

刘振东

GIS 应用水平考试（一级）

串讲资料



全国信息化工程师
GIS应用水平

2008年4月

GIS 应用水平考试（一级）考前串讲

1、关于考试

考试内容

考试的主要内容为计算机基础知识、地图学、遥感学、测绘学、GIS 的原理与方法、GIS 在信息化建设中的应用及案，以及 GIS 最新的行业知识与新闻。

考试题型

单选题、多选题、判断题、简答题、综合题

考试方法

全国统一时间 2008-4-26 上午，闭卷笔试。考试时间长度：150 分钟。

试卷满分 150 分，90 分以上为合格，120 分以上为优秀。合格即发证。

2、关于考题

科目比重

考试涉及的科目	所占百分比
计算机基础	20%
遥感技术	10%
测绘技术	10%
地图学	10%
地理信息系统原理	50%

题量分布

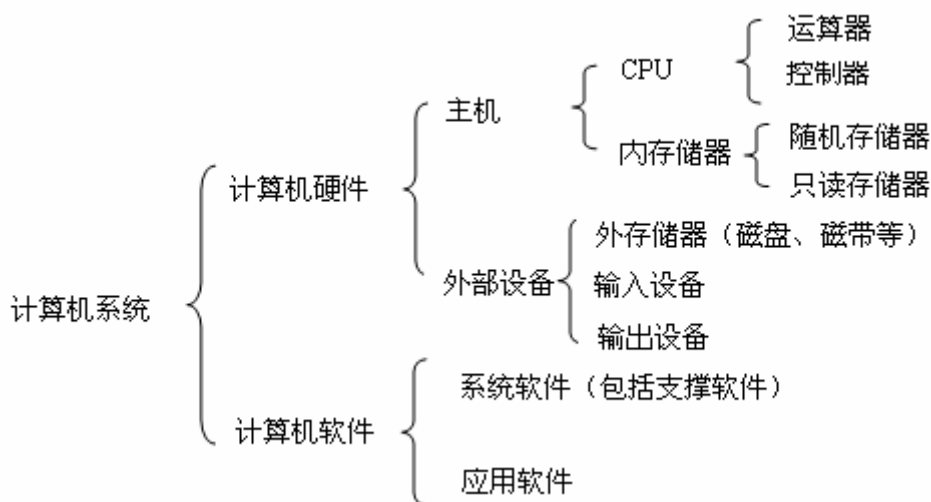
	题目数	单题分值	总分值
单选题	60	1	60
多选题	10	2	20
判断题	10	1	10
简答题	6	5	30
论述题	3	10	30
合 计	89		150

第一部分 计算机基础

第一章 计算机硬件概述

1.1 计算机的组成

微型计算机系统(简称微机系统)与传统的计算机系统一样,也是由硬件系统和软件系统两大部分组成。



计算机硬件系统由控制器、运算器、存储器、输入设备、输出设备五大基本部件构成。

软件系统由系统软件和应用软件两大部分组成。系统软件包括操作系统、语言处理程序、数据库管理系统、网络通信管理程序等部分。应用软件包括的面非常广,它包括用户利用系统软件提供的系统功能、工具软件和其它实用软件开发的各种应用软件。

1.2 计算机软件概述

从第一台计算机上第一个程序出现到现在,计算机软件已经发展成为一个庞大的系统。从应用的观点看,软件可以分为三类,即系统软件、支撑软件和应用软件。

1. 系统软件

操作系统是系统软件的典型代表。

当前使用比较多的操作系统有 UNIX、Linux, MS-DOS 和 Windows 操作系统。

UNIX 是 1969 年由美国电话电报公司(AT&T)的贝尔实验室推出的一种多用户操作系统,它可运行在不同厂商制造的各种型号的微型机或大型机上。MS—DOS 是美国微软公司开发的一种用于个人计算机的操作系统, MVS 是运行在 IBM 大中型计算机上的一个操作系统。Windows 是美国微软公司推出的具有多窗口和图形化界面的系统。

2. 支撑软件

支撑软件是支持其它软件的编制和维护的软件。例如测试工具、编辑工具、项目管理工具、数据流图编辑器、语言转换工具、界面生成工具等。

3. 应用软件

应用软件是为计算机在特定领域中的应用而开发的专用软件。例如各种管理信息系统、飞机订票系统、地理信息系统、CAD 系统等等。

1.3 存储器的基本概念

1.3.1 内存储器

内存又称为主存，分为随机存储 RAM 和只读存储器 ROM 两种，速度比外存快。RAM 是一种读写存储器，其内容可以随时根据需要读出，也可以随时重新写入新的信息。这种存储器可以分为静态 RAM 和动态 RAM 两种。静态 RAM 的特点是只要存储单元上加有工作电压，它上面存储的信息就会保持。动态 RAM 由于利用 MOS 管极间电容保存信息的，因此随着电容的漏电，信息会逐渐丢失，为了补偿信息的丢失，要每隔一定时间对存储单元的信息进行刷新，不论是静态 RAM 还是动态 RAM，当电源电压去掉时，RAM 中保存的信息都将会丢失。RAM 在微机中主要用来存放正在执行的程序和临时数据。由于静态存储器成本较高，通常在存储器较小的存储系统中采用，以省去刷新电路。在存储量较大的存储系统中宜用动态存储器，以降低成本。

ROM 是一种内容只能读出而不能写入和修改的存储器，其存储的信息是在制作该存储器时就被写入的。计算机断电后，ROM 中的信息不会丢失，ROM 常用来存放一些固定的程序、数据和系统软件等，如检测程序、ROMBIOS 等。只读存储器除了 ROM 外，还有 PROM、EPROM 等类型。PROM 是可编程只读存储器，但只可编写一次。与 PROM 器件相比，EPROM 器件是可以反复多次擦除原来写入的内容，重新写入新的内容的只读存储器。

由于 CPU 比内存速度快，目前，在计算机中还普遍采用了一种比主存储器存取速度更快的超高速缓冲存储器，即 Cache，置于 CPU 与主存之间，以满足 CPU 对内存高速访问的要求。有了 Cache 以后，CPU 每次读操作都先查找 Cache，如果找到，可以直接从 Cache 中高速读出；如果不在 Cache 中再由主存中读出。

衡量内存的常用指标有容量与速度。目前微机内存容量主要有 512MB、1GB、2GB 等。微机内存的速度是指读或写一次内存所需的时间，数量级以纳秒-ns 衡量。

1.3.2 外存储器

计算机系统都要配置外存储器。外存储器又称为辅助存储器，它的容量一般都比较大，而且大部分可以移动，便于不同计算机之间进行信息交流。

在微型计算机中，常用的外存有磁盘、光盘和磁带，磁盘又可以分为硬盘和软盘。

衡量硬盘的常用指标有容量、转速、硬盘自带 Cache（高速缓存）的容量等。容量越大，存储信息量越多；转速越高，存取信息速度越快；Cache 大，计算机整体速度越快。

目前微机常用硬盘容量在 10GB 以上，普通硬盘转速 5400 转，高速硬盘 7200 转，普通硬盘 256KB 的 Cache，而高速硬盘有 2M Cache。

光盘的存储介质不同于磁盘，它属于另一类存储器。由于光盘的容量大、存取速度快、不易受干扰等特点，光盘的应用越来越广泛。光盘根据其制造材料和记录信息的方式的不同一般分为三类：只读光盘、一次性写入光盘和可擦写光盘。

1.4 计算机中的信息表示

1.4.1 进位计数制

1.4.1.1 数制的概念

数制是用一组固定的数字和一套统一的规则来表示数目的方法。

按照进位方式计数的数制叫进位计数制，简称进制。十进制即逢十进一，生活中也常常遇到其它进制，如六十进制（每分钟 60 秒、每小时 60 分钟，即逢 60 进 1），十二进制，十六进制等。

进位计数涉及基数与各数位的位权。

1.5.1.2 基数

基数是指该进制中允许选用的基本数码的个数。

每一种进制都有固定数目的计数符号。

十进制：基数为 10，10 个记数符号，0、1、2、……9。每一个数码符号根据它在这个数中的位置(数位)，按“逢十进一”来决定其实际数值。

二进制：基数为 2，2 个记数符号，0 和 1。每个数码符号根据它在这个数中的数位，按“逢二进一”来决定其实际数值。

八进制：基数为 8，8 个记数符号，0、1、2、……7。每个数码符号根据它在这个数中的数位，按“逢八进一”来决定其实际的数值。

十六进制：基数为 16，16 个记数符号，0-9，A，B，C，D，E，F。其中 A~F 对应十进制的 10~15。每个数码符号根据它在这个数中的数位，按“逢十六进一”决定其实际的数值。

1.5.1.3 位权

一个数码处在不同位置上所代表的值不同，如数字 6 在十位数位置上表示 60，在百位数上表示 600，而在小数点后 1 位表示 0.6，可见每个数码所表示的数值等于该数码乘以一个与数码所在位置相关的常数，这个常数叫做位权。位权的大小是以基数为底、数码所在位置的序号为指数的整数次幂。十进制的个位数位置的位权是 10^0 ，十位数位置上的位权为 10^1 ，小数点后 1 位的位权为 10^{-1} 。

1.4.2 二进制数码

二进制数码有两个基本特征：

用 0、1 两个不同的符号组成的符号串表示数量；

相邻两个符号之间遵循“逢 2 进 1”的原则，即左边的一位所代表的数目是右边紧邻同一符号所代表的数目的 2 倍。

在计算机中数据的最小单位是 1 位二进制代码，简称为位(bit)。8 个连续的 bit 称为一个字节(byte)。

1.4.4 计算机中数据及编码

1.5.4.2 数据的单位

计算机中数据的常用单位有位、字节和字。

1. 位(Bit)

计算机中最小的数据单位是二进制的—一个数位，简称为位(英文名称为 bit，读音为比特)。计算机中最直接、最基本的操作就是对二进制位的操作。

2. 字节(Byte)

字节简称为 B，人们采用 8 位为 1 个字节。1 个字节由 8 个二进制数位组成。

字节是计算机中用来表示存储空间大小的基本容量单位。

1B=8bit

1KB=1024B=2¹⁰B

1MB=1024KB=2¹⁰KB=2²⁰B=1024×1024B

1GB=1024MB=2¹⁰MB=2³⁰B=1024×1024KB

1TB=1024GB=2¹⁰GB=2⁴⁰B=1024×1024MB

1KB=1024B 字节，“K”的意思是“千”

1MB=1024KB 字节，“M”读“兆”。

1GB=1024MB 字节，“G”读“吉”。

1TB=1024GB 字节，“T”读“梯”。

要注意位与字节的区别：位是计算机中最小数据单位，字节是计算机中基本信息单位。

3. 字(word)

每个字中二进制位数的长度，称为字长。字长越长，精度就越高。

注意字与字长的区别，字是单位，而字长是指标，指标需要用单位去衡量。正象生活中重量与公斤的

关系，公斤是单位，重量是指标，重量需要用公斤加以衡量。

1.5 ODBC 连接

ODBC (Open Database Connectivity, 开放数据库互连) 是微软公司开放服务结构 (WOSA, Windows Open Services Architecture) 中有关数据库的组成部分，它建立了一组规范，并提供了一组对数据库访问的标准 API (应用程序编程接口)。ODBC 本身也提供对 SQL 语言的支持，用户可以直接将 SQL 语句送给 ODBC。

一个基于 ODBC 的应用程序对数据库的操作不依赖任何 DBMS，不直接与 DBMS 打交道，所有的数据库操作由对应的 DBMS 的 ODBC 驱动程序完成。也就是说，不论是 FoxPro、Access，MySQL 还是 Oracle 数据库，均可用 ODBC API 进行访问。

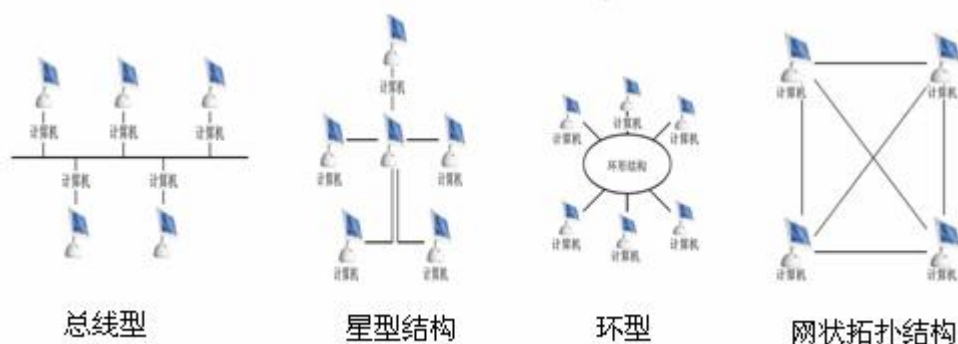
第二章 计算机网络基础

可以划分成局域网、城域网和广域网。

常用的网络传输媒介可分为两类：一类是有线的：一类是无线的。

有线传输媒介主要有双绞线、同轴电缆和光纤；无线媒介有微波、无线电、激光和红外线等。

网络拓扑是指网络上各个节点相互间的连接形式。基本的拓扑结构有三种：总线型、星型和环型。实际网络的拓扑结构可能是由这些基本拓扑混合构成的。



网络硬件组成：计算机、交换机、网线、等。

软件配置：包括网络协议、IP 设置、DNS 设置、网段、网关等。

第三章 数据库基础

3.1 数据库系统概述

1. 数据 (Data)

数据是数据库中存储的基本对象。

2. 数据库 (Database, 简称 DB)

数据库，是按一定的格式存放数据的仓库，数据库是一个结构化的数据集合。

3. 数据库管理系统 (Database Management System 简称 DBMS)

数据库管理系统是位于用户与操作系统之间的一层数据管理软件。它的主要功能包括以下几个方面：

(1) 数据定义功能

DBMS 提供数据定义语言 (Data Definition Language, 简称 DDL)，用户通过它可以方便地对数据库

中的数据对象进行定义。

(2) 数据操纵功能

DBMS 还提供数据操纵语言 (Data Manipulation Language, 简称 DML), 用户可以使用 DML 操纵数据实现对数据库的基本操作, 如查询、插入、删除和修改等。

(3) 数据库的运行管理

数据库在建立、运用和维护时由数据库管理系统统一管理、统一控制, 以保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

(4) 数据库的建立和维护功能

它包括数据库初始数据的输入、转换功能, 数据库的转储、恢复功能。数据库的重组织功能和性能监视、分析功能等。这些功能通常是由一些实用程序完成的。

数据库管理系统是数据库系统的一个重要组成部分。

4. 数据库系统 (Database System, 简称 DBS)

数据库系统是指在计算机系统中引入数据库后的系统, 一般由数据库、数据库管理系统 (及其开发工具)、应用系统、数据库管理员和用户构成。应当指出的是, 数据库的建立、使用和维护等工作只靠一个 DBMS 远远不够, 还要有专门的人员来完成, 这些人被称为数据库管理员 (Database Administrator, 简称 DBA)。

3.2 关系数据库标准语言 SQL

3.2.1 SQL 概述

SQL 语言是一个综合的、功能极强、简单易学的语言。SQL 语言包含数据查询 (Data Query)、数据操纵 (Data Manipulation)、数据定义 (Data Definition) 和数据控制 (Data Control) 功能。

3.2.2 数据定义

SQL 的数据定义功能包括定义表、定义视图和定义索引。

视图是基于基本表的虚表, 索引是依附于基本表的,

3.2.2.1 创建、删除与修改基本表

一、创建基本表

建立数据库最重要的一步就是创建一些基本表。SQL 语言使用 CREATE TABLE 语句创建基本表, 其一般格式如下:

```
CREATE TABLE <表名> (<列名><数据类型>[列级完整性约束条件]
                        [, <列名><数据类型>[列级完整性约束条件]].....
                        [, <表级完整性约束条件>]);
```

其中<表名>是所要定义的基本表的名字, 它可以由一个或多个属性 (列) 组成。

二、修改基本表

随着应用环境和应用需求的变化, 有时需要修改已建立好的基本表, SQL 语言用 ALTER TABLE 语句修改基本表, 其一般格式为:

```
ALTER TABLE <表名> -
[ADD<新列名><数据类型> (完整性约束) ]
[DROP<完整性约束名>]
```

[**MODIFY** <列名><数据类型>];

其中<表名>是要修改的基本表，**ADD** 子句用于增加新列和新的完整性约束条件，**DROP** 子句用于删除指定的完整性约束条件，**MODIFY** 子句用于修改原有的列定义，包括修改列名和数据类型。

三、删除基本表

当某个基本表不再需要时，可以使用 **DROP TABLE** 语句删除它。其一般格式为：

DROP TABLE <表名>

基本表定义一旦删除，表中的数据、此表上建立的索引和视图都将自动被删除掉。因此执行删除基本表的操作一定要格外小心。

注意：有的系统，如 Oracle，删除基本表后建立在此表上的视图定义仍然保留在数据字典中。但是，当用户引用时就报错。

3.2.2.2 建立与删除索引

建立索引是加快查询速度的有效手段。用户可以根据应用环境的需要，在基本表上建立一个或多个索引。

一、建立索引

在 SQL 语言中，建立索引使用 **CREATE INDEX** 语句，其一般格式为：

CREATE[**UNIQUE**][**CLUSTER**]**INDEX**<索引名>

ON<表名> (<列名>[<次序>][, <列名>[<次序>]].....);

其中，<表名>是要建索引的基本表的名字，索引可以建立在该表的一列或多列上，各列名之间用逗号分隔。每个<列名>后面还可以用<次序>指定索引值的排列次序，可选 **ASC**（升序）或 **DESC**（降序），缺省值为 **ASC**。**UNIQUE** 表明此索引的每一个索引值只对应唯一的数据记录。

CLUSTER 表示要建立的索引是聚簇索引。所谓聚簇索引是指索引项的顺序与表中记录的物理顺序一致的索引组织。

用户可以在最常查询的列上建立聚簇索引以提高查询效率。显然在一个基本表上最多只能建立一个聚簇索引。建立聚簇索引后，更新索引列数据时，往往导致表中记录的物理顺序的变更，代价较大，因此对于经常更新的列不宜建立聚簇索引。

二、删除索引

在 SQL 语言中，删除索引使用 **DROP INDEX** 语句，其一般格式为：

DROP INDEX<索引名>

3.2.3 查询

数据库查询是数据库的核心操作。SQL 语言提供了 **SELECT** 语句进行数据库的查询，该语句具有灵活的使用方式和丰富的功能。其一般格式为：

SELECT [**ALL**][**DISTINCT**<目标列表表达式>[, <目标列表表达式>]...

FROM<表名或视图名>[, 表名或视图名]...

[**WHERE** <条件表达式>]

[**GROUP BY**<列名 1>[**HAVING** <条件表达式>]]

[**ORDER BY**<列名 2>[**ASC**][**DESC**]];]

整个 **SELECT** 语句的含义是，根据 **WHERE** 子句的条件表达式，从 **FROM** 子句指定的基本表或视图中找出满足条件的元组。再按 **SELECT** 子句中的目标列表表达式，选出元组中的属性值形成结果表。如果有 **GROUP** 子句，则将结果按<列名 1>的值进行分组，该属性列值相等的元组为一个组。通常会在每组中作用集函数。如果 **GROUP** 子句带 **HAVING** 短语，则只有满足指定条件的组才予输出。如果有 **ORDER** 子句，则结果表还要按<列名 2>的值的升序或降序排序。

3.2.3.1 单表查询

单表查询是指仅涉及一个表的查询。

一、选择表中的若干列

选择表中的全部列或部分列，这就是投影运算。

1. 查询指定列

例 1：查询全体学生的学号与姓名。

```
SELECT Sno, Sname
```

```
FROM Student;
```

例 2：查询全体学生的姓名、学号、所在系。

```
SELECT Sname, Sno, Sdept
```

```
FROM Student;
```

2. 查询全部列

例 3 查询全体学生的详细记录。

```
SELECT *          等价于：SELECT Sno, Sname, Ssex, Sage, Sdept
```

```
FROM Student
```

```
FROM Student;
```

3. 查询经过计算的值

例 4：查全体学生的姓名及其出生年份。

```
SELECT Sname, 2008-Sage
```

```
FROM Student;
```

例 5：查询全体学生的姓名、出生年份和所有系，要录用小写字母表示所有系名。

```
SELECT Sname, 'Year of Birth: ', 2008-Sage, ISLOWER (Sdept)
```

```
FROM Student
```

二、选择表中的若干元组

1. 消除取值重复的行

两个本来并不完全相同的元组，投影到指定的某些列上后，可能变成相同的行了。

例 6：查询选修了课程的学生学号。

```
SELECT Sno FROM SC;
```

该查询结果里包含了许多重复的行。如果想去掉结果表中的重复行，必须指定 **DISTINCT** 短语：

```
SELECT DISTINCT Sno FROM SC;
```

2. 查询满足条件的元组

查询满足指定条件的元组可以通过 **WHERE** 子句实现。**WHERE** 子句常用的查询条件如表 3.1 所示。

表 3.1 常用的查询条件

查询条件	谓词
比较	=, >, <, >=, <=, !=, <>, !>, !<; NOT+ 上述运算符
确定范围	BETWEEN AND, NOT BETWEEN AND
确定集合	IN, NOT IN
字符匹配	LIKE, NOT LIKE
空值	IS NULL, IS NOT NULL
多重条件	AND, OR

(1) 比较大小

用于进行比较的运算符一般包括：

=（等于），>（大于），<（小于），>=（大于等于），<=（小于等于），!=或<>（不等于）。有些产品

中还包括: !> (不大于), !< (不小于)。

逻辑运算符 NOT 可与比较运算符同用, 对条件求非。

例 7: 查询计算机系全体学生的名单。

```
SELECT Sname FROM Student WHERE Sdept='CS';
```

例 8: 查询所有年龄在 20 岁以下的学生姓名及其年龄。

```
SELECT Sname, Sage FROM Student WHERE Sage<20;
```

例 9: 查询考试成绩有不及格的学生的学号。

```
SELECT DISTINCT Sno FROM Course WHERE Grade<60;
```

这里使用了 DISTINCT 短语, 当一个学生有多门课程不及格, 他的学号也只列一次。

(2) 确定范围

谓词 BETWEEN...AND...和 NOT BETWEEN ...AND....可以用来查找属性值在(或不在)指定范围内的元组, 其中 BETWEEN 后是范围的下限(即低值), AND 后是范围的上限(即高值)。

例 10: 查询年龄在 20-23 岁(包括 20 岁和 23 岁)之间的学生的姓名、系别和年龄。

```
SELECT Sname, Sdept, Sage FROM Student WHERE Sage BETWEEN 20 AND 30;
```

例 11: 查询年龄不在 20-23 岁之间的学生姓名、系别和年龄。

```
SELECT Sname, Sdept, Sage FROM Student  
WHERE Sage NOT BETWEEN 20 AND 23;
```

