

普通高等院校地理信息科学系列教材

# 地理空间数据库原理

(第二版)

崔铁军 等 编著



科学出版社

普通高等院校地理信息科学系列教材

# 地理空间数据库原理

(第二版)

崔铁军 等 编著

天津市品牌专业经费资助



科学出版社

北京

## 内 容 简 介

地理空间数据管理是地理信息系统的核心技术之一。本书全面、系统地论述了地理空间数据库的基本概念、原理与方法。全书共分十三章,内容包括:绪论、数据物理组织、地理空间数据组织、地理空间数据索引、数据模型、地理空间数据模型、地理空间数据查询与操作、地理空间数据库引擎、地理空间数据库体系结构、地理空间数据库管理系统、地理空间数据库设计、地理空间数据库建设、地理空间数据集成。本书内容丰富、组织严谨,原理和方法结合密切,丰富的图表和应用实例便于读者自学。

本书既可作为高等院校地理信息科学专业或相关专业本科生和研究生的教材,也可供从事信息化建设、信息系统开发等有关科研、企事业单位的科技工作者阅读参考。

### 图书在版编目(CIP)数据

地理空间数据库原理/崔铁军等编著. —2版. —北京:科学出版社,2016.11  
普通高等院校地理信息科学系列教材

ISBN 978-7-03-050355-8

I. ①地… II. ①崔… III. ①地理信息系统-数据库系统-高等学校-教材 IV. ①P208 ②TP311.13

中国版本图书馆 CIP 数据核字(2016)第 261156 号

责任编辑:杨 红 程雷星/责任校对:杜子昂

责任印制:张 伟/封面设计:陈 敬

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>

北京教图印刷有限公司 印刷

科学出版社发行 各地新华书店经销

\*

2007年4月第 一 版 开本:787×1092 1/16

2016年11月第 二 版 印张:24

2016年11月第九次印刷 字数:598 000

定价:69.00 元

(如有印装质量问题,我社负责调换)

## 第二版前言

《地理空间数据库原理》出版近十年了，承蒙读者厚爱，出版社已多次印刷。这次再版，主要基于两点考虑：一是近几年地理空间数据库技术发展很快，书的内容有些陈旧，需要更新；二是这次再版书主要是作为地理信息科学本科专业教材，书的内容要与其他教材配套，一些章节需要增删。

第一版中第二章地理空间现象计算机表达内容移到《地理信息科学基础理论》第四章。第一版中第三章基本数据结构，由于现在地理信息科学本科都开设了“数据结构”课程，所以本版删除了相关内容。本版对空间数据索引、时空数据模型、三维数据模型和空间数据引擎等内容进行了大量修改。

在本书再版过程中，解放军信息工程大学郭黎副教授、天津师范大学郭继发副教授、王辉博士认真校阅了全书，并提出了修改意见；在读研究生协助完成了插图绘图和初稿校对等工作。对此，作者向他们表示衷心的感谢。还需要说明的是，本书在编著过程中吸收了大量国内外有关论著的理论和技术成果，书中仅列出了部分参考文献，未公开出版的文献尽量在正文当页下方作了脚注，部分资料可能来自某些网站，但未能够注明其出处，请被引用资料的作者谅解。

值此成书之际，感谢天津师范大学城市与环境科学学院领导和老师的支持；感谢历届博士生、硕士生在地信科学研究方面所作出的不懈努力。

作者

2016年7月1日于天津

# 第一版序

自古以来，地图都是地理空间信息的主要载体和传播工具。随着社会发展和科学技术的进步，人们认知的地理空间的范围越来越大，对地图品种和数量的需求与日俱增，作为地理空间信息载体的地图，其存储、管理与分发成为突出的问题。

对于纸质的模拟地图，通过地图库存储与管理，这种方式一直沿用至今。

目前，以地理空间数据库、计算机地图制图和计算机网络技术为支撑的数字化地图制图已经取代了传统的手工地图制图，并正向以地理空间信息服务为核心的信息化数字化地图制图和地理信息系统转变，地图制图的思想观念、技术手段、产品形式和服务方式等都在发生深刻的变化，地理空间数据的存储与管理也一直是业界关注的问题。

地理空间数据的管理技术经历了许多变化，地理空间数据从由文件管理发展到用数据库管理可以追溯到 20 世纪 60 年代中期，到目前为止，它已从第一代的层次、网状数据库系统，第二代的关系数据库系统发展到第三代的以面向对象为主要特征的新一代数据库系统。与此同时，空间数据仓库技术也备受重视，它作为一种支持决策过程的、面向主题的、集成的、稳定的、不同时间的地理空间数据的集合，是一个多种异构数据源在单个站点以统一的模式组织的存储，其根本目的是服务于决策支持。

地理空间数据库具有地理空间信息科学技术、计算机科学技术等多学科交叉的特点，地理空间数据库的研究必须具备扎实、雄厚的信息科学技术知识和很强的地理空间抽象能力及丰富的工程实践经验。崔铁军教授于 1979~1983 年在中国人民解放军信息工程大学（简称解放军信息工程大学）测绘学院专修计算机地图制图，毕业后留校任教，后在德国联邦国防军大学获工学博士学位，1998 年进解放军信息工程大学测绘学院博士后科研工作站，长期从事地理空间数据库和地理信息系统理论、方法和技术的科研、教学和工程实践，在地理空间数据库方面有较深的造诣和丰富的实践经验，这些为撰写《地理空间数据库原理》一书奠定了坚实的基础。

