分析推理。由于空间数据本身的复杂性,为了提高对空间数据查询的支持和对空间数据分析推理的处理效率,空间数据库必须利用有效的索引机制。与传统的数据库管理系统相比,空间数据库涉及对现实世界大量空间目标的处理。一般的来说,无论哪一类数据库系统均需要一种索引机制帮助它根据数据的空间定位迅速地检索数据对象。

空间数据库索引的提出是由两方面决定的。

(1) 由于计算机的体系结构将存储器分为内存、外存两种,访问这两种存储器一次所花费的时间一般为  $30\sim 40\text{ns}$  和  $8\sim 10\text{ms}$ ,可以看出两者相差十万倍以上,尽管现在有“内存数据库”的说法,但绝大多数数据还是存储在外存磁盘上的,如果对磁盘上数据的位置不加以记录和组织,每查询一个数据项就要扫描整个数据文件,这种访问磁盘的代价就会严重影响系统的效率,因此系统的设计者必须将数据在磁盘上的位置加以记录和组织,通过在内存中的一些计算来取代对磁盘漫无目