(3) 确定集合

谓词 IN 可以用来查找属性值属于指定集合的元组。

例 12: 查询信息系(IS)、数学系(MA)和计算机科学系(CS)学生的姓名和性别。

```
SELECT Sname, Ssex FROM Student WHERE Sdept IN ('IS', 'MA', 'CS');
```

与 IN 相对的谓词是 NOT IN, 用于查找属性值不属于指定集合的元组。

例 13: 查询既不是信息系、数学系, 也不是计算机科学系的学生的姓名和性别。

```
SELECT Sname, Ssex FROM Student WHERE Sdept NOT IN ('IS', 'MA', 'CS');
```

(4) 字符匹配

谓词 LIKE 可以用来进行字符串的匹配。其一般语法格式如下:

```
[NOT] LIKE '<匹配串>'[ESCAPE '<换码字符>']
```

其含义是查找指定的属性列值与<匹配串>相匹配的元组。<匹配串>可以是一个完整的字符串, 也可以含有通配符%和_。其中:

● % (百分号) 代表任意长度(长度可以为 0)的字符串。例如 a%b 表示以 a 开头, 以 b 结尾的任意长度的字符串。如 acb, addgb, ab 等都满足该匹配串。

● _ (下横线) 代表任意单个字符。

例如 a_b 表示以 a 开头, 以 b 结尾的长度为 3 的任意字符串。如 acb, afb 等都满足该匹配串。

例 14: 查询学号为 95001 的学生的详细情况。

```
SELECT * FROM Student WHERE Sno LIKE '95001';
```

等价于 SELECT * FROM Student WHERE Sno = '95001';

如果 LIKE 后面的匹配串中不含通配符, 则可以用 = (等于) 运算符取代 LIKE 谓词, 用 != 或 <> (不等于) 运算符取代 NOT LIKE 谓词,

例 15: 查询所有姓刘的学生的姓名、学号和性别。

```
SELECT Sname, Sno, Ssex FROM Student WHERE Sname LIKE '刘%';
```

例 16: 查询姓“欧阳”且全名为三个汉字的学生的姓名。

```
SELECT Sname FROM Student WHERE Sname LIKE '欧阳__';
```

注意: 一个汉字要占两个字的位置, 所以匹配串欧阳后面需要跟两个_。

例 17: 查询名字中第 2 个字为“阳”字的学生的姓名和学号。

```
SELECT Sname, Sno FROM Student WHERE Sname LIKE '__阳%';
```

例 18: 查询所有不姓刘的学生姓名。

```
SELECT Sname FROM Student WHERE Sname NOT LIKE '刘%';
```

如果用户要查询的字符串本身就含有%或_, 这时就要使用 ESCAPE '<码字符>' 短语对通配符进行转义了。

例 19: 查询 DB_Design 课程的课程号和学分。

```
SELECT Cno, Ccredit FROM Course  
WHERE Cname LIKE 'DB_Design' ESCAPE '\';
```

ESCAPE '\ ' 短语表示 \ 为换码字符, 这样匹配串中紧跟在 \ 后面的字符 “_” 不再具有通配符的含义, 转义为普通的 “_” 字符。

例 20: 查询以 “DB_” 开头, 且倒数第 3 个字符为 i 的课程的具体情况。

```
SELECT * FROM Course WHERE Cname LIKE 'DB_%i_' ESCAPE '\';
```

这里的匹配串 “DB_%i_” 中, 第 1 个 _ 前面有换码字符, 所以它被转义为普通的 _ 字符。而 % 前、第 2 和第 3 个 _ 前面均没有换码字符, 所以它们仍作为通配符。

(5) 涉及空值的查询

例 21: 某些学生选修课程后没有参加考试, 所以有选课记录, 但没有考试成绩。查询缺少成绩的学生的学号和相应的课程号。

```
SELECT Sno, Cno FROM SC WHERE Grade IS NULL;
```

注意这里的 “IS” 不能用等号 (=) 代替。

例 22: 查询所有有成绩的学生学号和课程号。

```
SELECT Sno, Cno FROM SC WHERE Grade IS NOT NULL;
```

(6) 多重条件查询

逻辑运算符 AND 和 OR 可用来联结多个查询条件。AND 的优先级高于 OR 但用户可以用括号改变优先级。

例 23: 查询计算机系年龄在 20 岁以下的学生姓名。

```
SELECT Sname FROM Student WHERE Sdept = 'CS' AND Sage < 20;
```

三、对查询结果排序

用户可以用 ORDER BY 子句对查询结果按照一个或多个属性列的升序 (ASC) 或降序 (DESC) 排列, 缺省值为升序。

例 24: 查询选修了 3 号课程的学生学号及其成绩, 查询结果按分数的降序排列。

```
SELECT Sno, Grade FROM SC WHERE Cno = '3' ORDER BY Grade DESC;
```

对于空值, 若按升序排, 含空值的元组将最后显示。若按降序排, 空值的元组将最先显示。

例 25: 查询全体学生情况, 查询结果按所在系的系号升序排列, 同一系中的学生按年龄降序排列。

```
SELECT * FROM Student ORDER BY Sdept, Sage DESC;
```

四、使用集函数

为了进一步方便用户, 增强检索功能, SQL 提供了许多函数, 主要有:

COUNT ([DISTINCT ALL]*)	统计元组个数
COUNT ([DISTINCT ALL]<列名>)	统计一列中值的个数
SUM ([DISTINCT ALL]<列名>)	计算一列值的总和 (此列必须是数值型)
AVG ([DISTINCT ALL]<列名>)	计算一列值的平均值 (此列必须是数值型)
MAX ([DISTINCT ALL]<列名>)	求一列值中的最大值
MIN ([DISTINCT ALL]<列名>)	求一列值中的最小值

如果指定 DISTINCT 短语, 则表示在计算时要取消指定列中的重复值。如果不指定 DISTINCT 短语或指定 ALL 短语 (ALL 为缺省值), 则表示不取消重复值。

例 26: 查询学生总人数。

```
SELECT COUNT (*) FROM Student;
```

例 27: 查询选修了课程的学生人数。

```
SELECT COUNT (DISTINCT Sno) FROM SC;
```

学生每选修一门课, 在 SC 中都有一条相应的记录。一个学生要选修多门课程, 为避免重复计算学生人数, 必须在 COUNT 函数中用 DISTINCT 短语。

例 28: 计算 1 号课程的学生平均成绩。

```
SELECT AVG (Grade) FROM SC WHERE Cno = '1';
```

例 29: 查询选修 1 号课程的学生最高分数。

```
SELECT MAX (Grade) FROM SC WHERE Cno = '1';
```

五、对查询结果分组

GROUP BY 子句将查询结果表按某一列或多列值分组。值相等的为一组。

对查询结果分组的目的是为了细化集函数的作用对象。如果未对查询结果分组, 集函数将作用于整个查询结果, 如上面的例 26, 例 27, 例 28, 例 29。分组后集函数将作用于每一个组, 即每一组都有一个函数值。

例 30: 求各个课程号及相应的选课人数。

```
SELECT Cno, COUNT (Sno) FROM SC GROUP BY Cno;
```

该语句对查询结果按 Cno 的值分组, 所有具有相同 Cno 值的元组为一组, 然后对每一组作用集函数 COUNT 计算, 以求得该组的学生人数。

如果分组后还要求按一定的条件对这些组进行筛选, 最终只输出满足指定条件的组, 则可以使用 HAVING 短语指定筛选条件。

例 31: 查询选修了 3 门以上课程的学生学号。

```
SELECT Sno FROM SC GROUP BY Sno HAVING COUNT (*) > 3;
```

这里先用 GROUP BY 子句按 Sno 加进行分组, 再用集函数 COUNT 对每一组计数。HAVING 短语指定选择组的条件, 只有满足条件 (即元组个数>3, 表示此学生选修的课超过 3 门) 的组才会被选出来。

WHERE 子句与 HAVING 短语的区别在于作用对象不同, WHERE 子句作用于基本表或视图, 从中选择满足条件的元组。HAVING 短语作用于组, 从中选择满足条件的组。

3.2.3.2 连接查询

前面的查询都是针对一个表进行的。若一个查询同时涉及两个以上的表, 则称之为连接查询, 连接查询是关系数据库中最主要的查询, 包括等值连接、自然连接、非等值连接查询、自身连接查询、外连接查询和复合条件连接查询。

一、等值与非等值连接查询

连接查询中用来连接两个表的条件称为连接条件或连接谓词, 其一般格式为:

[<表名一>].<列名> <比较运算符> [<表名二>].<列名>

其中比较运算符主要有: =、>、<、>=、<=、!=。

此外连接谓词还可以使用下面形式:

[<表名一>].<列名> BETWEEN [<表名二>].<列名一> AND [<表名二>].<列名二>

当连接运算符为=时, 称为等值连接。使用其他运算符称为非等值连接。

连接谓词中的列名称为连接字段。连接条件中的各连接字段类型必须是可比的, 但不必是相同的。

例 32: 查询每个学生及其选修课程的情况。

学生情况存放在 Student 表中, 学生选课情况存放在 SC 表中, 所以本查询实际上涉及 Student 与 SC 两个表。这两个表之间的联系是通过公共属性 Sno 实现的。

```
SELECT Student.*, SC.* FROM Student, SC WHERE Student.Sno = SC.Sno;
```

本例中, SELECT 子句与 WHERE 子句中的属性名前都加上了表名前缀, 这是为了避免混淆。如果属性名在参加连接的各表中是唯一的, 则可以省略表名前缀。

连接运算中有两种特殊情况, 一种为自然连接, 另一种为广义笛卡尔积 (连接)。

广义笛卡尔积是不带连接谓词的连接。两个表的广义笛卡尔积即是两表中元组的交叉乘积。其连接的结果会产生一些没有意义的元组，所以这种运算实际很少使用。

若在等值连接中把目标列中重复的属性列去掉则为自然连接。

例 33：对例 32 用自然连接完成。

```
SELECT Student.Sno, Sname, Ssex, Sage, Sdept, Cno, Grade
FROM Student, SC WHERE Student.Sno = SC.Sno;
```

例中，由于 Sname, Ssex, Sage, Sdept, Cno 和 Grade 属性列在 Student 表与 SC 表中是唯一的，因此引用时可以去掉表名前级，而 Sno 在两个表都出现了，因此引用时必须加上表名前级。

二、自身连接

连接操作不仅可在两个表之间进行也可是一个表与自己进行连接，称为表的自身连接。

例 34：查询每一门课的间接先修课（即先修课的先修课）。

在 Course 表关系中，只有每门课的直接先修课信息，而没有先修课的先修课。要得到这个信息，必须先对一门课找到其先修课，再按此先修课的课程号，查找它的先修课程。这就要将 Course 表与其自身连接。

为清楚起见，可以为 Course 表取两个别名，一个是 FIRST，另一个是 SECOND。

完成该查询的 SQL 语句为：

```
SELECT FIRST.Cno, Second.Pcno FROM Course FIRST, Course SECOND
WHERE FIRST.Cpno = SECOND.Cno;
```

三、外连接

在通常的连接操作中，只有满足连接条件的元组才能作为结果抽出。

我们想以 Student 表为主体列出每个学生的基本情况及其选课情况，若某个学生没有选课，只输出其基本情况信息，其选课信息为空值即可，这时就需要使用外连接（Outer Join），可以如下改写例 33：

```
SELECT Student.Sno, Sname, Ssex, Sage, Sdept, Cno, Grade
FROM Student, SC WHERE Student.Sno = SC.Sno (*);
```

外连接的表示方法为，在连接谓词的某一边加符号*（有的数据库系统中用+）。外连接就好像是为符号*所在边的表（本例是 SC 表）增加一个“万能”的行，这个行全部由空值组成。它可以和另一边的表（本例是 Student 表）中所有不满足连接条件的元组进行连接。本例中即与 Student 表的 95003 和 95004 元组进行连接，由于这个“万能”行的各列全部是空值，因此在连接结果中，95003 和 95004 两行中来自 SC 表的属性值全部是空值。

如果外连接符出现在连接条件的右边，称其为右外连接（如本例）。如果外连接符出现在连接条件的左边，则称为左外连接。

四、复合条件连接

上面各个连接查询中，WHERE 子句中只有一个条件：即连接谓词。WHERE 子句中可以有多个连接条件，称为复合条件连接。

例 35：查询选修 2 号课程且成绩在 90 分以上的所有学生。

```
SELECT Student.Sno, Sname FROM Student, SC
WHERE SC.Cno = '2' AND Grade > 90;
```

连接操作除了可以是两表连接，一个表与其自身连接外，还可以是两个以上的表进行连接，后者通常称为多表连接。

例 36：查询每个学生的学号、姓名、选修的课程名及成绩。

本查询涉及到三个表，完成该查询的 SQL 语句如下：

```
SELECT Student.Sno, Sname, Cname, Grade
FROM Student, SC, Course
WHERE Student.Sno = SC.Sno AND SC.Cno = Course.Cno;
```


3.2.3.2 嵌套查询

在 SQL 语言中，一个 SELECT-FROM-WHERE 语句称为一个查询块。将一个查询块嵌套在另一个查询块的 WHERE 子句或 HAVING 短语的条件中的查询称为嵌套查询。例如：

```
SELECT Sname FROM Student WHERE Sno IN  
    (SELECT Sno FROM SC WHERE Cno = '2');
```

本例中，下层查询块 SELECT Sno FROM SC WHERE Cno = '2' 是嵌套在上层查询块 SELECT Sname FROM Student WHERE Sno IN 的 WHERE 条件中的。上层的查询块称为外层查询或父查询，下层查询块称为内层查询或子查询。SQL 语言允许多层嵌套查询，即一个子查询中还可以嵌套其他子查询。需要特别指出的是，子查询的 SELECT 语句中不能使用 ORDER BY 子句，ORDER BY 子句只能对最终查询结果排序。

嵌套查询一般的求解方法是由里向外处理。即每个子查询在上一级查询处理之前求解，子查询的结果用于建立其父查询的查找条件。

嵌套查询使得多个简单查询构成复杂的查询，从而增强 SQL 的查询能力。以层层嵌套的方式来构造程序正是 SQL (Structured Query Language) “结构化”的含义所在。

一、带有 IN 谓词的子查询

在嵌套查询中，子查询的结果往往是一个集合，所以谓词 IN 是嵌套查询中最经常使用的谓词。

例 37：查询与“刘晨”在同一个系学习的学生。

先分步来完成此查询，然后再构造嵌套查询。

1、确定“刘晨”所在系名

```
SELECT Sdept FROM Student WHERE Sname = '刘晨';
```

2、查找所有在 IS 系学习的学生

```
SELECT Sno, Sname, Sdept FROM Student WHERE Sdept = 'IS';
```

将第 1 步查询嵌入到第 2 步查询的条件中，构造嵌套查询。SQL 语句如下：

```
SELECT Sno, Sname, Sdept FROM Student WHERE Sdept IN (  
    SELECT Sdept FROM Student WHERE Sname = '刘晨');
```

DBMS 求解该查询时，实际上也是分步去做的，类似于上面写的分步过程。

二、带有比较运算符的子查询

带有比较运算符的子查询是指父查询与子查询之间用比较运算符进行连接。当用户能确切知道内层查询返回的是单值时，可以用 >, <, =, >=, <=, != 或 <> 等比较运算符。

例如在例 37 中，由于一个学生只可能在一个系学习，也就是说内查询的结果是一个值，因此可以用 = 代替 IN，其 SQL 语句如下：

```
SELECT Sno, Sname, Sdept FROM Student WHERE Sdept = (  
    SELECT Sdept FROM Student WHERE Sname = '刘晨');
```

需要注意的是，子查询一定要跟在比较符之后，下列写法是错误的：

```
SELECT Sno, Sname, Sdept FROM Student  
WHERE ( SELECT Sdept FROM Student WHERE Sname = '刘晨' ) = Sdept;
```

三、带有 ANY 或 ALL 谓词的子查询

子查询返回单值时可以用比较运算符，而使用 ANY 或 ALL 谓词时则必须同时使用比较运算符。其语义为：

>ANY	大于子查询结果中的某个值
>ALL	大于子查询结果中的所有值
<ANY	小于子查询结果中的某个值
<ALL	小于子查询结果中的所有值

.... ..

以此类推。

3.2.4 数据更新

SQL 中数据更新包括插入数据、修改数据和删除数据三条语句。

3.2.4.1 插入数据

SQL 的数据插入语句 INSERT 通常有两种形式。一种是插入一个元组，另一种是插入子查询结果。后者可以一次插入多个元组。

一、插入单个元组

插入单个元组的 INSERT 语句的格式为：

```
INSERT
INTO <表名> [ (<属性列 1>[, <属性列 2>].....) ]
VALUES (<常量 1>[, <常量 2>].....)
```

其功能是将新元组插入指定表中。其中新记录属性列 1 的值为常量 1, 属性列 2 的值为常量 2…。INTO 子句中没有出现的属性列, 新记录在这些列上将取空值。

但必须注意的是, 在表定义时说明了 NOT NULL 的属性列不能取空值, 否则会出错。

如果 INTO 子句中没有指明任何列名, 则新插入的记录必须在每个属性列上均有值。

例 1: 将一个新学生记录 (学号: 95020; 姓名: 陈冬; 性别: 男; 所在系: IS; 年龄: 18 岁) 插入到 Student 表中。

```
INSERT INTO Student VALUES ('95020', '陈冬', '男', 'IS', 18);
```

例 2: 插入一条选课记录 ('95020', '1')

```
INSERT INTO SC (Sno, Cno) VALUES ('95020', '1');
```

新插入的记录在 Grade 列上取空值。

二、插入子查询结果

子查询不仅可以嵌套在 SELECT 语句中, 用以构造父查询的条件 (如 3.2.3.3 节所述), 也可以嵌套在 INSERT 语句中, 用以生成要插入的批量数据。插入子查询结果的 INSERT 语句的格式为:

```
INSERT
INTO <表名> [ (<属性列 1>[, <属性列 2>].....) ]
```

子查询;

例 3: 对每一个系, 求学生的平均年龄, 并把结果存入数据库。

首先在数据库中建立一个新表, 其中一列存放系名, 另一列存放相应的学生平均年龄。

```
CREATE TABLE Deptage (Sdept CHAR (15) Avgage SMALL INT);
```

然后对 Student 表按系分组求平均年龄, 再把系名和平均年龄存入新表中。

```
INSERT INTO Deptage (Sdept, Avgage) SELECT Sdept, AVG (Sage)
FROM Student GROUP BY Sdept
```

3.2.4.2 修改数据

修改操作语句的一般格式为:

```
UPDATE <表名>
SET <列名>=<表达式>[, <列名>=<表达式>]...
[WHERE <条件>];
```

其功能是修改指定表中满足 WHERE 子句条件的元组, 其中 SET 子句给出<表达式>的值用于取代相应的属性列值。如果省略 WHERE 子句, 则表示要修改表中的所有元组。

一、修改某一个元组的值

例 4: 将学生 95001 的年龄改为 22 岁。

```
UPDATE Student SET Sage = 22 WHERE Sno = '95001';
```

二、修改多个元组的值

例 5：将所有学生的年龄增加 1 岁。

```
UPDATE Student SET Sage = Sage + 1;
```

三、带子查询的修改语句

子查询也可以嵌套在 UPDATE 语句中，用以构造修改的条件。

例 6：将计算机科学系全体学生的成绩置 0。

```
UPDATE SC SET Grade = 0 WHERE 'CS' = (SELECT Sdept FROM Student
WHERE Student.Sno = SC.Sno);
```

3.2.4.3 删除数据

删除语句的一般格式为：

```
DELETE
FROM <表名>
[WHERE <条件>]
```

DELETE 语句的功能是从指定表中删除满足 WHERE 子句条件的所有元组。如果省略 WHERE 子句，表示删除表中全部元组，但表的定义仍在字典中。也就是说，DELETE 语句删除的是表中的数据，而不是关于表的定义。

一、删除某一个元组的值

例 7：删除学号为 95019 的学生记录。

```
DELETE FROM Student WHERE Sno = '95019';
```

二、删除多个元组的值

例 8：删除所有的学生选课记录。

```
DELETE FROM SC;
```

这条 DELETE 语句将使 SC 成为空表，它删除了 SC 的所有元组。

三、带子查询的删除语句

子查询同样也可以嵌套在 DELETE 语句中，用以构造执行删除操作的条件。

例 9：删除计算机科学系所有学生的选课记录。

```
DELETE FROM SC WHERE 'CS' = (SELECT Sdept
FROM Student WHERE Student.Sno = SC.Sno);
```

四、更新操作与数据库的一致性

增删改操作只能对一个表操作。这会带来一些问题。例如，比如 95019 学生被删除后，有关其选课信息也应同时删除，而这只能通过两条语句进行。

第 1 条语句删除 95019 学生：

```
DELETE FROM Student WHERE Sno = '95019';
```

第 2 条语句删除 95019 学生的选课记录：

```
DELETE FROM SC WHERE Sno = '95019';
```

在执行了第 1 条 DELETE 语句之后，数据库中的数据已处于不一致状态，因为这时实际上已没有学号为 95019 的学生了，但 SC 表中仍然记录着关于 95019 学生的选课信息，即数据的参照完整性受到破坏。只有执行了第 2 条 DELETE 语句之后，数据才重新处于一致状态。但如果执行完一条语句之后，机器突然出现故障，无法再继续执行第 2 条 UPDATE 语句，则数据库中的数据将永远处于不一致状态。因此必须保证这两条语句要么都做，要么不做。为解决这一问题，数据库系统通常都引入了事务（Transaction）的概念。

当向参照表中插入元组，如往 SC 中插入选课记录（95020 111）。系统将自动地检查被参照表（Student 和 Course）中是否存在相应的元组（95020 号学生的记录，1 号课程的记录）如果都存在则操作成功，否则操作失败。

当删除主表中的元组（如学生 95019），可以有以下的策略：

- (1) 系统自动地删除参照表（如 SC）中相应的元组（学生 95019 的选课记录）。
- (2) 系统检查参照表中是否存在相应的元组，如果存在，则操作失败。

解释下面语句的含义：UPDATE 职工表 SET 工资=工资*2.0 WHERE 部门编号='A8'
AND 工资<= (SELECT AVG (工资) FROM 职工表);

第二部分 遥感技术基础

第四章 遥感技术基本概念

4.1 遥感基本概念

遥感（Remote Sensing），广义上泛指从远处探测、感知物体或事物的技术。即不直接接触物体，从远处通过仪器（传感器）探测和接收来自目标的信息（如电场、磁场、电磁波、地震波等），经过信息的传输及处理分析，识别物体的属性及其分布等特征的技术。遥感具有“千里眼、顺风耳”的特征。