该书内容丰富，既包含了数据库的基本理论，又扩展到了地理空间信息范畴；既把重点放在地理空间数据库，又扩展到了空间数据仓库；既突出了地理空间数据库的核心——空间数据模型，又把地理空间数据库的设计和建设放在了重要位置；既论述了地理空间数据库的体系结构，又介绍了关系数据库的接口技术和空间数据库引擎，是一部完整地论述地理空间数据库的好书，值得一读。有幸，我作为第一个读者，受益匪浅。

目前，我国有关地理空间数据库方面的书还太少，真诚地期待着更多的青年学者撰写这类著作，共同繁荣地理空间信息科学这块前程似锦的园地。

中国工程院院士 王家耀

2006 年夏

# 第一版前言

地理空间数据库技术是地理信息系统数据组织的核心技术，也是地理科学、测绘科学、计算机科学和信息科学相结合的产物。地理空间数据库技术已经代替传统的文件管理方式，逐步成为地理空间数据管理的主流技术。由于地理空间数据的特殊复杂性，地理空间数据管理在为计算机和信息科学作贡献的同时，也如饥似渴地汲取计算机主流技术的各项最新成果，成为计算机科技领域中应用研究技术内容最丰富的分支之一。

地理空间数据库的主要任务是研究地理空间物体的计算机数据表示方法、数据模型，在计算机内的数据存储结构和建立空间索引方法，以及如何以最小的代价高效地存储和处理空间数据，正确维护空间数据的现实性、一致性和完整性，为用户提供现实性好、准确性高、完备、开放和易用的地理空间数据。地理空间数据库是理论性和实践性很强的学科，理解起来也非常抽象。帮助有兴趣的读者更好地了解地理空间数据库的基本概念和建设方法，是作者编写本书的用心所在。

多年来，撰写一部有关地理空间数据库原理的书是作者的愿望，但现代科技高速发展，使地理空间数据库内容也更新很快，章节很难固定下来，加上人到中年工作繁忙，能专注写书的时间有限，而且由于知识积累与资料准备等，一直未能如愿。

作者长期致力于地理空间数据库理论、技术和方法的研究，工程实践，教学和研究生培养等工作。1987~1992年为地图数据库，特别是1:25万地图数据库技术方案论证做了大量工作，主要负责地图数据库管理系统的结构设计、功能模块划分和部分软件的编写及总体系统调试工作。1998年在博士后工作期间，承担大型地理信息系统研制任务，负责系统的需求分析、总体设计和详细设计等工作。近几年，先后参加和主持了多项国家和军队科研课题，从底层自主设计与开发了基于Oracle关系数据库的空间数据库管理系统，实现了对海量多源空间数据的管理与综合应用，主要功能有空间数据库定义，空间数据的录入、编辑与处理，检索与查询，可视化，备份和恢复等。所取得的研究成果，为本书的撰写奠定了坚实的基础。因此，可以说，本书也是上述工作的结晶。

本书是在作者所讲授地图数据库课程内容的基础上编著而成的。第三章基本数据结构主要参考了韩丽斌教授编著的《地图数据库原理与技术》（1994年，内部教材）；第五章数据模型和第九章地理空间数据库体系结构的部分内容参考了毋河海教授编著的《地图数据库系统》。其他章节也参阅、吸收了国内外有关论著的理论和技术成果，书末仅列出了部分参考文献，未公开出版的文献在正文当页下方作脚注，这里向所有参考文献作者致谢！

在本书撰写过程中，研究生邹方磊、和万礼和张威等协助完成了初稿校对等工作。陈应东副教授和郭健副教授等提出了宝贵的意见。在此，向他们表示衷心的感谢！

值此成书之际，感谢导师王家耀院士、Kurt Brunner教授、刘家豪教授、刘光运教授、韩丽斌教授和已故杨启和教授的培养教育；感谢解放军信息工程大学测绘学院训练部和地图学与地理信息工程系的几任领导的一贯支持；感谢地图数据库教研组董延春、郭黎、姚