人类通过大量的实践，发现地球上每一个物体都在不停地吸收、发射和反射信息和能量，其中有一种是电磁波，并发现不同物体的电磁波特性不同。遥感就是根据这个原理探测地表物体反射和发射的电磁波，从而提取这些物体的信息，完成远距离识别物体。

常见的传感器有：航空摄影机（航摄影）、全景摄影机、多光谱摄影机、RBV（Return Beam Vidicon）反射光道摄像机、MSS（Multispectral Scanner System）多光谱扫描系统、TM（Thematic Mapper）专题制图仪器、ETM（Enhance Thematic Mapper）增强型专题制图仪器、ETM+（Enhance Thematic Mapper Plus）增强型专题制图仪器、HRV（High Resolution Visible range instruments）扫描仪、合成孔径侧视雷达（Side-Looking Airborne Radar, SLAR）等。

MSS 有四个波段，MSS 图像覆盖范围是 $185\text{km} \times 185\text{km}$ ，空间分辨率是 $79\text{m} \times 57\text{m}$ ，优点是费用少，缺点是有时一个探测器会留下空白或者相同的信号难以区别，从而形成条带或者斑纹。

TM 有 7 个波段。

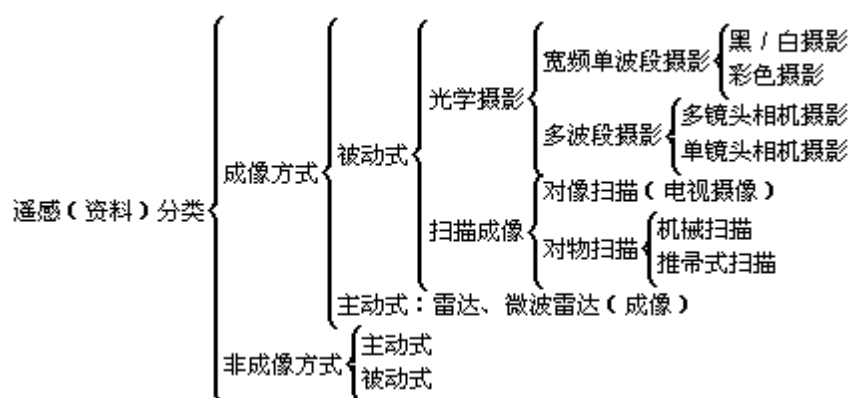
4.2 遥感的特点与分类

4.2.1 遥感的特点

- 1、感测范围大，具有综合、宏观的特点。
- 2、信息量大，具有手段多，技术先进的特点。
- 3、获取信息快，更新周期短，具有动态监测特点。

4.2.2 遥感的分类

- 1、按遥感平台的高度大体上可分为航天遥感、航空遥感和地面遥感。
- 2、按所利用的电磁波的光谱段可分为可见光/反射红外遥感，热红外遥感、微波遥感。
- 3、按研究对象分类可分为资源遥感与环境遥感两大类。
- 4、按应用空间尺度分类 可分为全球遥感、区域遥感和城市遥感。
- 5、根据遥感资料的记录方式和传感器工作方式，又可作下列分类。



成像方式（或称图像方式）就是将所探测到的强弱不同的地物电磁波辐射（反射或发射），转换成深浅不同的（黑白）色调构成直观图像的遥感资料形式，如航空像片、卫星图像等。非成像方式（或非图像方式）则是将探测到的电磁辐射（反射或发射），转换成相应的模拟信号（如电压或电流信号）或数字化输出，或记录在磁带上而构成非成像方式的遥感资料。如陆地卫星 CCT 数字磁带等。

主动式遥感或被动式遥感则是按传感器工作方式的不同所作的分类。所谓主动式是指传感器带有能发射讯号（电磁波）的辐射源，工作时向目标物发射，同时接收目标物反射或散射回来的电磁波，以此所进行的探测。被动式遥感则是利用传感器直接接收来自地物反射自然辐射源（如太阳）的电磁辐射或自身发出的电磁辐射，而进行的探测。

光学摄影亦指通常的摄影，即将探测接收到的地物电磁波依据深浅不同的色调直接记录在感光材料上。扫描方式是将所探测的视场（或地物）划分为面积相等顺序排列的像元，传感器则按顺序以每个像元为探测单元记录其电磁辐射强度，并经转换、传输、处理，或转换成图像显示在屏幕或胶片上，或制成扫描数字产品。

通常，遥感数字图像有以下三种数据格式：

1、BSQ（Band Sequential Format）格式

按波段顺序记录图像数据，该格式最适于对单个波段中任何部分的二维空间存取。

2、BIP（Band Interleaved by Pixel）格式

按像元顺序记录图像数据，即首先存储第一个像元的所有的波段数据，接着是第二个像元的所有波段数据，直到最后一个像元为止。这种格式为图像数据波谱维的存取提供最佳性能，也便于进行图像波谱分析。

3、BIL（Band Interleaved by Line）格式

按波段分行记录图像数据，该格式存储的图像先存储第一个波段的第一行、接着是第二个波段的第一行……，最后波段的第一行；再记录各波段的第二行；每个波段随后的行按照类似的方式交叉存取。这种格式提供了空间和波谱处理之间一种折中方式。

4.2.3 大气窗口

通常把通过大气而较少被反射、吸收或散射的透射率较高的电磁辐射波段称为大气窗口。因此，遥感传感器选择的探测波段应包含在大气窗口之内。

4.2.4 地物光谱的特性

当电磁辐射能量入射到地物表面上，将会出现三种过程：一部分入射能量被地物反射；一部分入射能量被地物吸收，成为地物本身内能或部分再发射出来，一部分入射能量被地物透射。

第五章 遥感图像处理

5.1 遥感图像的几何校正

5.1.1 几何精校正原理

遥感图像几何校正,就是利用控制点改正原始图像的几何变形,产生一幅符合某种地图投影或图件表达要求的新图像。它利用若干个控制点数据确立一个模拟图像几何畸变的数学模型,以此来建立原畸变图像(待校正图像)空间与标准图像(或称参考图像,如地图等)空间的某种对应关系,即所谓映射变换函数或校正变换函数。再依据这种变换函数把待校正图像空间中的全部像元变换到标准空间中去,从而实现图像的几何校正,在我国校正标准空间为高斯-克吕格投影空间。

实现上述两个空间的转换通常采用两种方法:即直接变换法与重采样法。

(一) 直接变换法

首先求出原始图像空间(变形空间)中各个像元在标准空间中的对应位置,然后将相应的亮度值转移过去。从上图可以看出,变形空间网格上的点,直接投影到标准空间后,点的规则派流被打乱,求标准空间网格点的灰度值就可以用插值与网格化方法来实现。

(二) 重采样法

用校正图像空间(标准空间)中的点反求出其在畸变空间(变形空间)的共轭点,然后再利用某种方法确定这一共轭点的亮度值,该亮度值就是校正空间共轭点的亮度。可以有三种方法求解:

1、最邻近点插值法:即把变形点空间中离共轭点最近像元点的灰度值作为共轭的灰度值。

2、双线性内插值法:该方法原理是假设共轭点四周的四个点组成区间内的灰度变换是线性的,从而利用线性插值的方法求得共轭点的灰度值。

3、三次卷积内插法:以实际位置邻近的 16 个像元灰度值来确定输出像元的灰度值的。

三种方法的优缺点如下表:

类别	优点	缺点
最邻近点插值法	算法简单且保持原光谱信息不便	几何精度较差,图像灰度具有不连续性,边界出现锯齿状。
双线性内插值法	计算较简单,图像灰度具有连续性且采样精度比较精确。	细节丧失,图像出现模糊
三次卷积内插法	图像灰度具有连续性且采样精度比较精确	计算量太大

5.1.2 几何精校正步骤

(一) 选取控制点

一般选择在图像和地形图上都容易识别定位的明显地物点,如道路、河流等交叉点,田块拐角,桥头等。控制点要有一定的数量,并且要求分布比较均匀,丘陵山区应尽可能选在高程相似的地段。

初选控制点后还需进行控制点对匹配精度的检验,剔除那些匹配误差超限的像元(误差限额一般定为 0.5,至多一个像元),必要时补充采集控制点,以保证控制点数量与质量。

控制点数量应当超过校正函数多项式系数的个数,一般是 $(n+1) \times (n+2) / 2$,这里 n 是多项式系数个数。

有时待校正的图像比例尺很小,准确识别定位控制点十分困难,还有时工作地区为荒漠或有很大的水面,根本找不到特征鲜明的控制点。在这些情况下可使用一些补救办法建立伪控制点或称模拟控制点。下图(a)给出了称之为重心法求控制点的示意图,假定图中阴影部分为湖水区,把它变成二值图并计算其重

心坐标(x , y)来作为伪控制点。(b)给出了使用延长线段得到人为交点作为伪控制点的示意图。

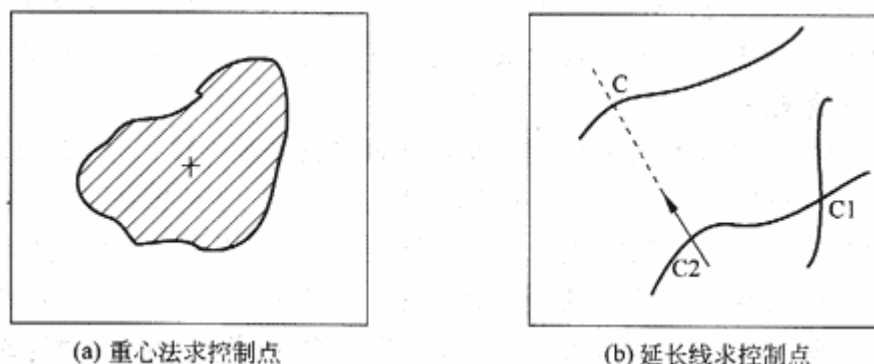


图 2-1 伪控制点产生示意图

(二) 获取校正多项式

选好控制点后即可进入第二步,进行图像的空间变换。多项式近似法是目前实践中最常用的一种方法,原理简单易于理解,计算也较方便,适用于各种遥感类别的图像,特别适合于平坦地区,只要控制点选得好,就可保证取得较高的校正精度。这种方法回避了成像的几何过程,直接对图像畸变本身进行函数逼近,也就是说,校正前后图像相应点之间的坐标关系可以选用一个适当的多项式来表达。

(三) 图像变换

取得变换后图像各像元的灰度值,即对图像进行变换。

在几何校正的最后阶段,必须根据几何畸变的校正式对输入图像进行变换,以便得到无畸变的图像,通常有两种方法:

(1) 对输入的图像的各个像元在变换后的输出图像坐标系上的相应位置进行计算,把各个像元的数据投影到该位置。

(2) 对输出图像的各个像元在输入图像坐标系的相应位置进行逆运算,求出该位置上的像元数据,该方法通常采用的方法。

校正函数可有多种选择:多项式法、共线方程法、随机场内插法等等,其中多项式方法的应用最为普遍。

多项式校正的基本思想就是把图像的变化规律可以看成是平移、缩放、旋转、仿射、偏扭、弯曲等形变的组合。利用有限的控制点的已知坐标,求解多项式的系数,确定变换函数,然后将各个像元带入多项式进行计算,得到校正后的坐标。

专业的遥感图像处理软件有: MAPGIS-RSP、ERDAS IMAGINE、ERMAPPER、PCI、ENVI 等。

第六章 遥感图像分析解译

6.1 遥感图像的计算机自动识别分类

图像的分类在数学或计算机处理上可归结为对一模式或数组的分类和识别,或是以适当的判别函数加以判别分类。

监督分类和非监督分类的根本区别在于是否利用训练场地来获取先验的类别知识,监督分类根据训练场提供的样本选择特征参数,建立判别函数,对待分类点进行分类。

非监督分类不需要更多的先验知识,它根据地物的光谱统计特性进行分类。

6.1.1 非监督分类

非监督分类又称边学习边分类法。

非监督分类主要采用聚类分析方法，其常用方法如下：

(1) 分级集群法 (Hierarchical Clustering)

(2) 动态聚类法

6.1.2 监督分类

监督分类又称训练场地法或先学习后分类法。

监督分类中常用的具体分类方法包括：

1、最小距离分类法

包括最小距离判别法和最近邻域分类法。

最小距离分类法原理简单，分类精度不高，但计算速度快，适合在快速浏览分类概况中使用。

2、多级切割分类法

多级分割法要求分割面总是与各特征轴正交，需要先进行主成分分析或采用其他方法对各轴进行相互独立的正交变换，然后进行多级分割。

3、特征曲线窗口法

4、最大似然比分类法

第三部分 测绘基础

第七章 数字测图概述

7.1 测绘学的内容和任务

测绘学按照研究范围研究对象及采用技术手段的不同，分为以下几个分支学科：大地测量学、摄影测量学、地图学、工程测量学、海洋测绘学。

1. 大地测量学

大地测量学是研究和确定地球的形状、大小、重力场、整体与局部运动和地表面点的几何位置以及它们的变化的理论和技术的学科。大地测量学是测绘学各分支学科的理论基础，基本任务是建立地面控制网、重力网，精确测定控制点的空间三维位置，为地形测图提供控制基础，为各类工程施工测量提供依据，为研究地球形状、大小、重力场及其变化，地壳形变及地震预报提供信息。现代大地测量学包括三个基本分支：几何大地测量学、物理大地测量学和空间大地测量学。

2. 摄影测量学

摄影测量学是研究、确定被摄物体的形状、大小和空间位置的一门学科。

按距离远近分有航空摄影测量、航天摄影测量、地面摄影测量、近景摄影测量和显微摄影测量；按用途分有地形摄影测量和非地形摄影测量；按技术处理方法分，有模拟法摄影测量、解析法摄影测量和数字摄影测量。

3. 地图学

地图学是研究模拟和数字地图的基础理论、设计、编绘、复制的技术方法以及应用的学科。地图学由理论部分、制图部分和地图应用部分组成。地图是测绘工作的重要产品。

4. 工程测量学

工程测量学是一门应用学科，按其研究对象可分为：建筑工程测量、水利工程测量、矿山测量、铁路工程测量、公路工程测量、输电线路与输油管道测量、桥梁工程测量、隧道工程测量、港口工程测量、军事工程测量、城市建设测量以及三维工业测量、精密工程测量、工程摄影测量等。

5. 海洋测绘学

海洋测绘学是以海洋水体和海底为对象，研究海洋定位、测定海洋大地水准面和平均海面、海底和海面地形、海洋重力、海洋磁力、海洋环境等自然和社会信息的地理分布及编制各种海图的理论和技术的学科。内容包括海洋大地测量、海道测量、海底地形测量和海图编制。

测量学又称为普通测量学或地形测量学，其主要内容包括角度测量、距离测量、水准测量、控制测量、地形图测绘及地形图的应用。

7.2 数字测图的发展概况

传统的地形测量是利用测量仪器对地球表面局部区域内的各种地物、地貌特征点的空间位置进行测定，以一定的比例尺并按图示符号将其绘制在图纸上，即通常所称的白纸测图。这种测图方法的实质是图解法测图。在测图过程中，数字的精度由于刺点、绘图、图纸伸缩变形能够因素的影响会大大降低，而且工序多、劳动强度大、质量管理难。在今天的信息时代，纸质地形图已难承载诸多图形信息，更新也极不方便，难以适应信息时代经济建设的需要。

随着科学技术的进步和计算机技术的发展及其各个领域的渗透，以及电子全站仪、GPS-RTK 技术等先进测量仪器和技术的广泛应用，地形测量向自动化和数字化方向发展，数字化测图技术应运而生。数字测图与图解法测图相比，以其特有的高自动化、全数字化、高精度的显著优势而具有广阔的发展前景。

数字测图实质上是一种全解析机助测图方法，在地形测量发展过程中这是一次根本性的技术变革。这种变革主要表现在：图解法测图的最终成果是地形图，图纸是地形信息的唯一载体；数字测图地形信息的载体是计算机的存储介质（磁盘或光盘），其提交的成果可供计算机处理、远距离传输、多方共享的数字地形图数据文件，通过数控绘图仪可输出地形图。另外，利用数字地形图可生成电子地图和数字地面模型（DTM）。

广义的数字测图包括：利用全站仪或其他测量仪器进行野外数字化测图；利用手持数字化仪或扫描数字化仪对纸质地形图的数字化；利用航摄、遥感相片进行数字化测图等技术。

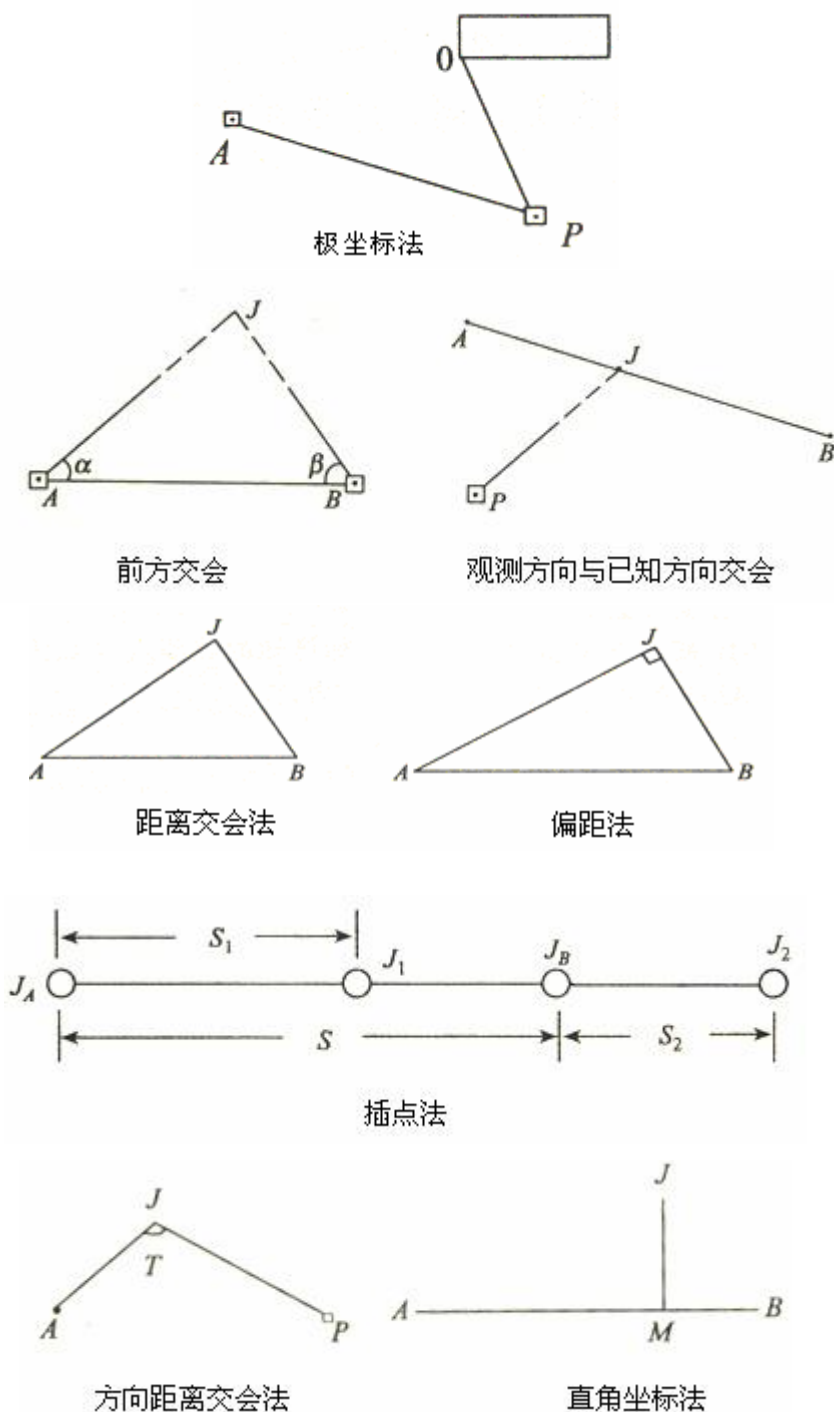
在 20 世纪 80 年代，摄影测量经历了模拟法、解析法发展为数字摄影测量。数字摄影测量是把摄影所获得的影象进行数字化得到数字化影象，由计算机进行数字处理，从而提供数字地形图或专题图、数字地面模型等各种数字化产品。

大比例尺地面数字测图是 20 世纪 70 年代电子速测仪问世后发展起来的，80 年代初全站型电子速测仪的发展加速了数字测图的研究和应用。

20 世纪 90 年代出现了载波相位差分技术，又称 RTK（Real Time Kinematic）实时动态定位技术，这种测量模式是位于基准站（已知的基准点）的 GPS 接收机通过数据链将其观测值及基准站坐标信息一起发给流动站的 GPS 接收机，流动站不仅接收来自参考站的数据，还直接接收 GPS 卫星发射的观测数据组成相位差分观测值，进行实时处理，能够实时提供测点在指定坐标系的三维坐标成果，在 20KM 测程内可达到厘米级的测量精度。实时差分观测时间短，并能实时给出定位坐标。

7.3 测定碎部点坐标的基本方法

在地面数字测图中，测定碎部点坐标的基本方法主要有极坐标法、方向交会法、量距法、方向距离交会法、直角坐标法等。



第八章 地籍测量

城镇地籍是用来记载城镇土地及其附着物的位置、界址、类型、数量、权属和用途等基本状况的簿册。自古以来我国历代政府均有编制，如明代的《鱼鳞图册》，就详细记载了每乡每户土地的亩数和方圆四至，并绘制成图，作为征收田赋和解决土地纷争的主要依据。当前，随着社会经济和科学技术的进步与发展，地籍已成为国家依法管理土地、巩固土地的社会主义公有制、保护土地所有者和使用者的合法权益、制定社会经济发展计划、珍惜和合理利用土地、促进国民经济协调、健康、持续发展的重要依据。

8.1 地籍测量概述

初始地籍测量是指在初始土地权属调查的基础上,利用测绘仪器,以科学的方法,在调查区域内,建立地籍控制网,测量每宗土地的地籍要素,绘制地籍图,为土地登记提供依据。初始地籍测量的内容包括:地籍平面控制测量、地籍细部测量。地籍测量不同于普通的地形测量。地形测量是在工作范围内测量某一时点的地物和地貌;地籍测量除测量地籍调查范围内的地物外,还主要测量反映土地权属的地籍要素,而反映地貌的高程等要素不是地籍测量的重点。地籍测量应随着每宗地的土地登记的变更而不断地更新,时时保证地籍资料的现势性。地籍测量是测绘技术与法律的综合应用,它以测定界址为重点,实地无论有无明显界线,都必须查明、测量界址点线并标定于地籍图上。因此,地籍测绘人员不仅要具有熟练的测绘技能,而且还应熟知相关的法律法规。

8.1.1 地籍控制测量概述

由控制点构成的几何图形,称为控制网,对控制网进行布设、观测、记录的测量工作称为控制网测量,根据控制网测量成果,进行控制点坐标的计算称为控制网平差。控制网测量与控制网平差统一称为控制测量。我们一般将用常规测角、测距手段建立的控制网称为地面控制网。应用 GPS 卫星定位技术建立的控制网称为 GPS 控制网(简称 GPS 网)。因此,根据建立控制网的手段不同,控制网测量分为地面控制网测量和 GPS 控制网测量。

界址点是指宗地权属界线的转折点,即拐点,它是标定宗地权属界线的重要标志。一块土地周围的界址点确定了,则其位置、形状、面积和权属界线也就确定了。在进行宗地权属调查时,界址点应由宗地相邻双方指界人在现场共同认定。确认的界址点上要设置界标,进行编号,并精确测定其位置,以防止日后界标被破坏时,能用测量方法准确地在实地恢复权属界址。

8.1.2 地籍细部测量概述

地籍细部测量是地籍测量的核心,是在地籍平面控制网的基础上,测量每宗土地的权属界线、位置、形状及地类界线等,并计算面积、测绘地籍图、绘制宗地图。准确地测定宗地界址点是管理土地产权的基本保证。宗地图是土地证书的附图和宗地档案的附图。地籍细部测量成果经审核依法登记后,具有法律效力。地籍细部测量的内容包括测定界址点、测绘地籍图、制作宗地图和面积量算等。地籍细部测量方法有解析法、图解勘丈法两种。

这两种方法的主要区别在于测定界址点所采用的方法不同。

(一) 解析法地籍测量的特征

所有界址点都用实测元素按公式解析计算其坐标。为了保证地籍调查的精度,有条件的地区,应尽量采用解析法进行地籍细部测量。

(二) 图解勘丈法地籍测量的特征

不用实测元素计算界址点的坐标,界址点位置应用勘丈的数据确定,界址点的坐标可以在图上图解获得。采用图解勘丈法进行地籍细部测量精度较低,很难满足地籍管理的需要,因此,在条件成熟时,应采用解析法进行地籍细部更新测量。

8.2 地籍控制测量

8.2.1 地籍控制网基本要求

地籍控制测量坐标系统尽量采用国家统一坐标系统。地籍控制测量坐标系最好选国家统一的 3° 带平面直角坐标系,使城镇地籍控制网成为国家网的组成部分,使地籍测量能充分利用国家控制点的成果。在条件不具备的地区,地籍控制网可用地方坐标系或任意坐标系。用任意坐标系时,起算数据应在较大比例尺的地形图或土地利用现状图上图解获取。

在进行地籍控制测量时,应将实地观测值统一投影到高斯正射投影平面上,进行各项改正。为使不同高度海平面的观测值在统一的平面上计算,要求把各项观测值归化至参考椭球面上(或平均海平面上),

防止引起距离变形。在这一因素的影响下，换算到参考椭球面上（或平均海平面上）的两点坐标反算出的距离，往往与实地上两点间的水平距离不一致（未顾及测量误差的影响），这就是坐标系统的长度变形问题。地籍平面控制网的任何两点坐标的要求长度变形小于某个限值，例如，每 km 长度变形小于 2.5cm（即相对变形小于 1/40000）时，这有利于正确测定界址点的坐标、计算面积等。因此，各地区应根据当地的具体情况，选择合适的坐标系统。

8.2.2 首级地籍控制网的布设

首级地籍控制网应能长期使用，因此布设首级地籍控制网的范围应覆盖中长期的城市规划区域。随着全球定位系统（GPS）技术的广泛应用以及 GPS 定位技术具有精度高、速度快、费用省、操作简便、控制点间无需通视等优势，首级平面控制网应优先以 GPS 网形式布设，采用 GPS 接收机测定控制点的坐标。特殊情况下，也可以用导线网、边角网、三角网等地面控制网布设方法，采用全站仪等测定控制点的坐标。首级地籍控制网的精度，要能保证四等网中最弱相邻点的相对点位中误差，以及四等以下各等级控制点相对于上级控制点的点位中误差不超过 $\pm 5\text{cm}$ 。布设首级地籍控制网时，必须先制定技术设计方案，经上级业务主管部门批准后方可实施。

8.2.3 加密控制网的布设

加密控制网应按地籍细部测量的要求安排计划，可分期、分片布设，也可以一次整体布设完成。加密控制网可以采用 GPS 网或导线网的形式布设。当调查区域范围较大，并要求一次整体布设加密控制网时，一般多采用 GPS 网形式布设，布设导线网时，导线宜布设成直伸形状，当复合导线长度超过《城镇地籍调查规程》规定时，应布设成结点网。结点与结点、结点与高级点之间的导线长度，不应超过复合导线长度的 0.7 倍。由于目前全站仪和 GPS 接收机的广泛应用，GPS 网和地面控制网计算平差软件的功能增强，因此，加密控制网的等级一般不再分级，计算时应整体平差。与地形测量相比，地籍测量要求平面控制点有较高的密度。一般说来，地籍平面控制点的密度每 2 km 不少于 10 点。

8.2.4 地籍图根控制网的布设

为满足地籍细部测量和日常地籍管理的需要，在基本控制（首级网和加密控制网）点的基础上，加密的直接供测图及测量界址点使用的控制网称为地籍图根控制网。

（一）地籍图根控制网的特点

与地形测绘的图根控制网相比，地籍图根控制网有下述特点：

（1）地形测绘的图根控制网布设规格（点位密度、精度等）由当时的测图比例尺决定，不同成图比例尺图根控制网的规格相差很大。地籍图根控制网布设规格，应满足测量界址点坐标的精度要求，与地籍图的比例尺大小基本无关。

（2）地形测绘的图根控制点，是为地形细部测量而布设的，测图（整个项目）完成后，便失去了其作用。因此，埋点时原则上设临时性标志。而地籍图根控制点不仅要为当前的地籍细部测量服务，同时还要为日常地籍管理（各种变更地籍测量、土地有偿使用过程中的测量等）服务，因此地籍图根控制点原则上应埋设永久性或半永久性标志。地籍图根控制点在室内处理时，应有示意图、点之记描述。

（3）由于地籍图根控制点密度是根据界址点位置及其密度决定的，几乎所有的道路上都要敷设地籍图根导线。一般说来，地籍图根控制点密度比地形图根控制点密度要大，通常每 2km 应布设 100-400 个地籍图根控制点。

（二）地籍图根控制网布设方式

在城镇建成区，通常采用导线布设地籍图根控制网。为减少图根控制点的二次扩展，应优先布设导线网，以一个或几个街区为单位，布设一级地籍图根导线网，然后采用二级复合导线或导线网加密。在建筑物稀少、通视良好的地区，可以布设地籍图根三角网。

（三）关于地籍图根导线布设的几点特殊规定

(1) 当导线长度小于允许长度的 $1/3$ 时, 只要求导线全长的绝对闭合差小于 13cm , 而不作导线相对闭合差的检查。

(2) 当单导线中的边长短于 10m 时, 允许不作导线角度闭合差检查, 但不得用该导线的边长及方位作为起算数据布设低一级导线或支点。

(3) 当用电磁波测距仪或电子全站仪测量导线的边长时, 导线总长允许放宽。但这时导线全长绝对闭合差不得大于 $\pm 22\text{cm}$, 而相对闭合差: 一级地籍图根导线不得大于 $1/5\,000$, 二级地籍图根导线不得大于 $1/3\,000$ 。

8.3 地籍图测绘

地籍图是土地管理的专题图, 它首先要反映包括行政界线、地籍街坊界线、界址点、界址线、地类、地籍号、面积、座落、土地使用者或所有者及土地等级等地籍要素; 其次要反映与地籍有密切关系的地物及文字注记, 一般不反映地形要素。地籍图是制作宗地图的基础图件。一个国家的整个国土范围由于被占用或使用、利用而被分割成许多地块和土地权属单位, 并且无一遗漏, 那么整个国土面积, 不论城镇、农村, 还是边远地区, 均必须测设地籍图, 因此地籍图具有国家基本图的特性。

地籍图测量方法分模拟法和数字法两种。数字法测绘地籍图是从野外数据采集及数据组织到绘制地籍图都由计算机等电子设备辅助完成。模拟法测绘地籍图基本靠手工完成。不论采用何种方法测绘地籍图都应充分利用宗地草图数据、权属调查勘丈数据及解析界址点坐标数据。

利用地形图编绘地籍图时, 地形图的比例尺、现势性及精度应满足要求, 在制作地籍图之前, 必须利用宗地草图中的勘丈值全面检核地形图的正确性。重点检核与界址有关的地物的正确性, 如果发现原地形图有错, 就利用勘丈值修改。在已有地形图成图到制作地籍图期间新建的建筑物, 图上没有反映, 应予补测。采用解析法、部分解析法或图解勘丈法测定界址点的地籍细部测量, 都可以利用地形图编绘地籍图。

地籍图的分幅与地形图的分幅相似, 分幅地籍图的幅面采用 $50\text{cm} \times 50\text{cm}$ 或 $50\text{cm} \times 40\text{cm}$ 。地籍图的分幅编号方法与地形图的分幅编号方法相似。地籍图分幅编号应以小比例尺图件为基础, 逐级编定较大比例尺的地籍图图幅号。当测区已有相应比例尺地形图时, 地籍图的分幅与编号方法可沿用地形图的分幅与编号, 并于编号后加注图幅内较大单位名称或以著名地理名称命名的图名。

一、地籍图的内容

地籍图是专题图, 内容主要有: 地籍要素、数学要素及地物要素。

(一) 地籍要素

在地籍图上应表示的地籍要素包括: 行政界线、界址点、界址线、地类号、地籍号、座落、土地使用者或所有者及土地等级等。

(二) 数学要素

在地籍图上应表示的数学要素包括: 大地坐标系、内外图廓线、坐标格网线及坐标注记、控制点点位及其注记、地籍图比例尺、地籍图分幅索引图、本幅地籍图分幅编号、图名及图幅整饰等内容。

(三) 地物要素

在地籍图上应表示的地物要素包括: 建筑物、道路、水系、地貌、土壤植被、注记等。

二、地籍图的分幅及精度要求

(一) 地籍图的分幅

图幅规格为 $40\text{cm} \times 50\text{cm}$ 的矩形图幅或 $50\text{cm} \times 50\text{cm}$ 的正方形图幅。基本地籍图分幅编号按图廓西南角坐标 (整 10m) 数编码, X 坐标在前, Y 坐标在后, 中间短线连接, 当勘丈区已有相应比例尺地形图时, 基本地籍图的分幅与编号方法亦可沿用地形图的分幅与编号。

(二) 地籍图的精度要求

地籍铅笔原图或地籍电子底图的基本精度要求: 相邻界址点间距、界址点与邻近地物点关系距离的中误差不得大于图上 $\pm 0.3\text{mm}$; 依勘丈数据装绘的上述距离的误差不得大于图上 $\pm 0.3\text{mm}$; 宗地内部与

界址边不相邻的地物点，不论采用何种方法勘丈，其点位中误差不得大于图上 $\pm 0.5\text{mm}$ ；邻近地物点间距中误差不得大于图上 $\pm 0.4\text{mm}$ 。

8.4 宗地图的制作

宗地图和分幅地籍图是宗地现状的直观描述。宗地图是以宗地为单位编绘的地籍图，分幅地籍图是以地图标准分幅为单位编绘的地籍图。宗地图与地籍图上的内容必须统一。

宗地图是土地证书的附图，通过具有法律手续的土地登记过程的认可，是土地所有者或使用者持有的具有法律效力的图件凭证，是处理土地权属问题时具有法律效力的图件。

一、宗地图内容

宗地图的主要内容：图幅号、地籍号；本宗地号、地类号、门牌号、面积及单位名称；本宗地界址点、界址点号（含与邻宗地共用的界址点）、界址线及界址边长；本宗地内建筑物、构筑物；邻宗地界址线（示意）；相邻宗地、道路、街巷及其名称；比例尺、指北方向、图廓线、制图单位、制图员、审核员及日期等。

二、宗地图绘制方法

宗地图绘制方法有蒙绘法、缩放绘制法、复制法、计算机输出法。

三、宗地图的要求

宗地图的要求是：宗地图一般用 32K、16K、8K 纸，宗地过大时，原则上可按分幅图整饰；宗地图必须依比例尺真实蒙绘；宗地图上界址边长必须注记齐全，界址边长注记采用实丈边长和解析反算边长均可；宗地图指北方向必须与相应的地籍图指北方向一致；宗地图的整饰、注记规格同地籍图。

8.5 面积量算

面积量算系指水平面积量算。面积量算的内容包括宗地面积、地类面积、宗地内建筑占地面积、建筑面积量算与面积汇总统计。

一、面积量算方法

面积量算方法一般分为解析法和图解法两种。

二、面积量算的要求

采用部分解析法进行地籍勘丈时，面积量算应采用坐标法计算每个街坊面积，用街坊面积数控制本街坊内各宗地面积之和。街坊内各宗地面积之和与街坊面积误差小于 $1/200$ 时，将误差按面积分配到各宗地，得出平差后的各宗地面积。完全用实测数据计算（量算）的宗地面积可以不参加平差。

采用图解法进行地籍勘丈时，面积量算应在聚酯薄膜原图上量算街坊面积，图面量算宜采用二级控制，首先以图幅理论面积为首级控制，图幅内各街坊及其他区块面积和图幅理论面积之差小于 $\pm 0.0025P$ （ P 为图幅理论面积）时，将闭合差按比例分配给各街坊及其他区块，得出平差后各街坊及各区块的面积。然后用平差后各街坊及各区块的面积去控制街坊内丈量的各宗地面积，其相对误差不得大于 $1/100$ ，在允许范围内将闭合差按比例分配给各宗地，得出平差后的宗地面积。全用实测数据计算的宗地面积可不参加平差。

第四部分 地图制图学基础

第九章 地图的基本概念

9.1 地图的定义和基本特性

地图是根据构成地图数学基础的数学法则和构成地理基础的综合法则将地球表面以符号系统和采用综合方法缩绘到平面上的表象，它反映各种自然和社会现象的空间分布、组合、联系及其在时间中的变化和发展，称之为地图。

地图一般具有以下三个基本特性：

- 1、特定的数学法则
- 2、特定的符号系统
- 3、制图综合的方法

9.2 地图的构成要素及分类

1. 地图的构成要素

构成地图的基本内容，叫做地图要素。它包括数学要素、地理要素和整饰要素（亦称辅助要素），所以又通称地图“三要素”。

（1）数学要素是决定图形分布位置和几何精度的数学基础，指构成地图的“骨架”。包括地图投影、坐标网、比例尺、控制点、高程系、地图分幅等。地图投影是用数学方法将地球椭球面上的图形转绘到平面上；坐标网是各种地图的数学基础，是地图上不可缺少的要素；比例尺表示坐标网和地图图形的缩小程度；大地控制点是保证将地球的自然表面转绘到椭球面上，再转绘到平面直角坐标网内时，具有精确的地理位置。高程系是描述平面地图上地貌要素的高度。

（2）地理要素，是指地图上表示的具有地理位置、分布特点的自然现象和社会现象。因此，又可分为自然要素（如江、河、地貌、土质、植被）和社会经济要素（如居民地、交通线、行政境界等），这是地图构成要素中的主体部分。

（3）整饰要素，也称为辅助要素。主要为图上补充辅助要素资料来丰富地图的内容。指便于读图和用图的某些内容。例如：图名、图号、图例和地图资料说明，以及图内各种文字、数字注记；绘制一些补充地图或剖面图、统计图；个别专题图还有一些表格或某一方面的重点文字说明。

2. 地图的分类

（1）按内容分类

按内容可将地图分为普通地图和专题地图两大类。

① 普通地图

普通地图基本上是以同等详细程度表示地表各种自然和社会现象的地图。即表达地图上的各种基本要素，如水系、地形、土质植被、居民地、交通网、境界和经济文化等要素。它比较全面地反映制图区域地自然环境、地区条件和人类改造自然地一般状况，也可以反映出自然、经济诸方面地相互联系和影响地基本规律。同时又是制作专题地图地底图。

普通地图按比例尺地大小，可分为地形图和地理图两类，大于 1：100 万（包括 1：100 万）比例尺的普通地图叫地形图，小于 1：100 万比例尺的普通地图叫地理图。

地形图:我国把 1:1 万、1:2.5 万、1:5 万、1:10 万、1:25 万、1:50 万、1:100 万七种比例尺的地形图规定为国家基本地形图,它们在内容、式样和规格上是一致的。这种地形图有严密的大地控制基础,采用统一投影、统一分幅编号,根据国家颁布的测绘规范和图式测制。

地理图:地理图是比例尺小于 1:100 万的地图。没有一定的系列,通常是根据用途需要而定,最常见的比例尺有:1:200 万、1:250 万、1:400 万、1:500 万、1:1000 万等。由于比例尺较小,包括的地区大,所以只能以高度概括的形式表示最主要的要素和区域的重要特征,其详细性和精确性都不如地形图。

②专题地图

专题地图是以普通地图做底图,着重表示某一种或几种专门内容的地图。专题地图的主题内容,可以是普通地图上所固有的一种或几种基本要素,也可以是专业部门特殊需要的内容。一般按地学标准分为自然地图和人文地图两大类。自然地图又可分为地质图、地貌图、水文图、气候图、土壤图、植被图、动物图等。人文地图可分为人口图、历史图、经济图、教育科学文化图等。

(2)按区域分类

首先是世界图,其次为大洲图和大洋图,在大洲(或大陆)内再按行政区或自然区划分。如按行政区划分时,则依国家或国内的一级行政区、二级行政区等等逐级划分。根据行政区划分类有较大意义,因为大多数地图都是按行政区划为限定范围进行编制的。另一种是按自然地理区域划分,如大洋图、流域图、海湾图、海峡图,或再细分如长江流域图,还可再分为三峡图、江汉平原图、长江下游平原图等等。

(3)按用途分类

地图按用途分类,就是给一定范围内的读者使用,解决他们的特定问题而提供的各种专门地图(如旅游地图、教学地图等)。一般说,许多地图都具有多方面的用途(如地形图),因此地图按用途分类在使用上受到一定限制。

地图按用途分类,可分军用地图和民用地图。军用图可以分为战术图战略图及各种军事专用地图;民用地图可以分为国民经济与管理地图、教育科学与文化地图。国民经济与管理地图可以进一步划分为自然资源及其评价地图、人口和劳动力资源及其评价地图、规划图、领航图和道路图等。教育科学与文化地图可进一步划分为教学地图、科学参考地图、文化宣传教育地图和旅游地图等。当然还可以再进一步细分,如教学地图可分为小学教学地图、中学教学地图和高等学校教学地图等。

第十章 地图的数学基础

10.1 地图比例尺

10.1.1 比例尺的定义

地图比例尺的定义:地图上某一长度与实地相应长度之比。

传统的概念仅适合于在大比例尺地图上使用。因此,完整、精确和**具有普遍意义的比例尺定义应该是:地图上沿某方向的微分线段和地面上相应微分线段水平长度之比。**

地图比例尺是一个比值,比例尺越大,图而精度越高;比例尺越小,图而精度越小,但概括性越强。当图幅大小相同时,比例尺越大,包括的地面范围越小;比例尺越小,包括的地面范围越大。在同一区域或同类的地图上,内容要素表示的详略程度和图形符号的大小,主要取决于地图比例尺。

10.1.2 地图比例尺的表现形式

比例尺有多种表示方法,如:数字比例尺,文字比例尺,直线比例尺等,我们最常见的是数字比例尺,数字比例尺是指用阿拉伯数字形式表示的比例尺。一般是用分子为 1 的分数形式表示,例如,1:10000(或 1/10000)其分母说明地图上线段已缩小了 10000 倍,在数字式比例尺的后面不应注有长度单位。其

中分子为图上距离，分母为实地距离。并且分子与分母必须用一个单位。

我国基本比例尺地形图有 1: 5 千、1: 1 万、1: 2.5 万、1: 5 万、1: 10 万、1: 25 万、1: 50 万和 1: 100 万八种。

普通地图通常按比例尺分为大、中、小三种：

大比例尺地形图：1: 5 千—1: 2.5 万比例尺地形图。

中比例尺地形图：1: 5 万—1: 25 万比例尺地形图。

小比例尺地形图：1: 50 万—1: 100 万比例尺地形图。

10.2 地图投影

10.2.1 地图投影概念

这种在球面与平面之间建立点与点之间对应函数关系的数学方法，称为地图投影。

10.2.2 地图投影的变形

1. 变形椭圆

地面上一个微分圆，面积小到可以忽略地球曲面的影响，作为平面看待，将它投影后变为椭圆，除个别为正圆外，一般皆为椭圆，这种椭圆被称为变形椭圆。

2. 长度变形

3. 面积变形

4. 角度变形

投影平面上任意两方向线所夹之角与球面上相应的两方向线夹角之差称为角度变形。

地图投影的变形，有角度变形、面积变形和长度变形。但不是所有投影都有这 3 种变形，等角投影就没有角度变形，等面积投影就没有面积变形，其他投影这 3 种变形都同时存在。