慧敏、吴正升和张斌等教师和历届博士生、硕士生在地空间数据库研究方面所作出的不懈努力。

本书内容横跨多个学科，加之作者水平所限，书中定有不少疏漏之处，恳请读者与专家们批评指正。

作 者

2006年7月于郑州

# 目 录

第二版前言

第一版序

第一版前言

第一章 绪论 .....	1
1.1 地理空间数据库起源 .....	1
1.2 地理现象数据表达 .....	4
1.3 地理空间数据库概述 .....	9
1.4 地理空间数据管理演变过程 .....	13
1.5 地理空间数据库研究内容 .....	20
1.6 地理空间数据库与其他课程关系 .....	23
1.7 本书主要内容及阅读指南 .....	25
第二章 数据物理组织 .....	27
2.1 物理组织的基本概念 .....	27
2.2 数据文件组织方法 .....	38
2.3 索引文件组织方法 .....	44
第三章 地理空间数据组织 .....	56
3.1 矢量数据组织 .....	56
3.2 栅格数据组织 .....	63
3.3 数字高程模型数据组织 .....	69
3.4 地物三维模型数据组织 .....	72
3.5 空间数据交换 .....	87
第四章 地理空间数据索引 .....	93
4.1 空间索引概述 .....	93
4.2 网格索引 .....	100
4.3 二叉树空间索引 .....	105
4.4 R 树索引 .....	110
4.5 四叉树索引 .....	113
4.6 空间填充曲线 .....	121
第五章 数据模型 .....	124
5.1 实体模型 .....	124
5.2 数据模型 .....	125
5.3 数据模型比较与转换 .....	141
5.4 面向对象数据模型 .....	145



<b>第六章 地理空间数据模型</b> .....	153
6.1 地理空间数据模型概述 .....	153
6.2 时空数据模型 .....	168
6.3 三维地理数据模型 .....	175
6.4 常用国内外数据模型 .....	181
<b>第七章 地理空间数据查询与操作</b> .....	189
7.1 全关系化地理空间数据存储 .....	189
7.2 地理属性数据查询与操作 .....	194
7.3 地理空间几何数据查询与操作 .....	200
7.4 栅格数据查询与操作 .....	213
<b>第八章 地理空间数据库引擎</b> .....	217
8.1 地理空间数据引擎概述 .....	217
8.2 关系数据库接口技术 .....	225
8.3 国内外 SED 技术分析 .....	237
<b>第九章 地理空间数据库体系结构</b> .....	244
9.1 地理空间数据库系统 .....	244
9.2 数据库系统体系结构 .....	256
9.3 空间数据库系统体系结构 .....	266
9.4 分布式空间数据库系统 .....	272
<b>第十章 地理空间数据库管理系统</b> .....	280
10.1 GDBMS 概述 .....	280
10.2 图形编辑操作 .....	286
10.3 地理空间数据维护 .....	299
<b>第十一章 地理空间数据库设计</b> .....	306
11.1 地理空间数据库设计概述 .....	306
11.2 地理信息建模方法 .....	310
11.3 地理空间数据库需求分析 .....	317
11.4 地理空间数据库模型设计 .....	319
<b>第十二章 地理空间数据库建设</b> .....	327
12.1 地理空间数据库建设概述 .....	327
12.2 地理矢量数据获取 .....	333
12.3 地理空间数据质量与检验 .....	339
12.4 国家基础地理空间数据库介绍 .....	348
<b>第十三章 地理空间数据集成</b> .....	355
13.1 地理空间数据集成概述 .....	355
13.2 多源地理空间数据集成方法 .....	361
13.3 地理空间数据元数据 .....	366
<b>主要参考文献</b> .....	373

# 第一章 绪 论

地理空间数据管理是地理信息系统的核心技术之一，也是地理信息技术与计算机数据库技术相结合的产物。地理空间数据管理在计算机文件系统( file system )和数据库系统( database system, DBS )的基础上，从基于文件存储、内存更新的单机、版本式管理模式逐步发展到基于数据库管理系统的多用户实时共享网络模式，成为地理信息系统的主流技术。由于地理空间数据的特殊性，地理空间数据管理在吸取计算机主流技术的各项最新成果的同时，也在为计算机和信息科学做贡献，成为计算机科技领域中应用研究技术内容最丰富的分支之一。

## 1.1 地理空间数据库起源

地理信息是描述地表形态及其所附的自然、人文地物特征与属性( attribute )的总称。地理信息是人们认知世界、利用自然不可缺少的媒介，也是经济社会发展的基础性、战略性资源。人类认识、利用乃至依赖地理信息和知识有着几千年的历史，最直接的证明便是地图。地图的出现甚至要早于文字。地图是运用一定的数学法则与地图语言，经过制图综合，将客观世界表现在平面上，实质上是公式化、符号化、抽象化地再现客观世界。通过地图，人类对自身生存的环境有了完整的认识。借助于地图这一简单却又有效的工具，人们可以认识未知的客观世界。地图传递信息和载负信息的能力是无法用其他手段替代的。因此，无论过去、现在还是将来，人们仍将大量使用地图，甚至依赖于地图。

20 世纪计算机的产生和发展，冲击着社会的各个领域，给许多行业带来了巨大的变化和深远的影响。计算机技术在信息处理领域的巨大功能和绝对优势，标志着社会进入了信息时代。古老的地图学在这个变革时期也正发生着巨大的变化，新技术的引用，不仅改变着传统的地图制作技术，还促使地图学理论与方法的研究不断深入。20 世纪 60 年代计算机应用于地图学产生了计算机地图制图技术，人们用计算机表示地图要素( feature )及其相互联系( relationship )，将连续的以模拟方式存在于纸质地图的空间物体离散化，以便计算机能够识别、存储和处理。