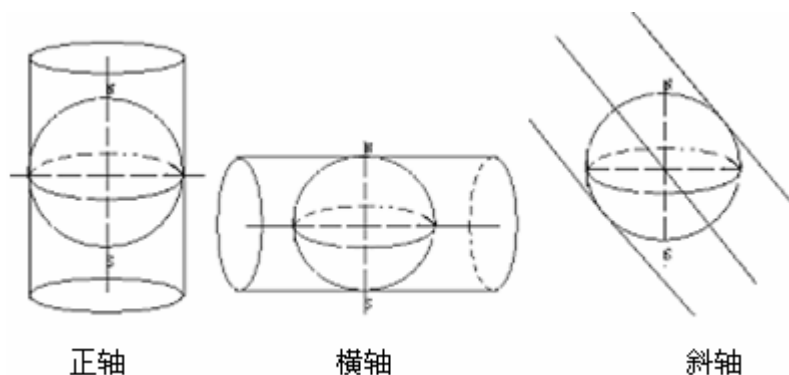
10.2.3 地图投影的分类

1. 按构成方法分类

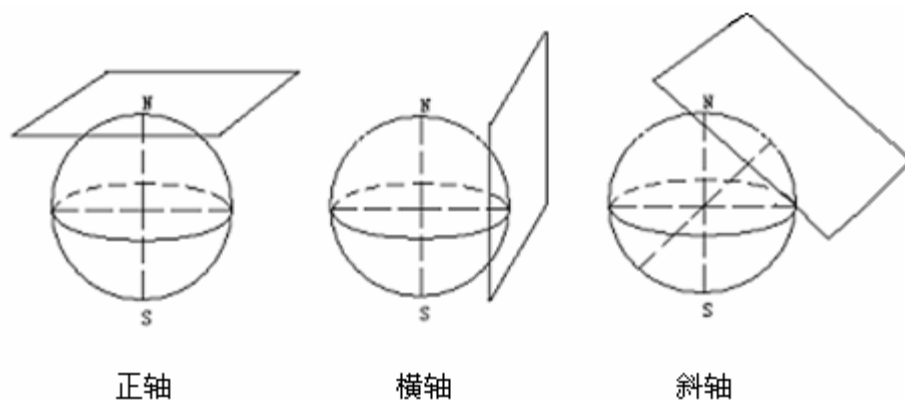
(1) 几何投影 圆柱投影、方位投影、圆锥投影

这种投影方法是把球面上的经纬线网投影到几何面上，然后将几何面展为平面而得到。根据几何面的形状，可以进一步分为下述几类：

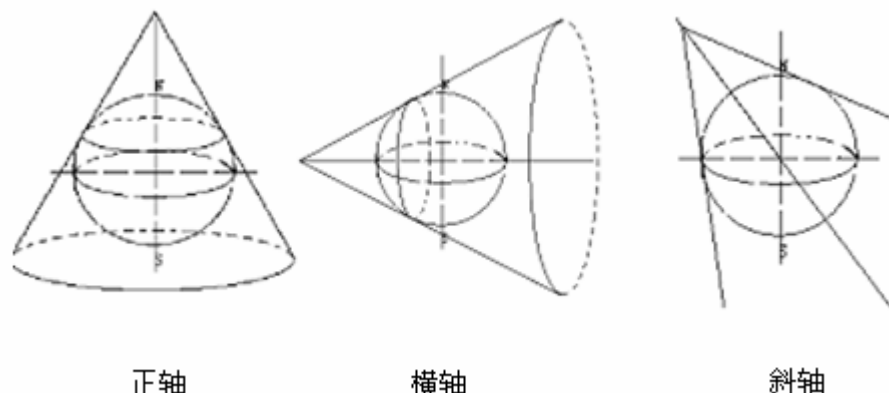
① 圆柱投影：以圆柱面作为投影面，使圆柱面与球面相切或相割，将球面上的经纬线投影到圆柱面上，然后将圆柱面展为平面而成。它可细分为正、横、斜轴圆柱投影。



② 方位投影：以平面作为投影面，使平面与球面相切或相割，将球面上的经纬线投影到平面上而成。它又可细分为正、横、斜轴方位投影，如下图所示：



③ 圆锥投影：以圆锥面作为投影面，使圆锥面与球面相切或相割，将球面上的经纬线投影到圆锥面上，然后将圆锥面展为平面而成。如下图所示：



(2) 非几何投影：不借助几何面，根据某些条件用数学解析法确定球面与平面之间点与点的函数关系。在这类投影中，一般按经纬线形状又分为下述几类：

① 伪方位投影：纬线为同心圆，中央经线为直线，其余的经线均为对称于中央经线的曲线，且相交于纬线的共同圆心。呈现辐射状

② 伪圆柱投影：纬线为平行直线，中央经线为直线，其余的经线均为对称于中央经线的曲线。

③ 伪圆锥投影：纬线为同心圆弧，中央经线为直线，其余经线均为对称于中央经线的曲线。

④ 多圆锥投影：纬线为同周圆弧，其圆心均为于中央经线上，中央经线为直线，其余的经线均为对称于中央经线的曲线。

2.按变形性质分类

(1) 等角投影：角度变形为零。同一方向长度比相同。

(2) 等积投影：面积变形为零。

(3) 等距投影：

等距投影的面积变形小于等角投影，角度变形小于等积投影。等积投影的形状变形比较大，等角投影的面积变形比较大。

3.按投影面与地球相割或相切分类

(1) 割投影

(2) 切投影

10.2.4 我国当前常用的地图投影

我国现行地图中，常采用的投影方式有：高斯克吕格投影、墨卡托投影、兰伯特投影。省区图多采用

墨卡托投影, 1: 100 万的地形图多采用兰伯特投影, 1: 1 万至 1: 50 万的地形图多采用高斯-克吕格投影。所以我们重点讲解高斯-克吕格投影。

1. 高斯-克吕格投影概念

高斯-克吕格投影是由高斯于 19 世纪 20 年代拟定, 后经克吕格补充而形成的一种地图投影方式, 它是一种等角横切圆柱投影, 简称高斯投影。

高斯-克吕格投影的中央经线和赤道为互相垂直的直线, 其他经线是均为凹向并对称于中央经线的曲线, 其他纬线是均为以赤道为对称轴的向两极弯曲的曲线, 经纬线成直角相交。

2. 高斯-克吕格投影的特征

- (1) 中央经线和地球赤道投影成为直线, 且为投影的对称轴;
- (2) 是等角投影, 经线和纬线投影后仍正交;
- (3) 中央经线长度比等于 1, 没有长度变形; 其余经线长度比均大于 1, 长度变形为正; 距中央经线越远, 变形越大;
- (4) 变形特征: 在同一条经线上, 最大变形在边缘经线与赤道的交点上, 但最大长度、面积变形分别仅为 +0.14% 和 +0.27% (6° 带), 变形极小。

3. 投影带的划分

为了保证地图的精度要求, 控制投影的变形不致过大, 高斯-克吕格投影采用分带投影的方法控制地图的变形。我国常采用 6° 分带法和 3° 分带法。我国规定 1: 1 万、1: 2.5 万、1: 5 万、1: 10 万、1: 25 万、1: 50 万比例尺地形图, 均采用高斯克吕格投影。1: 2.5 万至 1: 50 万的地形图, 采用经差 6° 分带方案, 全球共分为 60 个投影带; 我国位于东经 72° 到 136° 间, 共含 11 个投影带, 即 13 至 23 之间; 大于或等于 1: 1 万地图采用经差 3° 分带法。

6 度分带是从 0 度子午线起, 自西向东每隔经差 6 度为一投影带, 全球分为 60 带, 各带的带号用自然序数 1, 2, 3, ...60 表示。即以西经 0-6 为第 1 带, 其中央经线为 3 度, 东经 6-12 为第 2 带, 其中央经线为 9 度, 其余类推。

3 度分带是从东经 1 度 30 分的经线开始, 每隔 3 度为一带, 全球划分为 120 个投影带。

在高斯-克吕格投影分带上, 规定以中央经线为 X 轴, 赤道为 Y 轴, 两轴的交点为坐标原点。

X 坐标值在赤道以北为正, 以南为负; Y 坐标值在中央经线以东为正, 以西为负。我国在北半球, X 坐标皆为正值。Y 坐标在中央经线以西为负值, 运用起来很不方便。为了避免 Y 坐标出现负值, 将各带的坐标纵轴西移 500 公里, 即将所有 Y 值都加 500 公里。

由于采用了分带方法, 各带的投影完全相同, 某一坐标值 (x, y), 在每一投影带中均有一个, 在全球则有 60 个同样的坐标值, 不能确切表示该点的位置。因此, 在 Y 值前, 需冠以带号, 这样的坐标称为通用坐标。

10.2.5 我国地图坐标系统

① 1954 年北京坐标系

1954 年北京坐标系是原点在苏联普尔科沃。该坐标系采用克拉索夫斯基椭球作为参考椭球, 高程系统采用正常高, 以 1956 年黄海平均海面为基准。其缺点是椭球体面与我国大地水准面产生地误差较大, 提供的大地点坐标未经整体平差, 各点位之间的兼容性较差, 影响了坐标系本身的精度。

② 1980 年西安坐标系

1980 年国家大地坐标系是我国 1978 年 4 月召开的“全国天文大地网平差会议”上决定建立我国新的坐标系。采用 1975 年第十届国际大地测量即地球物理联合会 (IUGG/LAG) 推荐的新椭球体参数, 以陕西省西安市泾阳县永乐镇某点为国家大地坐标原点。其主要优点在于: 椭球体参数精度高; 椭球体面与我国大地水准面能够较好地吻合; 全国整体平差, 提高了平差结果地精度坐标统一; 大地原点选择在我国中部, 缩短了推算大地坐标的路程, 减少了推算误差地积累。

③ 高程及高程系

高程：地面点到大地水准面的垂直高度，称为高程，也叫绝对高程。地面点到任意水准面的垂直高度，称为相对高程。

高程系：高程控制网的建立，必须规定一个统一的高程基准面。我国的高程基准原来采用“1956 年黄海高程系”，由于黄海平均海面发生了微小的变化，因此启用了新的高程系，即“1985 年国家高程基准”。在采用新的高程基准后，对已有地图的等高线高程的影响可忽略不计。

第十一章 地图编制

11.1 机助制图

机助制图是将地图资料数字化，利用计算机进行编辑处理，通过绘图仪绘出新编的地图方法。

1. 机助制图原理

机助制图是计算机辅助制图的简称。为了使地图要素能为计算机识别和处理，必须将连续的图形转换成离散的数据，把地图上连续的图形转换成离散数据的过程，就是数字化。因此我们将图面上的所有要素都分为点状、线状、面状要素，通过数控绘图机以地图图形形式输出，就是计算机制图的全过程。

2. 机助制图工艺过程

机助制图的基本工艺过程包括数据输入、信息处理和图形输出 3 个主要环节。

(1) 数据输入

数据输入方式主要有：数字化仪输入、扫描矢量化输入、GPS 输入等。

(2) 信息处理

指在数据输入和图形输出这两个阶段之间，计算机对输入数据进行各种加工的过程。

一方面对输入的数据进行编辑加工、数据检查纠正、数据更新；另一方面对地图数据统一坐标、比例尺、投影变换；数据压缩和内容分类、要素符号化等，以获取符合编图要求的制图文件，并将制图文件进一步加工为绘图仪可以识别的绘图文件。

(3) 图形输出

是将计算机处理后的数据转换为图形形式，再通过各种输出设备输出图形的过程。可以直接制作胶片，然后再到印刷厂制版，出版地图。

还可以采用彩色喷墨绘图仪直接喷绘出彩色地图。

第十二章 计算机地图制图概述

12.1 计算机地图制图系统的构成

计算机地图制图系统由四个部分构成。计算机硬件系统、软件系统、地图数据、应用人员与组织机构。

12.1.1 硬件系统

硬件为计算机地图制图各种装置的总称，主要由计算机，数据、图形、图象的输入设备和输出设备 3 各部分构成。

1. 计算机硬件控制器、运算器、存储器、输入、输出设备五部分。

2. 数字地图输入设备：键盘、鼠标、磁盘（优盘、光盘、移动硬盘等）、扫描仪数字化仪等。

3. 地图输出设备：数据显示输出、纸质输出设备：打印机、绘图仪、处理控制设备。

12.1.2 软件系统

软件：使用计算机指令和算法语言编写的各种地图制图数据处理程序的总称。软件的组成，体现了机助制图系统的功能和特点。

目前多数机助制图系统软件具有的基本功能为：

- (1) 地图资料采集、编辑、管理程序，用于图形数字化过程中的各种误差改正，格式变换，数据合并和删除，生成数据文件等；
- (2) 基本绘图操作程序，用于控制图形输出设备的起始、终止、绘图头的驱动、换笔、直线或圆弧插补等；
- (3) 绘制地图各种点、线、面符号程序，如曲线光滑程序、面状晕线程序等；
- (4) 等值线程序，具有使用离散或网格数据点进行插值等值点、追踪排队以及输出等值线及其注记的功能；
- (5) 透视立体图程序，具有通过透视变换和消除隐藏线的处理，直观立体地表现自然和社会现象的功能；
- (6) 地图的数学基础和图外整饰以及投影变换程序；
- (7) 图表和统计地图程序；
- (8) 图形显示编辑程序；
- (9) 各种计算分析程序，如数据压缩、分类分级、图形简化、制图指标计算等。

制图软件按上述各类程序的性质可分为基本软件、功能软件和应用软件。

- (1) 基本软件指实现地图自动编绘过程中必须调用的基本子程序，如基本绘图操作程序。
- (2) 功能软件是指具有绘制某种简单地图图形或完成制图过程中某一独立功能的程序，如面状晕线程序、投影变换程序等。由于经常调用，功能软件也多设计为子程序形式，并具有多种控制参数。
- (3) 应用软件是解决地图制作某一方面完整功能的程序，如绘制等值线程序、输出统计地图程序等，这类程序直接面向用户。

12.1.3 地图数据

地图数据是编制专题地图主要的数据来源，包括各种比例尺的普通地图和根据实际用途不同而编制各种专题地图。

普通地图按比例尺和内容详细程度不同，分为地形图和地理图两种类型。

地图数据可以是纸质地图、电子地图、数字地图数据。电子地图数据的获取主要采用数字化的方法和数据格式的转换。数字化方法有手扶跟踪数字化方法和扫描数字化方法。在编制专题地图时常用的是扫描数字化方法。

地图数据分为空间数据和非空间数据两种类型。

12.1.4 应用人员与组织机构

电子地图的编制需要大量人力、物力的投入，还需要有一个完善的机构管理。

12.2 地理信息系统的发展过程

12.2.1 国际方面

国外 GIS 的发展经历了开拓期（60 年代）、巩固发展期（70 年代）、大发展时期（80 年代）、应用普及时代（90 年代以后）几个阶段。

12.2.2 国内方面

我国 GIS 的发展经历了四个阶段：即起步(1970-1980)、准备(1980-1985)、发展(1985-1995)、产业化(1996 年以后)阶段。

12.3 常见 GIS 介绍

对于国外软件来说，由于 GIS 技术研究起步早，软件产品已经相当成熟。美国环境研究所（ESRI）的 ArcGIS（Arc/Info）、MIS 公司的 MapInfo 等都是有名的国外 GIS 软件。它们在全球占有较大的市场，知名度较高。

对于国内软件来说，虽然 GIS 研究自上世纪八十年代初才开始，但经过九十年代的快速发展，已经产生了一批具有自主知识产权的 GIS 基础软件。这些软件在功能上与国外软件的差距正在缩小，部分性能甚至已经超越了国外软件，因而具有更高的性价比。目前，国内最具影响力的 GIS 软件主要有中地公司的 MapGIS、吉奥公司的 GeoStar 和超图公司的 SuperMAP。

第五部分 地理信息系统原理

第十三章 地理信息系统概论

13.1 地理信息系统基本概念

地理信息系统简称 GIS。GIS 在不同的应用领域、不同的专业，对它的理解是不一样的，目前国内外对 GIS 有许多定义。通常对 GIS 定义有：

- a. 地理信息系统是用于采集、存储、管理、处理、检索、分析和表达地理空间数据的计算机系统。
- b. 地理信息系统是由计算机系统、地理数据和用户组成，通过对地理数据的集成、存储、检索、操作和分析，生成并输出各种地理信息。
- c. 地理信息系统是在计算机软件 and 硬件支持下，运用系统工程和信息科学的理论，科学管理和综合分析具有空间内涵的地理数据，以提供对规划、管理、决策和研究所需信息的空间信息系统。
- d. 地理信息系统是以地理空间数据库为基础，采用地理模型分析方法，适时提供多种空间的和动态的地理信息，为地理研究和地理决策服务的计算机技术系统。

地理信息系统从外部来看，它表现为计算机软硬件系统；而其内涵是由计算机程序和地理数据组织而成的地理空间信息模型，是一个逻辑缩小的、高度信息化的地理系统。

上述定义均比较科学地阐明了 GIS 的对象、功能和特点。总之，地理信息系统是在计算机软硬件支持下，以采集、存储、管理、检索、分析和描述空间物体的定位分布及与之相关的属性数据，并回答用户问题等为主要任务的计算机系统。

13.2 地理信息系统的特点及功能

13.2.1 地理信息系统的特点

1 空间可视化

1) 空间地物轮廓特征的可视化

无论是在屏幕上展示一幅可以无级缩放和信息查询的地图，还是展现一幅三维的地形模型，都使我们对现实世界空间关系的认识更为直观、具体。

2) 具有空间参照特点的地物专题属性信息的可视化

2 空间导向

利用地理信息系统，我们不仅可以纵览研究区域的全域，还可以利用缩放和漫游等 GIS 所提供的基本功能深入到我们更感兴趣的区域去研究。

地理信息系统的空间导向功能还可以从空间查询功能中得到体现。

3 空间思维

地理信息系统的空间思维，就是要利用 GIS 数据库中已经存储的信息，通过 GIS 的工具(例如缓冲区分析、叠置分析)，生成 GIS 空间数据库中并求存储的信息。

13.2.2 地理信息系统的功能

1 数据采集与输入

主要有图形数据输入、栅格数据输入、GPS 数据输入、属性数据输入。

2 数据编辑与更新

数据编辑主要包括图形编辑和属性编辑。属性编辑主要与数据库管理结合在一起完成，图形主要包括拓扑关系建立、图形编辑、图形整饰、图幅拼接、图形变换、投影变换、误差校正等功能。

3 数据存储与管理

数据存储，即将数据以某种格式记录在计算机内部或外部存储介质上。

4 数据显示与输出

数据显示是中间处理过程和最终结果的屏幕显示，通常以人机交互方式来选择显示的对象与形式。GIS 不仅可以输出全要素地图，还可以根据用户需要，分层输出各种专题图、各类统计图、图表及数据。

5 空间查询与空间分析

GIS 的面向用户的应用功能不仅仅表现在它能提供一些静态的查询、检索数据，更有意义的在于用户可以根据需要建立一个应用分析的模式，通过动态的分析，从而为评价、管理和决策服务。

根据属性数据（包括组合条件）检索图形、根据图形（定位）检索属性以及图元间相互关系进行检索。

6 拓扑空间查询

空间目标之间的拓扑关系有两类：一种是几何元素的结点、弧段和面块之间的关联关系，用以描述和表达几何元素间的拓扑关系；另一种是 GIS 中地物之间的空间拓扑关系，可以通过关联关系和位置关系隐含表达，用户需通过特殊的方法查询。

空间关系有：面与面的关系，如检索与某个面状地物相邻的所有多边形及属性；线与线的关系，如检索与某一主干河相关联的所有支流；点与点的关系，如检索到某点一定距离内的所有点状地物；线与面的关系，如检索某公路所经过的县市或某县市内的所有公路；点与线的关系，如某河流上的所有桥梁；点与面的关系，如检索某市所有银行分布点。

7 缓冲区分析

缓冲区用以确定围绕某地要素绘出的定宽地区，以满足一定的分析条件。点的缓冲区是个圆饼，线的缓冲区是个条带状，多边形的缓冲区则是个更大的相似多边形。缓冲区分析是 GIS 中基本的空间分析功能之一，尤其对于建立影响地带是必不可少的。如道路规划中建立缓冲区以确定道路两边若干距离内的土地利用性质。

8 叠加分析

叠加分析提供根据两幅或两幅以上图层在空间上比较地图要素和属性的能力，通常有合成叠加和统计叠加之分，前者是根据两组多边形边界的交点建立具有多重属性的多边形，后者则进行多边形范围的属性特征统计分析。合成叠加得到一张新的叠加图，产生了许多新多边形，每个多边形都具有两种以上的属性。统计叠加的目的是统计一种要素在另一种要素中的分布特征。

9 选址分析

根据区域地理环境的特点,综合考虑资源配置、市场潜力、交通条件、地形特征、环境影响等因素,在区域范围内选择最佳位置。

10 网络分析

建立交通网络、地下管线网络等的计算机模型,研究交通流量、进行交通规则、处理地下管线突发事件(爆管、断路)等应急处理。警务和医疗救护的路径优选、车辆导航等也是 GIS 网络分析应用的实例。

11 距离分析及相邻相接分析

距离分析提供了在地图上距离的功能,相邻分析确定哪些地图要素与其它要素相接触或相邻,而相接分析则结合距离和相邻分析两者的针对性,提供确定地图要素间邻近或邻接的功能。

12 等高线分析

等高线图是人们传统上观测地形的主要手段,可以从等高线上精确地获得地形的起伏程度,区域内各部分的高程等。

13 坡度坡向分析

在 DTM 中计算坡度和坡向,派生出坡度坡向图供地形分析(如日照分析、土地适宜性分析等)。

14 断面图分析

用户可以在断面图上考察该剖面地形的起伏并计算剖面面积,以便用于工程设计和工程量算。

15 地形表面面积和填挖方体积计算

利用 DTM 数据,可以比较容易地求出所需要地区的地形表面面积以及施工区域内填挖方的体积(土石方量)。

第十四章 空间数据结构与空间数据库

在地理信息系统的空间数据结构中,有栅格结构和矢量结构两种方式。栅格结构的编码方式主要有直接栅格编码、链码、游程长度编码、块码、四叉树码等;矢量结构主要有坐标序列编码、树状索引编码和二元拓扑编码等编码方法。空间数据结构的选择对 GIS 的设计和建立起着非常关键的作用。

14.1 地理实体及其描述

14.1.1 空间实体的描述

通常需要从如下方面对地理实体进行描述:

- 1) 编码:是唯一的,用于区别不同的实体。
- 2) 位置:通常用坐标值的形式(或其它方式)给出实体的空间位置。
- 3) 类型:指明该地理实体属于哪一种实体类型,或由哪些实体类型组成。
- 4) 行为:指明该地理实体可以具有哪些行为和功能。
- 5) 属性:指明该地理实体所对应的非空间信息,如道路的宽度、路面质量、车流量、交通规则等。
- 6) 说明:用于说明实体数据的来源、质量等相关的信息。
- 7) 关系:与其它实体的关系信息。

14.1.2 空间数据的特征

空间数据具有三个基本特征: 1) 属性特征 2) 空间特征 3) 时间特征

14.1.3 空间数据的类型

根据空间数据的特征,可以把空间数据归纳为三类: 1) 属性数据 2) 几何数据 3) 关系数据

此外，还有元数据，它是描述数据的数据。

14.1.4 空间数据结构

数据结构即数据组织的形式，是适合于计算机存贮、管理、处理的数据逻辑结构。

空间数据结构是空间数据在计算机中的具体组织方式，主要以矢量数据结构和栅格数据结构两种形式存储。

14.2 栅格数据结构

14.2.1 栅格数据结构特点

栅格结构是最简单最直观的空间数据结构，是指将地球表面划分为大小均匀紧密相邻的网格阵列，每个网格作为一个象元或像素，由行、列号定义，并包含一个代码，表示该像素的属性类型或量值。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。

在栅格结构中，点用一个栅格单元表示；线状地物则用沿线走向的一组相邻栅格单元表示，每个栅格单元最多只有两个相邻单元在线上；面或区域用记有区域属性的相邻栅格单元的集合表示，每个栅格单元可有多于两个的相邻单元同属一个区域。

栅格结构的显著特点是：属性明显，定位隐含，即数据直接记录属性本身，而所在位置则根据行列号转换为相应的坐标给出。这种结构容易实现，算法简单，且易于扩充、修改，也很直观。

栅格结构表示的地表是不连续的，是量化和近似离散的数据。在栅格结构中，地表被分成相互邻接、规则排列的矩形方块(特殊的情况下也可以是三角形或菱形、六边形等，每个地块与一个栅格单元相对应。栅格数据的比例尺就是栅格大小与地表相应单元大小之比。在许多栅格数据处理时，常假设栅格所表示的量化表面是连续的，以便使用某些连续函数。

14.2.2 栅格结构的建立

要建立一个栅格数据结构需要明确三个内容：数据来源（即获取数据的途径），栅格系统的确定和栅格单元代码方式的确定。

（一）栅格数据的获取途径

1、来自于遥感数据

通过遥感手段获得的数字图像就是一种栅格数据。它是遥感传感器在某个特定的时间、对一个区域地面景象的辐射和反射能量的扫描抽样，并按不同的光谱段分光并量化后，以数字形式记录下来的像素值序列。

2、来自于对图片的扫描

通过扫描仪对地图或其它图件的扫描，可把资料转换为栅格形式的数据。具体为：扫描仪扫描专题图的图像数据得到每个像元的（行、列、颜色（灰度）），定义颜色与属性对应表，用相应属性代替相应颜色，得到每个像元的（行、列、属性），再进行栅格编码、存贮，即得到该专题图的栅格数据。

3、由矢量数据转换而来

通过运用矢量数据栅格化技术，把矢量数据转换成栅格数据。这种情况通常是为了有利于GIS中的某些操作，如叠加分析等，或者是为了有利于输出。

4、由手工方法获取

在专题图上均匀划分网格，逐个网格地确定其属性代码的值，最后形成栅格数据文件。

（二）栅格系统的确定

栅格系统的确定包括栅格坐标系的确定和栅格单元尺寸的确定。

1、栅格坐标系的确定

坐标系的确定实质是坐标系原点和坐标轴的确定。

栅格系统的起始坐标应与国家基本比例尺地形图公里网的交点相一致，并分别采用公里网的纵横坐标轴作为栅格系统的坐标轴。

2、栅格单元的尺寸

栅格单元的尺寸确定的原则是应能有效地逼近空间对象的分布特征，又减少数据的冗余度。格网太大，忽略较小图斑，信息丢失。一般讲实体特征愈复杂，栅格尺寸越小，分辨率愈高，然而栅格数据量愈大，按分辨率的平方指数增加，计算机成本就越高，处理速度越慢。

具体可采用保证最小多边形的精度标准来确定尺寸的方法。

(三) 栅格单元代码（属性值）的方式的确定

为了保证数据的质量，当一个栅格单元内有多个可选属性值时，决定栅格代码时尽量保持地表的真实性，保证最大的信息容量，按一定方法来确定栅格属性值。

在下图 3-2 所示的一块矩形地表区域，内部含有 A、B、C 三种地物类型，O 点为中心点，将这个矩形区域近似地表示为栅格结构中的一个栅格单元时，可根据需要，采取如下的方式之一来决定栅格单元的代码。

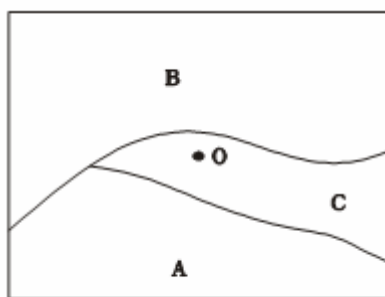


图 3-2 栅格单元代码的确定

①中心点法：用处于栅格中心处的地物类型或现象特性决定栅格代码。在下图所示的矩形区域中，中心点 O 落在代码为 C 的地物范围内，按中心点法的规则，该矩形区域相应的栅格单元代码应为 C，中心点法常用于具有连续分布特性的地理要素，如降雨量分布，人口密度图等。

②面积占优法：以占矩形区域面积最大的地物类型或现象特性决定栅格单元的代码。在右图所示的例中，显见 B 类地物所占面积最大，故相应栅格代码定为 B。面积占优法常用于分类较细，地物类别斑块较小的情况。

③重要性法：根据栅格内不同地物的重要性，选取最重要的地物类型决定相应的栅格单元代码、假设图 3-3 中 A 类为最重要的地物类型，即 A 比 B 和 C 类更为重要，则栅格单元的代码应为 A。重要性法常用于具有特殊意义而面积较小的地理要素，特别是点、线状地理要素，如城镇、交通枢纽、交通线、河流水系等，在栅格中代码应尽量表示这些重要地物。

④长度占优法：每个栅格单元的值由该栅格中线段最长的实体的属性来确定。

14.2.3 栅格数据的压缩编码

提高了栅格数据精度，就增加了数据量，数据冗余严重。为了解决这个难题，已发展了一系列栅格数据压缩编码方法，其类型又有信息无损编码和信息有损编码之分。

信息无损编码是指编码过程中没有任何信息损失，通过解码操作可以完全恢复原来的信息。

信息有损编码是指为了提高编码效率，最大限度地压缩数据，在压缩过程中损失一部分相对不太重要的信息，解码时这部分难以恢复。

常用的压缩编码方法有以下几种：

1、链式编码(Chain Codes)

链式编码又称为弗里曼链码(Freeman, 1961)或边界链码。链式编码主要是记录线状地物和面状地物的边界。它把线状地物和面状地物的边界表示为：由某一起始点开始并按某些基本方向确定的单位矢量链。基本方向可定义为：东=0，东南=1，南=2，西南=3，西=4，西北=5，北=6，东北=7等八个基本方

向（如图 3-4 所示）。

如果对于图 3-5 所示的线状地物确定其起始点为像元(1, 5)，则其链式编码为：

1, 5, 3, 3, 3, 3, 3, 3, 3

对于图 3-5 所示的面状地物，假设其原起始点定为像元（5, 8），则该多边形边界按顺时针方向的链式编码为：5, 8, 3, 2, 4, 4, 6, 6, 7, 6, 0, 2, 1

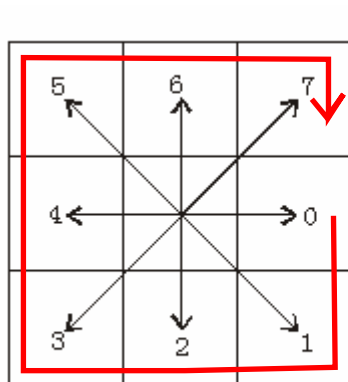


图 3-4 链式编码的方向代码

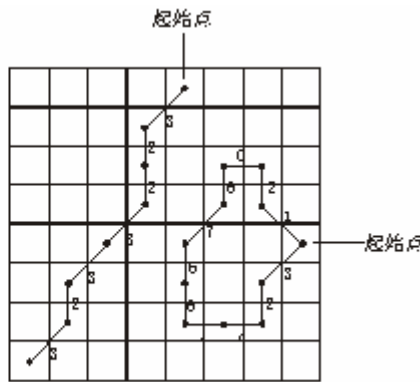


图 3-5 链式编码示意图

链式编码的前两个数字表示起点的行、列数，从第三个数字开始的每个数字表示单位矢量的方向，八个方向以 0—7 的整数代表。

链式编码对线状和多边形的表示具有很强的数据压缩能力，且具有一定的运算功能，如面积和周长计算等，探测边界急弯和凹进部分等都比较容易，类似矢量数据结构，比较适于存储图形数据。缺点是对叠置运算如组合、相交等则很难实施，对局部修改将改变整体结构，效率较低，而且由于链码以每个区域为单位存储边界，相邻区域的边界则被重复存储而产生冗余。

2、游程长度编码（run-length code）

游程长度编码是栅格数据压缩的重要编码方法，它的基本思路是：对于一幅栅格图像，常常有行(或列)方向上相邻的若干点具有相同的属性代码，因而可采取某种方法压缩那些重复的记录内容。其编码方案是，只在各行(或列)数据的代码发生变化时依次记录该代码以及相同代码重复的个数，从而实现数据的压缩。例如对图 3-6(a)所示的栅格数据，可沿行方向进行如下游程长度编码：

(9,4),(0,4),(9,3),(0,5),(0,1)(9,2),(0,1),(7,2),(0,2),(0,4),(7,2),(0,2),(0,4),(7,4),(0,4),(7,4),(0,4),(7,4),(0,4),(7,4)

游程长度编码对图 3-6(a)只用了 40 个整数就可以表示，而如果用前述的直接编码则需要 64 个整数表示，可见游程长度编码压缩数据是十分有效又简便的。事实上，压缩比的大小是与图的复杂程度成反比的，在变化多的部分，游程数就多，变化少的部分游程数就少，图件越简单，压缩效率就越高。

游程长度编码在栅格加密时，数据量没有明显增加，压缩效率较高，且易于检索，叠加合并等操作，运算简单，适用于机器存储容量小，数据需大量压缩，而又要避免复杂的编码解码运算增加处理和操作时间的情况。

3、块状编码(block code)

块码是游程长度编码扩展到二维的情况，采用方形区域作为记录单元，每个记录单元包括相邻的若干栅格，数据结构由初始位置(行、列号)和半径，再加上记录单元的代码组成。根据块状编码的原则，对图 3-6 (a) 所示图像可以用 12 个单位正方形，5 个 4 单位的正方形和 2 个 16 单位的正方形就能完整表示，具体编码如下：

(1,1,2,9),(1,3,1,9),(1,4,1,9),(1,5,2,0),(1,7,2,0),(2,3,1,9),(2,4,1,0),(3,1,1,0),(3,2,1,9),(3,3,1,9),(3,4,1,0),
(3,5,2,7), (3,7,2,0), (4,4,1,0),(4,2,1,0), (4,3,1,0), (4,4,1,0), (5,1,4,0), (5,5,4,7)

一个多边形所包含的正方形越大，多边形的边界越简单，块状编码的效率就越好。块状编码对大而简单的多边形更为有效，而对那些碎部较多的复杂多边形效果并不好。块状编码在合并、插入、检查延伸性、计算面积等操作时有明显的优越性。然而对某些运算不适应，必须在转换成简单数据形式才能顺利进

行。

4、四叉树编码(quad-tree code)

四叉树结构的基本思想是将一幅栅格地图或图像等分为四部分。逐块检查其格网属性值(或灰度)。如果某个子区的所有格网值都具有相同的值。则这个子区就不再继续分割, 否则还要把这个子区再分割成四个子区。这样依次地分割, 直到每个子块都只含有相同的属性值或灰度为止。

图 3-6 (b) 表示对图 3-6(a)的分割过程及其关系。

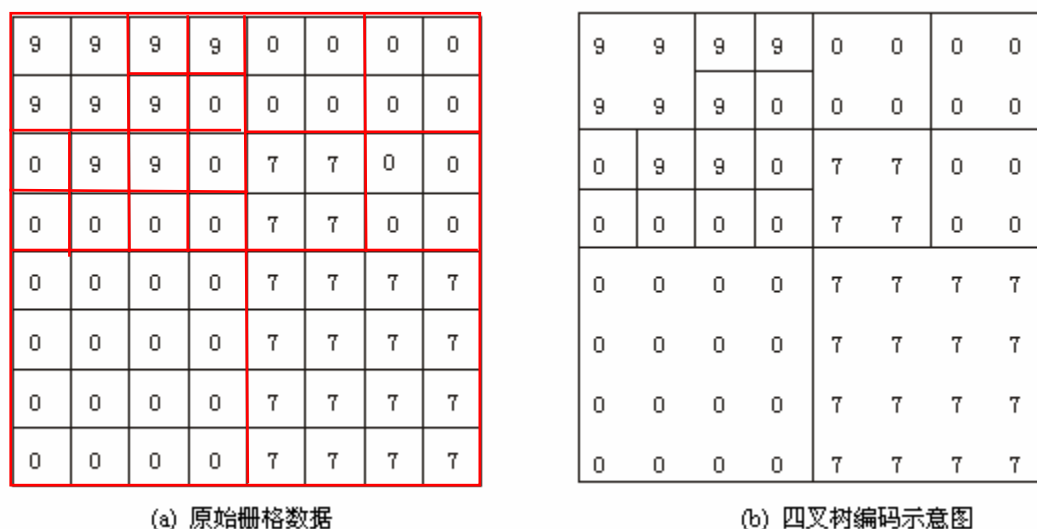


图 3-6 四叉树编码示意图

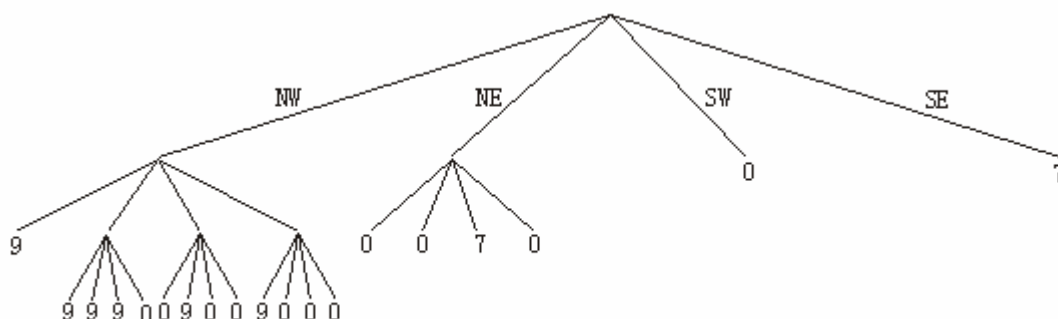


图 3-7 四叉树的树状表示

其中最上面的那个结点叫做根结点, 它对应整个图形。总共有 4 层结点, 每个结点对应一个象限, 如 2 层 4 个结点分别对应于整个图形的四个象限, 排列次序依次为按左上(NW)、右上(NE)、左下(SW), 右下(SE), 不能再分的结点称为终止结点(又称叶子结点), 可能落在不同的层上, 该结点代表的子象限具有单一的代码, 所有终止结点所代表的方形区域覆盖了整个图形。用一个树结构表示如图 3-7 所示。

由上面图形的四叉树分解可见, 四叉树中象限的尺寸是大小不一的, 位于较高层次的象限较大, 深度小即分解次数少, 而低层次上的象限较小, 深度大即分解次数多, 这反映了图上某些位置单一地物分布较广而另一些位置上的地物比较复杂, 变化较大。正是由于四叉树编码能够自动地依照图形变化而调整象限尺寸, 因此它具有极高的压缩效率。

采用四叉树编码时, 为了保证四叉树分解能不断地进行下去, 要求图像必须为 $2^n \times 2^n$ 的栅格阵列, n 为极限分割数, $n+1$ 为四叉树的最大高度或最大层数, 图 3-7 为 $2^3 \times 2^3$ 的栅格, 因此最多划分三次, 最大层数为 4, 对于非标准尺寸的图像需首先通过增加背景的方法将图像扩充为 $2^n \times 2^n$ 的图像。

四叉树编码具有可变的分辨率, 并且有区域性质, 压缩数据灵活, 许多运算可以在编码数据上直接实现, 大大地提高了运算效率, 是优秀的栅格压缩编码之一。

一般说来, 对数据的压缩是以增加运算时间为代价的。在这里时间与空间是一对矛盾, 为了更有效地利用空间资源, 减少数据冗余, 不得不花费更多的运算时间进行编码, 好的压缩编码方法就是要在尽可能

减少运算时间的基础上达到最大的数据压缩效率，并且是算法适应性强，易于实现。链码的压缩效率较高，已经近矢量结构，对边界的运算比较方便，但不具有区域的性质，区域运算困难；游程长度编码既可以在很大程度上压缩数据，又最大限度地保留了原始栅格结构，编码解码十分容易；块码和四叉树码具有区域性质，又具有可变的分辨率，有较高的压缩效率，四叉树编码可以直接进行大量图形图像运算，效率较高，是很有前途的方法。在此基础上已经开始发展了用于三维数据的八叉树编码等。

14.3 矢量数据结构

地理信息系统中另一种最常见的图形数据结构为矢量结构，即通过记录坐标的方式尽可能精确地表示点、线、多边形等地理实体。

14.3.1 矢量数据结构编码的基本内容

1、点实体

点实体包括由单独一对 x, y 坐标定位的一切地理或制图实体。在矢量数据结构中，除点实体的 x, y 坐标外还应存储其它一些与点实体有关的数据来描述点实体的类型、制图符号和显示要求等。

下图说明了点实体的矢量数据结构的一种组织方式。

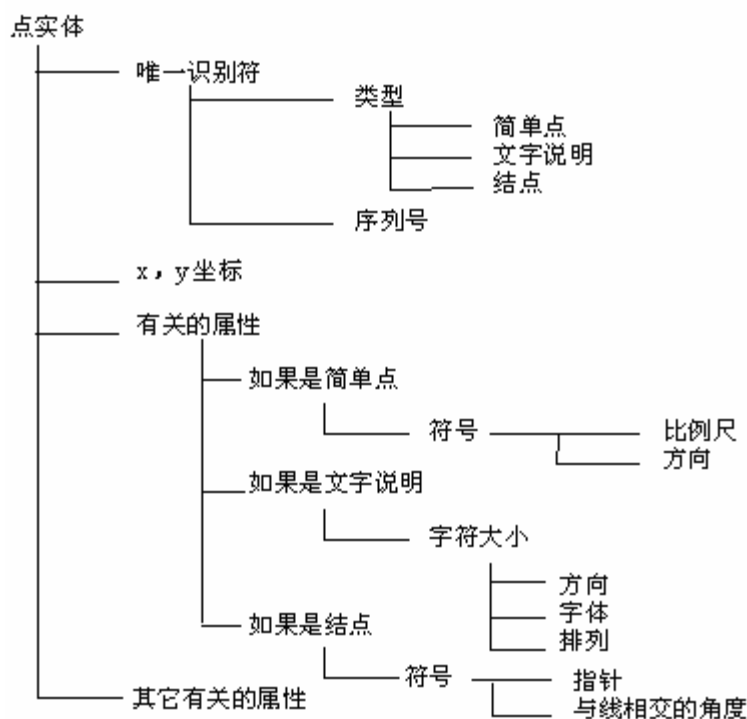


图 3-8 点实体的矢量数据结构

2、线实体

线实体由两对以上的 x, y 坐标定义。最简单的线实体只存储它的起止点坐标、属性、显示符等有关数据。例如，线实体输出时可能用实线或虚线描绘，这类信息属符号信息，它说明线实体的输出方式。虽然线实体并不是以虚线存储，仍可用虚线输出。

弧、链是 n 个坐标对的集合，这些坐标可以描述任何连续而又复杂的曲线。组成曲线的线元素越短， x, y 坐标数量越多，就越逼近于一条复杂曲线，既要节省存储空间，又要求较为精确地描绘曲线，唯一的办法是增加数据处理工作量。亦即在线实体的纪录中加入一个指示字，当起动显示程序时，这个指示字告诉程序：需要数学内插函数(例如样条函数)加密数据点且与原来的点匹配。于是能在输出设备上得到较精确的曲线。不过，数据内插工作却增加了。弧和链的存储记录中也要加入线的符号类型等信息。

线的网络结构：简单的线或链携带彼此互相连接的空间信息，而这种连接信息又是供排水网和道路网

分析中必不可少的信息。因此要在数据结构中建立指针系统才能让计算机在复杂的线网结构中逐线跟踪每一条线。指针的建立要以结点为基础。如建立水网中每条支流之间连接关系时必须使用这种指针系统。指针系统包括结点指向线的指针。每条从结点出发的线汇于结点处的角度等,从而完整地定义线网络的拓扑关系。

如上所述,线实体主要用来表示线状地物(公路、水系、山脊线)、符号线和多边形边界,有时也称为“弧”、“链”、“串”等,其矢量编码包括以下内容:

其中唯一标识是系统排列序号;线标识码可以标识线的类型;起始点和终止点可以用点号或直接用坐标表示;显示信息是显示线的文本或符号等;与线相联的非几何属性可以直接存储于线文件中,也可单独存储,而由标识码联接查找。

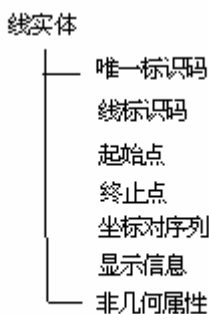


图 3-9 线实体矢量编码的基本内容

3、面实体

多边形(有时称为区)数据是描述地理空间信息的最重要的一类数据。在区实体中,具有名称属性和分类属性的,多用多边形表示,如行政区、土地类型、植被分布等;具有标量属性的有时也用等值线描述(如地形、降雨量等)。

多边形矢量编码,不但要表示位置和属性,更重要的是能表达区域的拓扑特征,如形状、邻域和层次结构等,以便使这些基本的空间单元可以作为专题图的资料进行显示和操作,由于要表达的信息十分丰富,基于多边形的运算多而复杂,因此多边形矢量编码比点和线实体的矢量编码要复杂得多,也更为重要。

在讨论多边形数据结构编码的时候,首先对多边形网提出如下的要求:

(1) 组成地图的每个多边形应有唯一的形状、周长和面积。

(2) 数据结构应能够记录每个多边形的邻域关系,其方法与水系网中记录连接关系一样。

(3) 专题地图上的多边形并不都是同一等级的多边形,而可能是多边形内嵌套小的多边形。例如,湖泊的水涯线在土地利用图上可算是个岛状多边形,而湖中的岛屿为“岛中之岛”。这种所谓“岛”或“洞”的结构是多边形关系中较难处理的问题。

14.3.2 矢量数据的获取方式

矢量数据的获取方式通常有:

- 1) 由外业测量获得,可利用测量仪器自动记录测量成果(常称为电子手簿),然后转到地理数据库中。
- 2) 由栅格数据转换获得,利用栅格数据矢量化技术,把栅格数据转换为矢量数据。
- 3) 跟踪数字化,用跟踪数字化的方法,把地图变成离散的矢量数据。

14.3.3 矢量数据的表示及数据结构

(一)、矢量数据表示

在 GIS 中,矢量数据表示时应考虑以下问题:

- 1) 矢量数据自身的存贮和处理。
- 2) 与属性数据的联系。
- 3) 矢量数据之间的空间关系(拓扑关系)。

(二)、简单数据结构

矢量数据的简单数据结构分别按点、线、面三种基本形式来描述。

1、标识码：按一定的原则编码，简单情况下可顺序编号。标识码具有唯一性，是联系矢量数据和与其对应的属性数据的关键字。

2、点结构中的 X, Y 坐标：是点实体的定位点，如果是有向点，则可以有两个坐标对。

3、线结构中的坐标对数 n：是构成该线(链)的坐标对的个数。X,Y 坐标串是构成线(链)的矢量坐标，共有 n 对。也可把所有线(链)的 X,Y 坐标串单独存放，这时只要给出指向该链坐标串的首地址指针即可。

4、面结构是链索引编码的面(多边形)的矢量数据结构，链数 n 指构成该面(多边形)的链的数目。链标识码集指所有构成该面(多边形)的链的标识码的集合，共有 n 个。