早期的计算机制图( 地图制图自动化 )只是把计算机作为工具来完成地图制图的任务，把人们从繁重的手工地图制图劳动中解脱出来，并由此带来了巨大的经济和社会效益。国家、军事部门和企业根据各自对地图数据的需要，投入了大量的人力、物力进行各种比例尺的地图数字化，产生了大量的地图数据。这些数据成为国家和军队的重要资源财富。与其他数据相比，地图数据特殊的数学基础、非结构化数据结构和动态变化的时间特征，给数据获取、处理和存储带来很大难度，如何妥善保存和科学管理这些地图数据是人们长期以来十分关注的课题( 毋河海，1990 )。伴随着计算机数据组织存储技术的发展，地图数据的维护、更新和管理经历了从低级向高级的发展过程，最初采用文件系统的形式，后来逐步发展为地图数据库系统( map database system, MDBS )。该系统由地图数据、地图数据管理系统( database management system, DBMS )、计算机硬件设备和地图数据库管理人员等组成。

地图数据的主要来源是普通地图,反之,生产地图也是早期地图数据库建设与发展的主要驱动力。因此,地图数据有以下几个特点。

(1) 地图比例尺影响。地图数据是某一特定比例尺的地图经数字化而产生的。地理物体表示的详细程度,不可避免地受地图综合的影响。经过了人为制图综合,地理物体的几何精度(形状)和质量特征已经不是现实世界中的真实反映,只能是现实世界的近似表达。为了满足地图应用的需要,不同比例尺地图建立不同地理数据库,如1:5万数据库、1:25万数据库和1:100万数据库等。

(2) 强调数据可视化,忽略了实体(entity)的空间关系。地图数据主要是为地图生产服务的,强调数据的可视化特征。主要采用“图形表现属性”的方式,地图上地理物体的数量特征和质量特征用大量的辅助符号表示,包括线型、粗细、颜色、纹理、文字注记、大小等数十种。地图数据是以相应的图式、规范为标准的,依然保留着地图的各项特征。数据不表示各种地理现象之间的空间位置关系,如道路两旁的植被或农田、与之相邻的居民地等,各种地理现象之间的关系是通过读图者的形象思维从地图上获取的。地理物体如道路、居民地和河流在空间关系上是相互联系的有机整体,但在地图数据表示中是相互孤立的。因此,地图数据不强调实体的关系表示。

(3) 按地图印刷色彩分层管理。为满足地图印刷的需求,依据地图制图覆盖理论,对地图数据按色彩分层管理,不是按照地理物体的自然属性进行分类分级。这种分层不仅割裂了地理物体之间的有机联系,还导致了同一个地物在不同层内重复存储,如河流两岸的加固陡坎隐含着河流的水涯线信息,道路与绿化带平行接壤使道路边沿线隐含着绿化带的边沿,河流、道路和铁路等线状地物可能隐含着区划界限。

(4) 地图图幅限制了数据范围。受印刷机械、纸张和制图设备的限制,传统的地图用图幅限制地图的大小,地图数据用图幅来组织和管理。地图图幅割裂了大区域地理物体的完整性和连续性,例如,一条境界线因为地图的分幅而断作几条记录存储在不同的图幅内。

随着信息科学技术和地图数据应用的深入,地图数据已不再局限于地图生产。地图数据与其他地理信息结合产生各种地理数据,包括资源、环境、经济和社会等领域的一切带有地理坐标的数据,研究解决各种地理问题。地理数据与地图数据相比,这种数据主要通过属性数据描述地理实体的定性特征,用数字表示空间实体的数量特征、质量特征和时间特征。

在数据获取手段上,不再局限于地图的数字化,获取空间物体信息的手段越来越多样化,特别是随着传感器技术、航空和航天平台技术、数据通信技术的发展,现代遥感技术已经进入一个能够动态、快速、准确、多手段提供多种对地观测数据的新阶段。新型传感器不断出现,已从过去的单一传感器发展到现在的多种类型的传感器,并能在不同的航天、航空遥感平台上获得不同空间分辨率、时间分辨率和光谱分辨率的遥感影像。遥感影像的空间分辨率已达到米级;光谱分辨率已达到纳米级,波段数已增加到数十甚至数百个;回归周期可达几天甚至十几小时;微波遥感已逐渐采用多极化技术、多波段技术及多种工作模式。全球定位系统(global positioning system, GPS)、惯性导航系统(inertial navigation system, INS)和激光雷达等高技术系统相结合的智能型实时地理信息获取系统步入了实用阶段,为地理数据的实时更新提供一个实用、简便、低廉的工具。

在地理信息表示方面,广大科学工作者开始思索如何利用它来反映自然和社会现象的分

布、组合、联系及其时空发展和变化,研究在计算机存储介质上如何科学、真实地描述、表达和模拟现实世界中地理实体或现象、相互关系及分布特征。初期的地图数据仅仅把各种空间实体简单地抽象成点、线和面,这远远不能满足实际需要,要想进一步拓宽应用前景,必须进一步研究它们之间的关系(空间关系)。空间关系通过一定的数据结构来描述与表达具有一定位置、属性和形态的空间实体之间的相互关系。当人们用数字形式描述空间物体,并使系统具有特殊的空间查询、空间分析等功能时,就必须把空间关系映射成适合计算机处理的数据结构,这时必须考虑数据的表示方法。

数据组织上,为了满足地理分析需求,在人们认识世界和改造世界的一定区域内(即现实世界地理空间)不管逻辑上还是物理上均为连续的整体,不受传统图幅划分的限制组织数据。

从空间尺度上讲,地物在地理空间只有唯一的地理数据表示,空间物体本身没有比例尺的含义,应尽可能详细、真实地描述物体形状、几何精度和属性。但人们对地理环境的认识往往需要一个从总体到局部、从局部到总体反复认识的过程。为了满足人们对地理空间这种认识需求,必须考虑空间物体的多尺度性,以满足不同的社会部门或学科领域的人群对空间信息的选择需求。

综上所述,从数据内容、获取手段、表示方法和数据组织上看这些数据已经超出了地图数据的表示范畴,为了与地图数据区分人们称之为地理信息数据(geo-information data)。地理信息数据的获取、处理、管理和分析及其在地学领域的应用促使了地理信息系统(geographic information system, GIS)的产生和发展。GIS的本质是以测绘技术获取地理信息为基础,以数据库储存和管理地理数据,以可视化为地理信息表达的主要手段,运用不同地理空间分析方法满足应用需求。所以,GIS的核心是地理信息数据库(geographic information database, GIDB)。它是在一定的地域内,将地理空间信息和一些与该地域地理信息相关的属性信息结合起来,实现对地理几何特征和属性信息的采集、更新和综合管理。

地图数据和地理信息数据都是带有地理坐标的数据,是地理空间信息两种不同的表示方法,地图数据强调数据可视化,采用“图形表现属性”的方式,忽略了实体的空间关系,而地理信息数据主要通过属性数据描述地理实体的数量和质量特征。地图数据和地理信息数据所具有的共同特征就是地理空间坐标,统称为地理空间数据(geospatial data)。地理空间数据库(geospatial database, GDB)的主要任务是研究地理空间物体的数据表示、数据模型(data model)及在计算机内数据的存储结构和空间索引的建立方法,如何以最小的代价高效巧妙地存储和处理空间数据,正确维护空间数据的现实性,为用户提供现实性好、准确性高、完备、开放和易用的地理空间数据。

地理空间数据库系统(geospatial database system, GDBS)的核心软件是空间数据库管理系统(geospatial database management system, GDBMS)。它是为了满足日益发展的空间数据管理的需要,在文件的基础上发展起来的一种空间数据管理技术。它按一定的方式组织和存储管理地理空间数据,具有较高的程序和数据独立性,能以较少的重复为多个用户或应用程序提供数据服务。

将计算机技术与数据库技术应用于地理空间数据的管理,需要解决地理空间数据的模型和索引等一系列复杂的问题,研究解决这些问题的理论方法推动了地理空间数据库技术的产生和发展。

## 1.2 地理现象数据表达

数据作为信息的载体,是用符号记录下来的、可识别的信息。从一般意义上来说,数据是描述现实世界中各种具体事务或抽象概念的、可存储并具有明确意义的信息。地理空间数据是描述地球表层(有一定厚度)一定范围内的地理事物及其关系的数据,是指用来表示空间实体的位置、形状、大小及其分布特征等方面信息的数据。

### 1. 地理空间

关于“空间”(spatial)的概念不同的学科有不同的解释。从物理学的角度看,空间是指宇宙在三个互相垂直的方向上所具有的广延性。从天文学的角度看,空间是指时空连续体系的一部分。地理学是研究地球表层空间分布规律的科学,因此地理学的空间是一个定义在地球表层空间实体集上的关系。在空间实体之间有无数种关系,物理距离只是这些关系中的一种度量;定义一种关系就自然定义了一种空间,而这个空间又是和几何关系联系在一起的,并且,几何关系是所有这些关系中的基础关系。也许正因为如此,今天大多数的学者都强调空间位置和拓扑关系。也就是说,地理空间(geospatial)是一个相对空间,是一个空间实体组合排列集(这些空间实体具有精确的空间位置),强调宏观的空间分布和空间实体间的相关关系(关系是以各单个地理空间实体为联结的结点或载体)。地理空间若想精确定位于地球上,还必须承认它有欧氏空间基础,有相对于地球坐标系的绝对位置。这样,通过地理空间和欧氏空间的统一,从而将地理现象的宏观特性和空间位置的精确特征紧密有机地联系在一起。其中,宏观特性主要体现在地理对象之间的拓扑关系与非拓扑关系(通过数据模型体现)上,其载体则是具有精确位置、起着联结结点作用的那些单个地理空间对象(通过单对象的数据结构体现)。

依附地理空间存在着各种事物或现象,它们可能是物质的,也可能是非物质的,这些事物和现象的一个典型特征是,与一定的地理空间位置有关,都具有一定的几何形态,这些事物或现象称为地理空间物体(geospatial object)。在地理空间中,物体不仅反映事物和现象的地理本质内涵,还反映它们在地理空间中的位置、分布状况及它们之间的相互关系。地理空间实体(geospatial entity)是根据分析应用的需要对空间物体进行的抽象表示。本书中,常用地理空间物体和地理空间实体区分客观存在的物体和数字表示的物体,前者如长江,后者如表示长江的曲线(坐标串数据)。