这种结构具有结构简单、直观、易实现以实体为单位的运算和显示的优点。由于面结构建立了链索引，一个面(多边形)就可由多条链构成，每条链的坐标可由线(链)的矢量数据结构获取。这种方法可保证多边形公共边的唯一性；但多边形的分解和合并不易进行；邻域处理比较复杂，需追踪出公共边；在处理“洞”或“岛”之类的多边形嵌套问题时较麻烦，需计算多边形的包含等。

由于拓扑关系简单，这种数据结构主要用于矢量数据的显示、输出，以及一般的查询和检索。

(三)拓扑数据结构

具有拓扑关系的矢量数据结构就是拓扑数据结构，拓扑数据结构是 GIS 的分析和应用功能所必需的。拓扑数据结构的表示方式没有固定的格式，还没有形成标准，但基本原理是相同的。

1、拓扑元素

矢量数据可抽象为点(结点)、线(链、弧段、边)、面(多边形)三种要素，即称为拓扑元素。

点(结点)：孤立点、线的端点、面的首尾点、链的连接点等。

线(链、弧段、边)：两结点间的有序弧段。

面(多边形)：若干条链构成的闭合多边形。

2、编码方式

拓扑数据结构的关键是拓扑关系的表示，而几何数据的表示可参照矢量数据的简单数据结构。在目前的 GIS 中，主要表示基本的拓扑关系，而且表示方法不尽相同。

14.3.4 矢量数据结构编码的方法

矢量数据结构的编码形式，按其功能和方法可分为：实体式、索引式、双重独立式和链状双重独立式。

1、实体式

实体式数据结构是指构成多边形边界的各个线段，以多边形为单元进行组织。按照这种数据结构，边界坐标数据和多边形单元实体一一对应，各个多边形边界都单独编码和数字化。例如对下图 3-10 所示的多边形 A、B、C、D、E，可以用下表 3-11 的数据来表示。

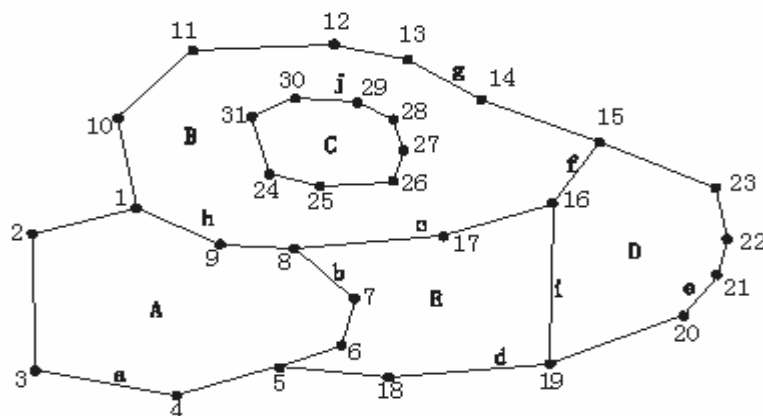


图 3-10 多边形原始数据

表 3-11 多边形数据文件

多边形	数 据 项
A	$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5), (x_6, y_6), (x_7, y_7), (x_8, y_8), (x_9, y_9), (x_1, y_1)$
B	$(x_1, y_1), (x_9, y_9), (x_{10}, y_{10}), (x_{11}, y_{11}), (x_{12}, y_{12}), (x_{13}, y_{13}), (x_{14}, y_{14}), (x_{15}, y_{15}), (x_{16}, y_{16}), (x_{17}, y_{17}), (x_{18}, y_{18}), (x_1, y_1)$
C	$(x_{24}, y_{24}), (x_{25}, y_{25}), (x_{26}, y_{26}), (x_{27}, y_{27}), (x_{28}, y_{28}), (x_{29}, y_{29}), (x_{30}, y_{30}), (x_{31}, y_{31}), (x_{24}, y_{24})$
D	$(x_{18}, y_{18}), (x_{20}, y_{20}), (x_{21}, y_{21}), (x_{22}, y_{22}), (x_{23}, y_{23}), (x_{25}, y_{25}), (x_{26}, y_{26}), (x_{18}, y_{18})$
E	

这种数据结构具有编码容易、数字化操作简单和数据编排直观等优点。但这种方法也有以下明显缺点：

- (1)相邻多边形的公共边界要数字化两遍，造成数据冗余存储，可能导致输出的公共边界出现间隙或重叠；
- (2)缺少多边形的邻域信息和图形的拓扑关系；
- (3)岛只作为一个单个图形，没有建立与外界多边形的联系。

因此，实体式编码只用在简单的系统中。

2、索引式

索引式数据结构采用树状索引以减少数据冗余并间接增加邻域信息，具体方法是对所有边界点进行数字化，将坐标对以顺序方式存储，由点索引与边界线号相联系，以线索引与各多边形相联系，形成树状索引结构。

树状索引结构消除了相邻多边形边界的数据冗余和不一致的问题，在简化过于复杂的边界线或合并多边形时可不必改造索引表，邻域信息和岛状信息可以通过对多边形文件的线索引处理得到，但是比较繁琐，因而给邻域函数运算、消除无用边、处理岛状信息以及检查拓扑关系等带来一定的困难，而且两个编码表都要以人工方式建立，工作量大且容易出错。

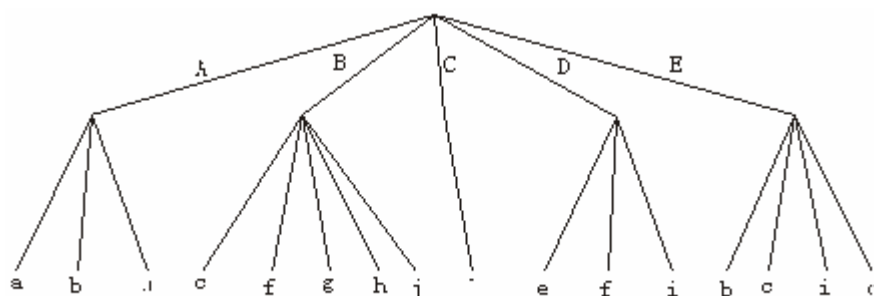


图 3-12 线与多边形之间的树状索引

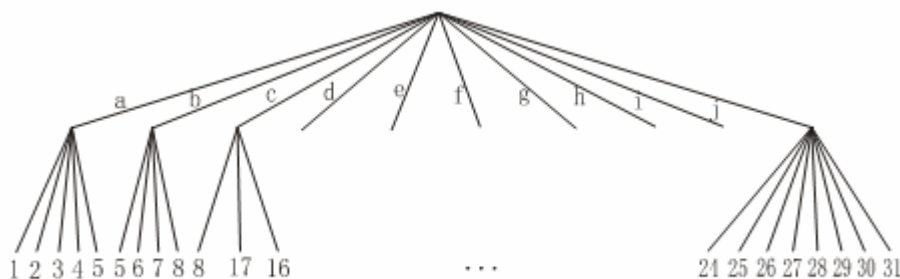


图 3-13 点与线之间的树状索引

3、双重独立式

这种数据结构最早是由美国人口统计局研制来进行人口普查分析和制图的，简称为 DIME(Dual Independent Map Encoding)系统或双重独立式的地图编码法。它以城市街道为编码的主体。其特点是采用

了拓扑编码结构。

双重独立式数据结构是对图上网状或面状要素的任何一条线段，用其两端的节点及相邻面域来予以定义。例如对下图 3-14 所示的多边形数据，用双重独立数据结构表示下表 3-15 所示：

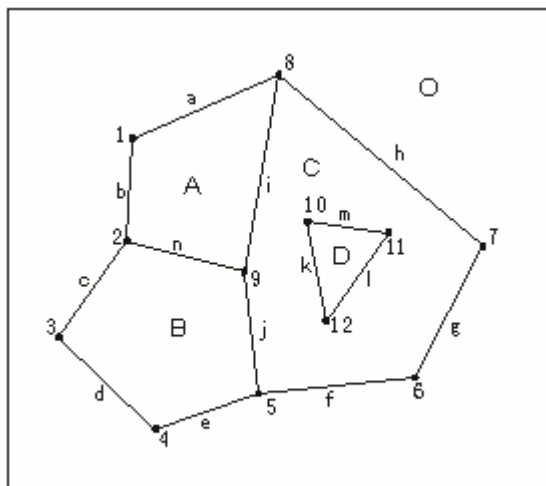


图 3-14 多边形原始数据

线号	左多边形	右多边形	起点	终点
a	0	A	1	8
b	0	A	2	1
c	0	B	3	2
d	0	B	4	3
e	0	B	5	4
f	0	C	6	5
g	0	C	7	6
h	0	C	8	7
i	C	A	8	9
j	C	B	9	5
k	C	D	12	10
l	C	D	11	12
m	C	D	10	11
n	B	A	9	2

表 3-15 双重独立式 (DIME) 编码

表中的第一行表示线段 a 的方向是从节点 1 到节点 8，其左侧面域为 0，右侧面域为 A。在双重独立式数据结构中，节点与节点或者面域与面域之间为邻接关系，节点与线段或者面域与线段之间为关联关系。这种邻接和关联的关系称为拓扑关系。利用这种拓扑关系来组织数据，可以有效地进行数据存储正确性检查，同时便于对数据进行更新和检索。因为在这种数据结构中，当编码数据经过计算机编辑处理以后，面域单元的第一个始节点应当和最后一个终节点相一致，而且当按照左侧面域或右侧面域来自动建立一个指定的区域单元时，其空间点的坐标应当自行闭合。如果不能自行闭合，或者出现多余的线段，则表示数据存储或编码有错，这样就达到数据自动编辑的目的。例如，从上表中寻找右多边形为 A 的记录，则可以得到组成 A 多边形的线及结点如下表，通过这种方法可以自动形成面文件，并可以检查线文件数据的正确性。

表 3-16 自动生成的多边形 A 的线及结点

线号	起点	终点	左多边形	右多边形
a	1	8	0	A
i	8	9	C	A
n	9	2	B	A
b	2	1	0	A

此外，这种数据结构除了通过线文件生成面文件外，还需要点文件，这里不在列出。

4、链状双重独立式

链状双重独立式数据结构是 DIME 数据结构的一种改进。在 DIME 中，一条边只能用直线两端点的序号及相邻的面域来表示，而在链状数据结构中，将若干直线段合为一个弧段（或链段），每个弧段可以有许多中间点。

在链状双重独立数据结构中，主要有四个文件：多边形文件、弧段文件、弧段坐标文件、结点文件。多边形文件主要由多边形记录组成，包括多边形号、组成多边形的弧段号以及周长、面积、中心点坐标及有关“洞”的信息等，多边形文件也可以通过软件自动检索各有关弧段生成，并同时计算出多边形的周长和面积以及中心点的坐标，当多边形中含有“洞”时则此“洞”的面积为负，并在总面积中减去，其组成的弧段号前也冠以负号；弧段文件主要有弧记录组成，存储弧段的起止结点号和弧段左右多边形号；弧段坐标文件由一系列点的位置坐标组成，一般从数字化过程获取，数字化的顺序确定了这条链段的方向。结点文件由结点记录组成，存储每个结点的结点号、结点坐标及与该结点连接的弧段。结点文件一般通过软件自动生成，因为在数字化的过程中，由于数字化操作的误差，各弧段在同一结点处的坐标不可能完全一致，需要进行匹配处理。当其偏差在允许范围内时，可取同名结点的坐标平均值。如果偏差过大，则弧段需要重新数字化。

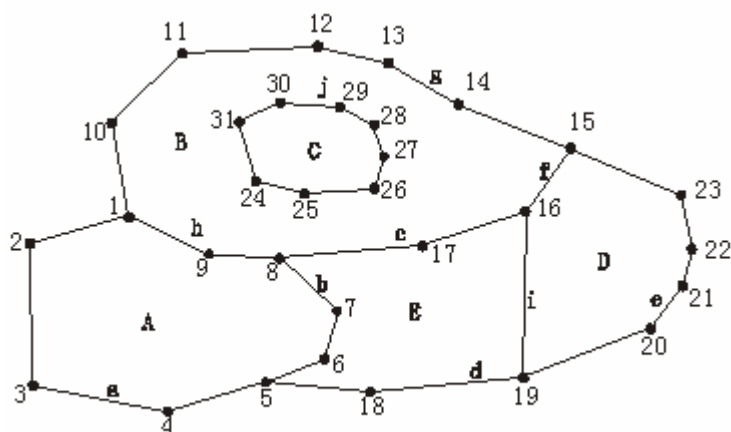


图 3-17 多边形原始数据

对如上图所示的矢量数据，其链状双重独立式数据结构的多边形文件、弧段文件、弧段坐标文件见以下三个表。

表 3-18 多边形文件

多边形号	弧段号	周长	面积	中心点坐标
A	h, b, a			
B	g, f, c, h, -j			
C	j			
D	e, i, f			
E	e, i, d, b			

表 3-19 弧段文件

弧段号	起始点	终结点	左多边形	右多边形
a	5	1	0	A
b	8	5	E	A
c	16	8	E	B
d	19	5	0	E
e	15	19	0	D
f	15	16	D	B
g	1	15	0	B
h	8	1	A	B
i	16	19	D	E
j	31	31	B	C

表 3-20 弧段坐标文件

弧段号	点 号
a	5,4,3,2,1
b	8,7,6,5
c	16,17,8
d	19,18,5
e	15,23,22,21,20,19
f	15,16,
g	1,10,11,12,13,14,15
h	8,9,1
i	16,19
j	31,30,29,28,27,26,25,24,31

链状双重独立式数据结构

14.4 两种数据结构的比较

栅格结构和矢量结构是模拟地理信息的两种不同的方法。栅格数据结构类型具有“属性明显、位置隐含”的特点，它易于实现，且操作简单，有利于基于栅格的空间信息模型的分析，如在给定区域内计算多边形面积、线密度，栅格结构可以很快算得结果，而采用矢量数据结构则麻烦的多；但栅格数据表达精度不高，数据存储量大，工作效率较低。如要提高一倍的表达精度（栅格单元减小一半），数据量就需增加三倍，同时也增加了数据的冗余。因此，对于基于栅格数据结构的应用来说，需要根据应用项目的自身特点及其精度要求来恰当地平衡栅格数据的表达精度和工作效率两者之间的关系。另外，因为栅格数据格式的简单性（不经过压缩编码），其数据格式容易为大多数程序设计人员和用户所理解，基于栅格数据基础之上的信息共享也较矢量数据容易。最后，遥感影像本身就是以象元为单位的栅格结构，所以，可以直接把遥感影像应用于栅格结构的地理信息系统中，也就是说栅格数据结构比较容易和遥感相结合。

表 3-21 栅格、矢量数据结构特点比较

比 较 内 容	矢 量 格 式	栅 格 格 式
数据量	小	大
图形精度	高	低
图形运算	复杂、高效	简单、低效
遥感影像格式	不一致	一致或接近

输出表示	抽象、昂贵	直观、便宜
数据共享	不易实现	容易实现
拓扑和网络分析	容易实现	不易实现

矢量数据结构类型具有“位置明显、属性隐含”的特点，它操作起来比较复杂，许多分析操作(如叠置分析等)用矢量数据结构难于实现；但它的数据表达精度较高，数据存储量小，输出图形美观且工作效率较高。两者的比较见上表。

14.5 空间数据管理

14.5.1 空间数据库的概念

地理信息系统中的数据库就是一种专门化的数据库，由于这类数据库具有明显的空间特征，所以有人把它称为空间数据库。

空间数据库（或称地图数据库）是地理信息系统的重要组成部分，因为地图是地理信息系统的主要载体。空间数据库用于存贮和管理地图信息；在数据处理系统中，它既是资料的提供者，也可以是处理结果的归宿处；在检索和输出过程中，它是形成绘图文件中或各类地理数据的数据源。

空间数据库是某一区域内关于一定地理要素特征的数据集合。

空间数据库与一般数据库相比，具有以下特点：

- 1) 数据量特别大，地理系统是一个复杂的综合体，要用数据来描述各种地理要素，尤其是要素的空间位置，其数据量往往很大。
- 2) 不仅有地理要素的属性数据（与一般数据库中的数据性质相似），还有大量的空间数据，即描述地理要素空间分布位置的数据，并且这两种数据之间具有不可分割的联系。
- 3) 数据应用广泛，例如地理研究、环境保护、土地利用与规划、资源开发、生态环境、市政管理、道路建设等。

14.5.2 数据库与数据库管理系统

数据库是存贮在计算机内的有结构的数据集合，是关于事物及其关系的信息组合。早期的数据库物体本身与其属性是分开存储的，只能满足简单的数据恢复和使用。数据定义使用特定的数据结构定义，利用文件形式存储，称之为文件处理系统。

文件处理系统是数据库管理最普遍的方法，但是有很多缺点：首先每个应用程序都必须直接访问所使用的数据文件，应用程序完全依赖于数据文件的存储结构，数据文件修改时应用程序也随之修改；另外的问题是数据文件的共享。由于若干用户或应用程序共享一个数据文件，要修改数据文件必须征得所有用户的认可。由于缺乏集中控制也会带来一系列数据库的安全问题。数据库的完整性是严格的，信息质量很差比没有信息更糟。

文件管理系统的优点：

- （1）数据文件是大量数据的集合形式。每个文件包含有大量的记录，每个记录包含若干个甚至多达几十个以上的数据项。文件和文件名面向用户并存储于计算机的存储设备上，可以反复利用。
- （2）面向用户的数据文件，用户可通过它进行查询、修改、插入、删除等操作。
- （3）数据文件与对应的程序具有一定的独立性，即程序员可以不关心数据的物理存储状态，只需考虑数据的逻辑存储结构，从而可以大量地节省修改和维护程序的工作量。
- （4）数据文件的缺点是只能对应于一个或几个应用程序，不能摆脱程序的依赖性。数据文件之间不能建立关系，呈现出无结构的信息集合状态，往往冗余度大，不易扩充，维护和修改。

数据库管理系统（Database Management System, DBMS）是在文件处理系统的基础上进一步发展的一个软件系统，用以维护数据库、接受并完成用户对数据库的一切操作。DBMS 在用户应用程序和数据文件之间起到了桥梁作用。DBMS 的最大优点是提供了两者之间的数据独立性，即应用程序访问数据文件时，

不必知道数据文件的物理存储结构。当数据文件的存储结构改变时，不必改变应用程序。

数据库管理系统的特点：

(1) 数据管理方式建立在复杂的数据结构设计的基础上，将相互关联的数据集—文件赋予某种固有的内在联系。各个相关文件可以通过公共数据项联系起来。

(2) 数据库中的数据完全独立，不仅是物理状态的独立，而且是逻辑结构的独立，即程序访问的数据只需提供数据项名称。

(3) 数据共享成为现实，数据库系统的并发功能保证了多个用户可以同时使用同一个数据文件，而且数据处于安全保护状态。

(4) 数据的完整性、有效性和相容性保证其冗余度最小，有利于数据的快速查询和维护

数据库系统指由硬件设备、软件系统、专业领域的数据库体和管理人员构成的一个运行系统。

14.5.3 数据库的主要特征

数据库方法与文件管理方法相比，具有更强的数据管理能力。数据库具有以下主要特征：

1. 数据集中控制特征

在文件管理方法中，文件是分散的，每个用户或每种处理都有各自的文件，不同的用户或处理的文件一般是没有联系的，因而就不能为多用户共享，也不能按照统一的方法来控制、维护和管理。数据库很好地克服了这一缺点，数据库集中控制和管理有关数据，以保证不同用户和应用可以共享数据。数据集中并不是把若干文件“拼凑”在一起，而是要把数据“集成”。因此，数据库的内容和结构必须合理，才能满足众多用户的要求。

2. 数据冗余度小的特征

数据的重复存储。在文件方式中，数据冗余太大。冗余数据的存在有两个缺点：一是增加了存储空间；二是易出现数据不一致。设计数据库的主要任务之一是识别冗余数据，并确定是否能够消除。在目前情况下，即使数据库方法也不能完全消除冗余数据。有时为了提高数据处理效率，也应该有一定程度的数据冗余。但是，在数据库中应该严格控制数据的冗余度。在有冗余的情况下，数据更新、修改时，必须保证数据库内容的一致性。

3. 独立性特征

数据独立是数据库的关键性要求。数据独立是指数据库中的数据与应用程序相互独立，即应用程序不因数据性质的改变而改变；数据的性质也不因应用程序的改变而改变。数据独立分为两级：物理级和逻辑级。物理独立是指数据的物理结构变化不影响数据的逻辑结构；逻辑独立意味着数据库的逻辑结构的改变不影响应用程序。但是，逻辑结构的改变必然影响到数据的物理结构。目前，数据逻辑独立还没有能完全实现。

4. 复杂的数据模型

数据模型能够表示现实世界中各种各样的数据组织以及数据间的联系。复杂的数据模型是实现数据集中控制、减少数据冗余的前提和保证。采用数据模型是数据库方法与文件方式的一个本质差别。数据库常用的数据模型有三种：层次模型，网络模型和关系模型。因此，根据使用的模型，可把数据库分成：层次型数据库，网络型数据库和关系型数据库。

5. 数据保护特征

数据保护对数据库来说是至关重要的，一旦数据库中的数据遭到破坏，就会影响数据库的功能，甚至使整个数据库失去作用、数据保护主要包括四个方面的内容：安全性控制、完整性控制、并发控制、故障的发现和恢复。

14.5.4 数据库的系统结构和数据组织方式

数据库是一个复杂的系统。数据库的基本结构可以分成三个层次：物理级、概念级和用户级。

1. 物理级：数据库最内的一层。它是物理设备上实际存储的数据集合(物理数据库)。它是由物理模式(也称内部模式)描述的。

2. 概念级：数据库的逻辑表示，包括每个数据的逻辑定义以及数据间的逻辑联系。它是由概念模式定义的，这一级也被称为概念模型。

3. 用户级：用户所使用的数据库，是一个或几个特定用户所使用的数据集(外部模型)，是概念模型的逻辑子集。它由外部模式定义。

数据组织方式：数据是现实世界中信息的载体，是信息的具体表达形式。为了表达有意义的信息内容，数据必须按照一定的方式进行组织和存储。数据库中的数据组织一般可以分为四级：数据项、记录、文件和数据库。

1. 数据项：

是可以定义数据的最小单位，也叫元素、基本项、字段等。数据项与现实世界实体的属性相对应，数据项有一定的取值范围，称为域。域以外的任何值对该数据项都是无意义的。如表示月份的数据项的域是1—12，13就是无意义的值。每个数据项都有一个名称，称为数据项目。数据项的值可以是数值的、字母的、汉字的等形式。数据项的物理特点在于它具有确定的物理长度，一般用字节数表示。

几个数据项可以组合，构成组合数据项。如“日期”可以由日、月、年三个数据项组合而成。组合数据项也有自己的名字，可以作为一个整体看待。

2. 记录：

由若干相关联的数据项组成。记录是应用程序输入—输出的逻辑单位。对大多数据库系统，记录是处理和存储信息的基本单位。记录是关于一个实体的数据总和，构成该记录的数据项表示实体的若干属性。