地理空间的数学描述可以表示为:设  $E_1, E_2, E_3, \dots, E_n$  为  $n$  个不同类的地理空间实体;  $R$  表示地理空间实体值的相互联系和相互制约关系;  $\Omega = \{E_1, E_2, E_3, \dots, E_n\}$  表示地理空间中各个组成部分(实体)的集合,那么地理空间可以表示为  $S = \{\Omega, R\}$  (王家耀, 2000)。



图 1.1 地理实体的描述

### 2. 地理空间实体

人们在认识自然和改造自然的活动中,长期以来用语言、文字与地图等手段表示描述自然现象、人文社会文化的发生与演变的空间位置、形状、大小范围及其分布特征等方面的地理信息(图 1.1)。空间实体是指具有确定的位置和形态特征并具有地理意义的地理空间

的物体。确定的位置和形态特征是指至少在给定的时刻,空间实体具有确定的形态,但是“确定的形态”并不意味着空间实体必须是可见的、可触及的,也可以是不可见的。地理意义是指在特定的地学应用环境中,被确认为有分析的必要。河流、道路、城市是看得见摸得着的空间实体,而境界、航线等则是不可见的空间实体。

地理学是研究地球表面自然和人文社会现象及发展规律的科学。人们借助于外感官,探讨地球表面众多现象、过程、特征,以及人类和自然环境的相互关系在空间及时间上的分布。地理现实世界是复杂多样的,要正确地认识、掌握与应用这种广泛而复杂的现象,需要进行去粗取精、去伪存真的加工,这就要求对地理环境进行科学的认识。对于复杂对象的认识是一个从感性认识到理性认识的抽象过程。对于同一客观世界,不同社会部门或学科领域的人群,往往在所关心的问题、研究的对象等方面存在着差异,这就会产生不同的环境映象。

### 3. 地理空间实体的抽象表达

人类自从有了语言就学会了抽象表达。人类借助于外感官了解外面的地理现象,在认识过程中,从感性认识上升到理性认识,把所感知事物的共同本质特点抽象出来,加以概括,就成为概念。概念层次的世界充满了复杂的形状、样式、细节。人类在表达概念的过程中形成语言,包括自然语言、文字和图形。用图形表示地理世界就有了地图。地图用简单的、抽象的地图符号描述复杂的地理现象。地图在抽象概括表达过程中用两种观点描述现实世界。

一是场的观点。地理现象在空间上是连续地充满地球表层空间的。地球表面的任何一点都处于三维空间,如果包含时间,是四维空间离散世界,如大气污染、大气降水、地表温度、土壤湿度及空气与水的流动速度和方向等。基于场的思想是把地理空间的事物和现象作为连续的变量来看待,借助物理学中场的概念表示一类具有共同属性值的地理实体或者地理目标的集合,根据应用的不同,场可以表现为二维或三维。一个二维场就是在二维空间中任何已知的点上,都有一个表现这一现象的值;而一个三维场就是在三维空间中对于任何位置来说都有一个值。一些现象,如空气污染物在空间中本质上讲是三维的。地理空间上任意给定的空间位置都对应一个唯一的属性值。根据这种属性分布的表示方法,基于场模型可分为图斑模型、等值线模型和选样模型。

图斑模型是将一个地理空间划分成一些简单的连通域,每个区域用一个简单的数学函数表示一种主要属性的变化。根据表示地理现象的不同,图斑模型可以对应不同类型的属性函数。比较简单情况,每个区域中的属性函数值保持一个常数。图斑模型常常被用于描述土壤类型、土地利用现状、植被及生物的空间分布。除了单一属性值,还有多属性值的情况。

等值线模型经常被视为由一系列等值线组成,一条等值线就是地面上所有具有相同属性值的点的有序集合。用一组等值线将地理空间划分成一些区域,每个区域中属性值的变化是相邻两条等值线的连续插值。等值线模型常表示等高线、等温线、大气压、地下水文线等。

选样模型是以有限的抽样数据表达地球表面无限的连续现象,地理现象在地理空间上任何一点的属性值是通过有限个点的属性值插值计算的。按采样点分为无规律的离散点和规则格网点。

二是对象观点。地球表层空间被散布的各种对象所填充,对象之间具有明确的边界,每一个对象都有一系列的属性。基于对象的思想是采用面向实体的构模方法,将地球表面的现实世界抽象为点、线、面、体的基本单元,每个基本单元表示为一个实体对象。每个实体对象的几何位置和形态用矢量坐标表示。每个实体对象均赋以唯一的标识来表示,并用属性表表示实体对象的质量和数量特征。

场观点和实体对象观点并不互相排斥，有些应用可以共存。例如，对于地面起伏的描述，既可采用场模型描述，如离散点、断面线、不规则三角形和规则三角形，也可以采用等高线对象表示。基于场的模型和基于实体的模型各有长处，应该恰当地综合运用这两种方法来建模。

#### 4. 空间实体的地图表示

人类很早就学会了用地图图形科学地、抽象概括地反映自然界和人类社会各种现象的空间分布、组合、相互联系及其随时间动态的变化和发展。地图是空间信息的载体，是客观地理世界的一种最有效的表示形式，是人们认识所生存的空间世界环境的最有力工具。地图对空间信息的反映，是通过对现实世界的科学抽象和概括，依据一定数学法则，运用地图语言——地图符号实现的。地图不是客观对象的完全再现，而是通过地图制图工作者“过滤”、概括和抽象出来的模型。对地物进行取舍、图形化简、数量和质量概括、地物协调，以及夸大表示等处理，是地图学者对地物进行深刻理解、综合分析并抽象概括等一系列复杂科学思维和创造性劳动的结果。地图是地表空间关系和空间形式的视觉图解表象，是客观世界的抽象符号描述，是客观世界的符号模型。

#### 5. 地理空间实体的数据描述

随着计算机技术的应用，为了使计算机能够识别、存储和处理地理实体，人们不得不将以连续的模拟方式存在于地理空间的物体离散化，用数字（数据）描述地球表面地理信息。空间物体离散化的基本任务就是将以图形模拟的空间物体表示成计算机能够接受的数字形式（数字化）。地理空间数据是指用来表示地理空间实体的位置、形状、大小及其分布特征诸多方面信息的数据，可以用来描述来自现实世界的目标，具有定位、定性、时间和空间关系等特性。

##### 1) 地理空间实体的数据抽象

数据是对事物的描述，可以以文字、数字、图形、影像、声音等多种形式存在，但数据不是事物本身。数据是以诸如人工统计、仪器测量、社会调查等多种方式获取的，这就必然导致甚至可能是必然存在各种误差（error），如人为差错、仪器的系统误差等。因此，数据只能从有限的方面描述事物，而不可能也没有必要全面、详尽、保真地复制事物本身。数据是被描述事物的另一种存在方式。这种转换经历了三个领域：现实世界、观念世界和数据世界，如图 1.2 所示。

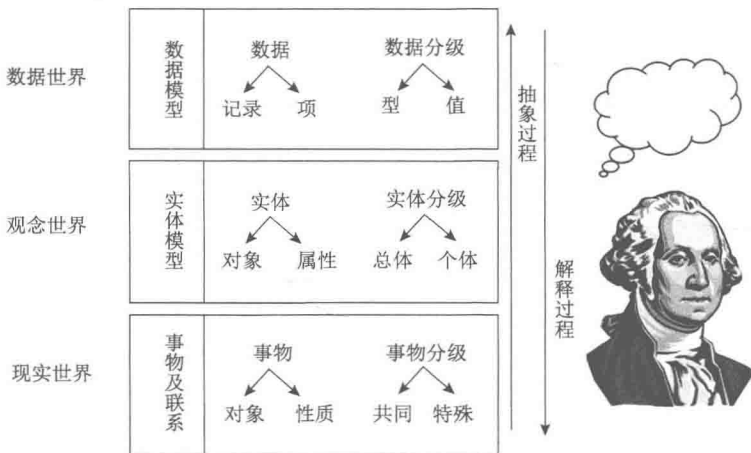


图 1.2 空间实体的数据抽象

(1) 现实世界是存在于人们头脑之外的客观世界,事物及其相互联系就处在这个世界之中。事物可分成“对象”与“性质”两大类,又分为“特殊事物”与“共同事物”两个重要级别。

(2) 观念世界是现实世界在人们头脑中的反映。客观事物在观念世界中称为实体,反映事物联系的是实体模型。

(3) 数据世界是观念世界中信息的数据化,现实世界中的事物及联系在这里用数据模型描述。

如前所述,人们使用抽象方式描述地理现实世界,只能描述对象的局部。同样,用数据描述地理世界也建立在抽象表达基础上,不能也没有必要全面、详尽、保真地复制地理现象本身。用数据世界描述地理世界有两种形式:①基于场的观点,表达地理连续现象的栅格数据;②基于对象观点,表达地理离散现象的矢量数据(图 1.3)。

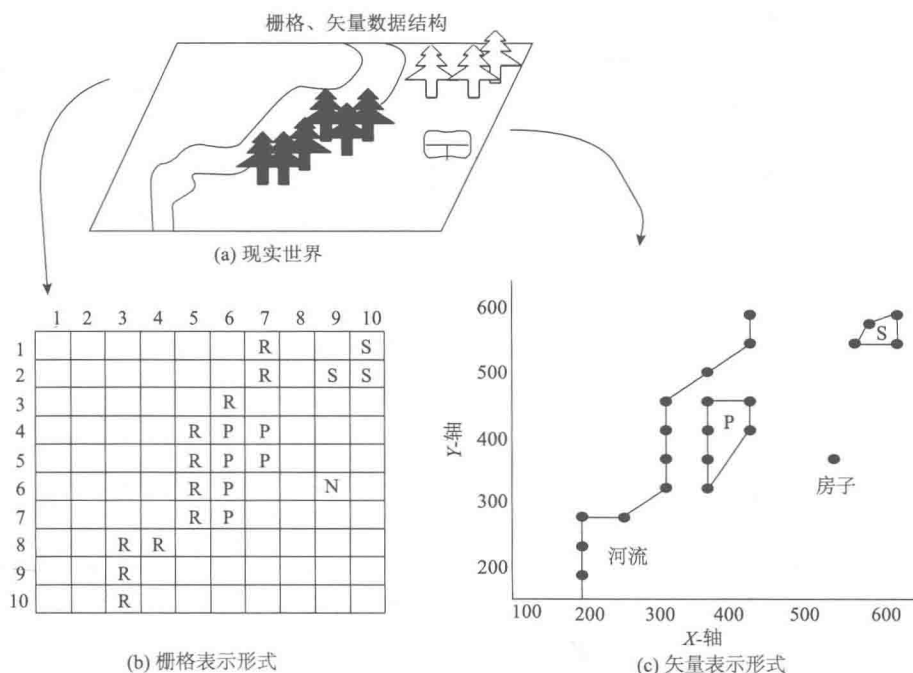


图 1.3 地理实体的数据描述

## 2) 矢量数据

矢量数据就是在直角坐标系中,用  $X$ 、 $Y$  坐标表示地图图形或地理实体的位置和形状的数据。通过记录实体坐标及其关系,尽可能精确地表现点、线、多边形等地理实体,坐标空间设为连续,允许任意位置、长度和面积的精确定义。矢量数据结构是利用欧几里得几何学中的点、线、面及其组合体来表示地理实体空间分布的一种数据组织方式。这种数据组织方式能最好地逼近地理实体的空间分布特征,数据精度高,数据存储的冗余度低,便于进行地理实体的网络分析。

在现实世界中许多地理事物和现象可以构成网络,如铁路、公路、通信线路、管线、自然界中的物质流、能量流和信息流等,都可以表示为相应的点之间的连线,由此构成现实世界中多种多样的地理网络。按照基于对象的观点,网络是由点对象和线对象之间的拓扑空间关系构成的。对于每一个具体的空间实体对象都直接赋有位置和属性信息,以及空间实体之间的关系说明。



### 3) 栅格数据

栅格数据就是按栅格阵列单元的行和列排列的有不同“值”的数据集。栅格结构是用大小相等分布均匀、紧密相连的像元(网格单元)阵列来表示空间地物或现象分布的数据组织,是最简单、最直观的空间数据结构,它将地球表面划分为大小、均匀、紧密相邻的网格阵列。每一个单元(像素)的位置由它的行列号定义,所表示的实体位置隐含在栅格行列位置中,数据组织中的每个数据表示地物或现象的属性或指向其属性的指针。点实体由一个栅格像元来表示;线实体由一定方向上连接成串的相邻栅格像元来表示;面实体(区域)由具有相同属性的相邻栅格像元的块集合来表示。

## 6. 地理空间数据特征

地理空间数据代表了现实世界地理实体或现象在信息世界的映射,是地理空间抽象的数字描述和离散表达。地理空间数据是描述地球表面一定范围(地理圈、地理空间)内地理事物的(地理实体)位置、形态、数量、质量、分布特征、相互关系和变化规律的数据。地理空间数据作为数据的一类,除具有空间特征、属性特征和时间特征外,还具备抽样性、时序性、详细性与概括性、专题性与选择性、多态性、不确定性(uncertainty)、可靠性(reliability)与完备性等特点。这些特点构成了地理空间数据与其他数据的差别。

### 1) 空间特征

在事务数据库中,数据记录一般是结构化的,即每一个记录有相同的结构和固定的长度,记录中每个字段表达的只能是原子数据(不可再分的数据),内部无结构,不允许嵌套记录。而这种结构化不能满足空间数据表示的要求。这是因为,地理实体都具有空间坐标,空间坐标不仅指示了地理实体的位置、大小和形状,还记录了拓扑信息,用以表达地理实体之间的关联、邻接、包含等空间关系。拓扑数据一方面方便了空间数据的空间查询和空间分析;另一方面也增加了空间数据一致性和完整性维护的复杂性。因此,空间数据的组织和管理不同于一般的事务性数据,要根据它的空间分布特征,建立相应的空间索引,从而实现空间数据高效快速的存储和提取。

### 2) 属性特征

属性特征,即通常所说的非几何属性或简称属性,是与地理实体相联系的地理变量或地理意义。属性分为定性和定量两种,前者包括名称、类型、特性等,后者包括数量和等级;定性描述属性如土壤种类、行政区划等,定量属性如面积、长度、土地等级、人口数量等。属性一般是经过抽象的概念,通过分类、命名、量算、统计得到的,任何地理实体至少有一个属性。

为了唯一识别地理实体,每一个地理实体均分配一个分类编码。这种编码可能是按国家标准、行业标准或地区标准进行的,也可能是全球性的。通过分类编码将空间数据和属性数据关联起来。

### 3) 空间关系特征

空间数据除了要描述地理实体的空间坐标和空间分布之外,还要描述地理实体之间的空间关系及实体组成元素之间的拓扑关系(如点与线、线与面等)。这给空间数据的一致性和完整性维护增加了困难。特别是某些几何对象并不直接记录其坐标信息,如面状目标仅记录组成它的弧段标识,因而在查找、显示和分析时均要操纵和检索多个数据文件。

### 4) 时间特征

时间特征用以描述事物或现象随时间的变化。这种变化表现为三种可能的形式:一是属