记录有“型”和“值”的区别。“型”是同类记录的框架，它定义记录，“值”是记录反映实体的内容。

为了唯一标识每个记录，就必须有记录标识符，也叫关键字。记录标识符一般由记录中的第一个数据项担任，唯一标识记录的关键字称主关键字，其它标识记录的关键字称为辅关键字。

3. 文件：

文件是一给定类型的(逻辑)记录的全部具体值的集合。文件用文件名称标识。文件根据记录的组织方式和存取方法可以分为：顺序文件、索引文件、直接文件和倒排文件等等。

4. 数据库：

是比文件更大的数据组织。数据库是具有特定联系的数据的集合，也可以看成是具有特定联系的多种类型的记录的集合。数据库的内部构造是文件的集合，这些文件之间存在某种联系，不能孤立存在。

14.5.5 数据间的逻辑联系

数据间的逻辑联系主要是指记录与记录之间的联系。记录是表示现实世界中的实体的，实体之间存在着一种或多种联系，这样的联系必然要反映到记录之间的联系上来。数据之间的逻辑联系主要有三种：

1、一对一的联系：简记为 1: 1。在 1: 1 的联系中，一个集合中的元素可以标识另一个集合中的元素。例如，地理名称与对应的空间位置之间的关系就是一种一对一的联系。

2、一对多的联系(1: N)：现实生活中以一对多的联系较多常见。行政区划就具有一对多的联系，一个省对应多个市，一个市有多个县，一个县又有多个乡。

3、多对多的联系(M: N)：这是现实中最复杂的联系 M: N 的联系，在数据库中往往不能直接表示出来，而必须经过某种变换，使其分解成两个 1: N 的联系来处理。地理实体中的多对多联系是很多的，例如土壤类型与种植的作物之间有多对多联系。同一种土壤类型可以种不同的作物，同一种作物又可种植在不同的土壤类型上。

14.6 数据库系统的数据模型

数据模型是数据库系统中关于数据和联系的逻辑组织的形式表示。每一个具体的数据库都是由一个相应的数据模型来定义。每一种数据模型都以不同的数据抽象与表示能力来反映客观事物，有其不同的处理数据联系的方式。

目前,数据库领域采用的数据模型有层次模型、网状模型和关系模型以及面向对象模型。其中应用最广泛的是关系模型。

14.6.1 层次模型

层次模型是数据处理中发展较早、技术上也比较成熟的一种数据模型。它的特点是将数据组织成有向有序的树结构。层次模型由处于不同层次的各个结点组成。除根结点外,其余各结点有且仅有一个上一层结点作为其“双亲”,而位于其下的较低一层的若干个结点作为其“子女”。结构中结点代表数据记录,连线描述位于不同结点数据间的从属关系(限定为一对多的关系)。

14.6.2 网状模型

网络数据模型是数据模型的另一种重要结构,它反映着显示世界中实体间更为复杂的联系,其基本特征是,结点数据间没有明确的从属关系,一个结点可与其它多个结点建立联系。不仅是双向的而且还可以是多对多的。

网络模型用连接指令或指针来确定数据间的显式连接关系,是具有多对多类型的数据组织方式,网络模型将数据组织成有向图结构。结构中结点代表数据记录,连线描述不同结点数据间的关系。

14.6.3 关系模型

在层次与网络模型中,实体间的联系主要是通过指针来实现的,即把有联系的实体用指针连接起来。而关系模型则采用完全不同的方法。

关系模型是根据数学概念建立的,它把数据的逻辑结构归结为满足一定条件的二维表形式。此处,实体本身的信息以及实体之间的联系均表现为二维表,这种表就称为关系。一个实体由若干个关系组成,而关系表的集合就构成为关系模型。

在生活中表示实体间联系的最自然的途径就是二维表格。表格是同类实体的各种属性的集合,在数学上把这种二维表格叫做关系。二维表的表头,即表格的格式是关系内容的框架,这种框架叫做模式,关系由许多同类的实体所组成,每个实体对应于表中的一行,叫做一个元组。表中的每一列表示同一属性,叫做域。

14.6.4 面向对象模型

面向对象的定义是指无论怎样复杂的事例都可以准确地由一个对象表示。每个对象都是包含了数据集和操作集的实体,即是说,面向对象的模型具有封装性的特点。

14.7 GIS 中空间数据库的组织方式

传统数据库系统管理地理空间数据有以下几个方面的局限性:

(1)传统数据库系统管理的是不连续的、相关性较小的数字和字符;而地理信息数据是连续的,并且具有很强的空间相关性。

(2)传统数据库系统管理的实体类型较少,并且实体类型之间通常只有简单、固定的空间关系;而地理空间数据的实体类型繁多,实体类型之间存在着复杂的空间关系,并且还能产生新的关系(如拓扑关系)。

(3)传统数据库系统存贮的数据通常为等长记录的数据;而地理空间数据通常由于不同空间目标的坐标串长度不定,具有变长记录,并且数据项也可能很大,很复杂。

(4)传统数据库系统只操纵和查询文字和数字信息;而空间数据库中需要有大量的空间数据操作和查询,如相邻、连通、包含、叠加等。

目前,大多数商品化的 GIS 软件都不是采取传统的某一种单一的数据模型,也不是抛弃传统的数据模

型, 而是采用建立在关系数据库管理系统(RDBMS)基础上的综合的数据模型, 归纳起来, 主要有以下三种:

一、混合结构模型 (Hybrid Model)

它的基本思想是用两个子系统分别存储和检索空间数据与属性数据, 其中属性数据存储常规的 RDBMS 中, 几何数据存储的空间数据管理系统中, 两个子系统之间使用一种标识符联系起来。在检索目标时必须同时询问两个子系统, 然后将它们的结果结合起来。

由于这种混合结构模型的一部分是建立在标准 RDBMS 之上, 故存储和检索数据比较有效、可靠。但因为使用两个存储子系统, 它们有各自的规则, 查询操作难以优化, 存储在 RDBMS 外面的数据有时会丢失数据项的语义; 此外, 数据完整性的约束条件有可能遭破坏, 例如在几何空间数据存储子系统中目标实体仍然存在, 但在 RDBMS 中却已被删除。属这种模型的 GIS 软件有 ARC / INFO、GENEMAP 等。

二、扩展结构模型 (Extended Model)

混合结构模型的缺陷是因为两个存储子系统具有各自的职责, 互相很难保证数据存储、操作的统一。扩展结构模型采用同一 DBMS 存储空间数据和属性数据。其做法是在标准的关系数据库上增加空间数据管理层, 即利用该层将地理结构查询语言(GeoSQL)转化成标准的 SQL 查询, 借助索引数据的辅助关系实施空间索引操作。这种模型的优点是省去了空间数据库和属性数据库之间的繁琐联结, 空间数据存取速度较快, 但由于是间接存取, 在效率上总是低于 DBMS 中所用的直接操作过程, 且查询过程复杂。

三、统一模数据型 (Integrated Model)

这种综合数据模型不是基于标准的 RDBMS, 而是在开放型 DBMS 基础上扩充空间数据表达功能。空间扩展完全包含在 DBMS 中, 用户可以使用自己的基本抽象数据类型(ADT)来扩充 DBMS。在核心 DBMS 中进行数据类型的直接操作很方便、有效, 并且用户还可以开发自己的空间存取算法。该模型的缺点是, 用户必须在 DBMS 环境中实施自己的数据类型, 对有些应用将相当复杂。

第十五章 数据的输入、编辑与输出

地理信息系统的数据来源非常广泛, 大致可以分为, 野外实测数据、航天航空遥感、全球定位系统(GPS)、原介质地图数据等。

15.1 GIS 数据输入

15.1.1 GIS 空间数据输入

1 野外数据的输入

GPS 数据输入可以直接读入文本数据; 全站仪数据输入利用测图软件系统直接接收及内存导入等方法。

2 原图数据输入

数据输入: 地理数据如何有效地输入到地理信息系统中是一项琐碎、费时、代价昂贵的任务, 大多数的地理数据是从纸质地图输入, GIS 常用的方法是数字化处理。

数字化的方法有两种: 数字化仪数字化; 扫描数字化。

2 数字化仪数字化是通过数字化输入子系统将底图资料数字化。采用方法是通过数字化仪由人工手扶游标跟踪, 将底图转化为图形数据。

2 扫描仪数字化的作用是将底图资料通过扫描仪上扫描后, 计算机获得栅格图形数据, 再通过矢量化系统将栅格数据转换成矢量数据。

扫描矢量化提供了矢量跟踪导向功能, 一般扫描矢量化软件都提供了交互式手动、半自动、全自动矢

量化方式供用户选择。

扫描矢量化输入法是目前地图输入的一种较有效的输入法。扫描矢量化子系统提供的扫描输入法是通过扫描仪直接扫描原图，以栅格形式存贮于图像文件中(如*.TIF 等)，然后经过矢量化将栅格数据转换成矢量数据，存入到文件中，再进行编辑、输出。

矢量化操作步骤如下：

(1) 图纸扫描

配置好扫描仪和计算机的连接之后就可以扫描图纸了，扫描的结果文件最好保存为无压缩的 tiff 格式，分辨率最好不低于 300dpi，并且图像底边尽量保持水平。

注意：如果扫描图像效果不好，宁可重新扫描，不可将就来校正和矢量化。

(2) 栅格图像校正

包括：非标准图幅与标准图幅校正的校正。

1) 非标准图幅校正方法（影像对影像校正与图框对影像校正）

2 影像对影像的校正

影像对影像的校正可以解决多幅影像拼接问题。处理方法是选择好要镶嵌的两幅影像，通过手工选取两幅影像图中相对应的控制点，至少必须添加 3 个控制点对。

添加完控制点后，点击预览查看影像匹配后的效果。

2 图框对影像的校正

矢量图框校正图像方法与图像校正图像的方法基本雷同。

矢量图框对影像的校正可以解决非标准比例尺图框对影像图校正套合与多幅影像拼接问题。处理方法是添加影像图及矢量图框，通过手工选取相对应的控制点，至少必须添加 4 个控制点对。

添加了控制点后，点击校正预览。

2) 标准图幅校正

标准图幅的校正通过读入图幅号信息，系统自动计算公里格网交点作为控制点文件。再根据图幅控制点文件逐一校正图像文件中的各个控制点。

校正完毕后得到 GeoTiff 文件，GeoTiff 是包含地理信息（坐标等信息）的一种 tiff 的格式文件。此时可将校正后的栅格数据与标准图幅矢量图框叠加显示。

(3) 读图分层

要素类分层的意义：图形矢量化以前，首先认真读图，对整个图形主要结构要有一个了解，然后根据一定的目的和分类指标，对底图上的图形要素进行分类。譬如，可按地理要素或地质要素分类，也可以按几何特征分类，或按用途与特征结合分类等。每一类作为一个图层（要素类），对每一个图层赋一个图层名（要素类名）。

对底图资料充分了解后，就可以拟一个矢量化工作提纲。它包括总的图层数，每一层的命名，每个数据文件的命名，工作的先后顺序，工作进度等等。

添加图层：栅格数据集层，添加要素类图层及注记类图层，这两个图层主要作为存储矢量化数据的图层；还可添加对象类图层、几何网络图层等。

(4) 制作图例板

图例版的制作可以直接提高作图人员的工作效率。根据系统提供制作图例板工具，将用户常用的几何图形符号分门别类设置好参数添加到同一个图例版中。使得用户在矢量化过程中只需选择图例板中预先设置好的图例即可，从而加快了用户选择几何图形的参数设置工作。

(5) 矢量化

使用矢量化工具能够将栅格数据方便的转化成矢量数据。

矢量化包括：交互式矢量化、全自动矢量化。

交互式矢量化(人工导向自动识别跟踪矢量化)：

对于那些干扰因素比较大如全要素图件，需要人工干预的图，要想追踪出比较理想的图，无条件全自动矢量化就显得力不从心了，此时人工导向自动识别跟踪矢量化正好解决这个问题。矢量化追踪的基本思

想就是沿着栅格数据线的中央跟踪，将其转化为矢量数据线。当进入到矢量化追踪状态后，即可以开始矢量跟踪，移动光标，选择需要追踪矢量化的线，屏幕上即显示出追踪的踪迹。每跟踪一段遇到交叉地方就会停下来，让你选择下一步跟踪的方向和路径。当一条线跟踪完毕后，即可以终止一条线，此时可以开始下一条线的跟踪。

全自动矢量化：它具有无需细化处理，处理速度快，不会出现细化过程中常见的毛刺现象，矢量化的精度高等特点。

全自动矢量化对于那些图面比较清洁，线条比较分明，干扰因素比较少的图，跟踪出来的效果比较好，但是对于那些干扰因素比较大的图（注释、标记特别多的图），就需要人工干预，才能追踪出比较理想的图。采用全自动矢量化对于图件来说最好是二值图（各要素分版清绘的图件）。

（6）保存文件

将矢量化得到的数据直接保存为文件存入计算机中。

3 电子数据转换

将已有的电子数据转换输入进 GIS 系统（如 autocad 数据导入到 mapgis 等），减轻信息收集、编码、输入的工作量。从计算机的角度来看难度虽不大，但在技术上需解决分类、编码、格式等标准化问题。从而实现了不同 GIS 系统间的数据转换，达到数据资源共享的目的。

15.1.2 GIS 属性数据输入

非空间关联属性有时称为特征编码或简单地称为属性，但它本身不属于空间数据类型。例如道路的非空间关联属性数据包括道路宽度、表面类型、建筑日期、水管、电线、特殊交通规则、每小时的车辆流量等，很显然这些数据都与道路这一空间实体相关，这些数据也可以有效地被存储和处理。若给予每一种数据一个公共识别符就可以有效地与空间数据（道路）连接起来。

属性数据的输入可在图形的适当位置键入。但数据量较大时一般都与空间数据分开输入且分别存储，将属性数据首先输入一个顺序文件，经编辑、检查无误后转存数据库的相应文件或表格。这是大数据量输入时的常用方法。

15.1.3 空间数据和非空间数据的连接

空间数据输入时虽然可以直接在图形实体上附加一个特征编码或识别符，但这样交互式地输入大量复杂地非空间数据，其效率就降低了，空间和非空间数据连接的最好方法是用特征编码或识别符把非空间属性数据与已数字化的点、线、面空间实体连接在一起。

非空间属性数据的数据项目最多，应把属于同一个实体的所有数据项放在同一个记录中，并将记录的序号或某一特征数据项作为该记录的识别符或关键字。它是空间与非空间数据的连接及相互检索的联系纽带。

15.2 GIS 数据编辑

15.2.1 图形数据编辑

空间和非空间数据数据输入时会产生一些误差，主要有：空间数据不完整或重复、空间数据位置不正确、空间数据变形、空间与非空间数据连接有误以及非空间数据不完整等。所以，在大多数情况下，空间和非空间数据输入以后，必须经过检核，然后才能进行交互式编辑。

一般进行图形数据交互式编辑有如下步骤：

- （1）读入数据
- （2）开窗显示图形数据检查错误
- （3）编辑修改

(4) 编辑工作完毕，存储数据

常用的编辑功能有：

- (1) 增加数据
- (2) 删除数据
- (3) 修改空间位置数据
- (4) 修改空间形状数据
- (5) 修改非空间数据：如图元颜色、点元符号、线元符号等。

15.2.2 属性数据编辑

属性数据是描述空间实体特征的数据集，这些数据主要用于描述实体要素类别、级别等分类特征和其他质量特征。

对于属性数据的输入与编辑，一般在属性数据处理模块中进行，但为了建立属性描述数据与几何图形的联系，通常需要在图形编辑系统中设计属性数据的编辑功能，主要是将一个实体的属性数据连接到相应的几何目标上，亦可在数字化及建立图形拓扑关系的同时或之后，对照一个几何目标直接输入属性数据。一个功能强的图形编辑系统可提供删除、修改、拷贝属性等功能。

15.3 GIS 数据输出

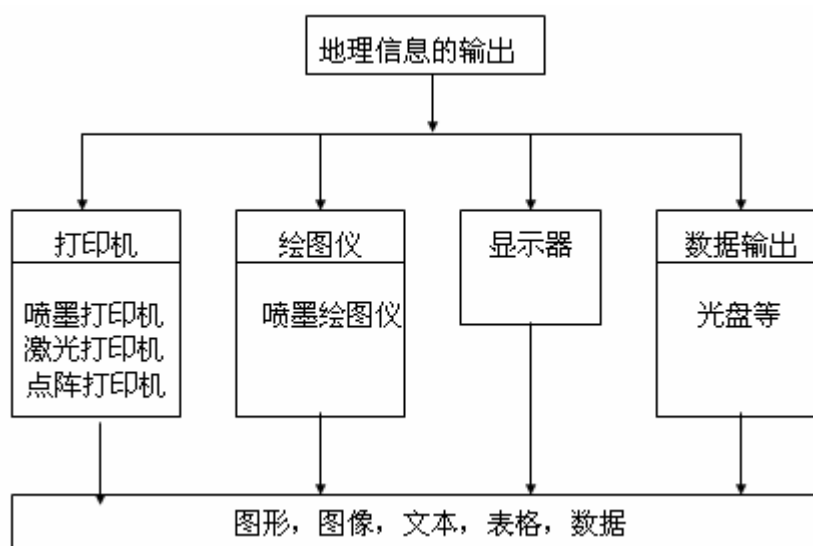
地理信息数据的输出除必须具备硬件设备以外，还必须有相应的软件支持。

15.3.1 输出的硬件设备

- 1 屏幕显示
- 2 打印输出
- 3 绘图仪输出

15.3.2 地理信息输出系统

地理信息输出系统是 GIS 平台软件不可缺少的重要模块。目前，地理信息系统软件都有输出图形、图像和属性数据报表功能，输出的方式主要有屏幕显示、打印输出、绘图仪输出和数据输出四种形式。GIS 常用的输出设备如下图所示。



15.3.3 地理信息系统产品的可视化输出

主要有以下四种类型：

- 1 地图
- 2 影像图
- 3 统计图表
- 4 晕渲图和三维场景

第十六章 空间分析原理与方法

空间分析：基于空间数据的分析技术，它以地学原理为依托，通过分析算法，从空间数据中获取有关地理对象的空间位置、空间分布、空间形态、空间形成、空间演变等信息。

地理信息系统的空间分析一般包括三大方面：空间分析、网络分析、DTM 分析。

16.1 空间分析

矢量数据空间分析的一般步骤是：

- 1) 明确分析的目的和评价准则；
- 2) 准备分析数据；
- 3) 进行空间分析操作；
- 4) 进行结果分析；
- 5) 解释、评价结果（如有必要，返回步骤 1）；
- 6) 结果输出（地图、表格和文档）。

矢量数据常用的空间分析基本原理与方法有以下几种：

16.1.1 叠加分析

地理信息系统的叠加分析是将有关主题层组成的数据层面，进行叠加产生一个新数据层面的操作，其结果综合了原来两层或多层要素所具有的属性。叠加分析不仅包含空间关系的比较，还包含属性关系的比较。地理信息系统叠加分析可分为以下几类：视觉信息叠加、点与多边形叠加、线与多边形叠加、多边形叠加等。

16.1.2 包含查询分析

包含分析确定要素之间是否存在直接的联系，即矢量点、线、面之间是否存在空间位置上的联系，这是地理信息分析处理中常要提出的问题，也是在地理信息系统中实现图形—属性对应检索的前提条件与基本的分析方法。例如，确定某区域内矿井的个数，这是点与面之间的包含分析，确定某一县境内公路的类型以及不同级别道路的里程，是线与面之间的包含分析。

16.1.3 矢量数据的缓冲区分析

缓冲区分析是研究根据数据库的点、线、面实体，自动建立其周围一定宽度范围内的缓冲区多边形实体。例如，城市的噪音污染源所影响的一定空间范围、交通线两侧所划定的绿化带，即可分别描述为点的缓冲区与线的缓冲带。而多边形面域的缓冲带有正缓冲区与负缓冲区之分。

这里举例说明如何利用建立缓冲区、拓扑叠加和特征提取，计算一条道路拓宽改建过程中的拆迁指标。具体分析步骤如下：

1) 明确分析的目的和标准

本例的目的是计算由于道路拓宽而需拆迁的建筑物的建筑面积和房产价值，道路拓宽改建的标准是：

- 1 道路从原有的 20m 拓宽至 60m；
- 1 拓宽道路应尽量保持直线；
- 1 部分位于拆迁区内的 10 层以上的建筑不拆除。

2) 准备进行分析的数据

本例需要涉及两类信息，一类是现状道路图；另一类为分析区域内建筑物分布图及相关信息。

3) 进行空间操作

首先选择拟拓宽的道路，根据拓宽半径，建立道路的缓冲区。

然后将此缓冲区与建筑物层数据进行拓扑叠加分析，产生一幅新图，此图包括所有部分或全部位于拓宽区内的建筑物信息。

4) 进行统计分析

首先对全部或部分位于拆迁区内的建筑物进行选择，凡部分落入拆迁区且楼层高于 10 层以上的建筑物，将其从选择组中去掉，并对道路的拓宽边界进行局部调整。

然后对所有需拆迁的建筑物进行拆迁指标计算。

5) 将分析结果以地图和表格的形式打印输出。

16.2 网络分析

1、基本概念

网络分析的主要用途是：选择最佳路径；选择最佳布局中心的位置。所谓最佳路径是指从始点到终点的最短距离或花费最少的路线；最佳布局中心位置是指各中心所覆盖范围内任一点到中心的距离最近或花费最小；网流量是指网络上从起点到终点的某个函数，如运输价格，运输时间等。网络上任意点都可以是起点或终点。

2、网络分析的基本方法

(1) 路径分析

- a. 静态求最佳路径：由用户确定网线权值关系后，即给定每条弧段的属性，当需求最佳路径时，读出路径的相关属性，求最佳路径。
- b. 动态分段技术：给定一条路径由多段联系组成，要求标注出这条路上的公里点或定位某一公路上的某一点，标注出某条路上从某一公里数到另一公里数的路段。
- c. N 条最佳路径分析：确定起点、终点，求代价较小的几条路径，因为在实践中往往仅求出最佳路径并不能满足要求，可能因为某种因素不走最佳路径，而走近似最佳路径。
- d. 最短路径：确定起点、终点和所要经过的中间点、中间连线，不考虑网线和转角的阻碍强度，求最短路径。
- e. 动态最佳路径分析：实际网络分析中权值是随着权值关系式变化的，而且可能会临时出现一些障碍点，所以往往需要动态地计算最佳路径。

(2) 地址匹配

地址匹配实质是对地理位置的查询，它涉及到地址的编码。地址匹配与其它网络分析功能结合起来，可以满足实际工作中非常复杂的分析要求。所需输入的数据，包括地址表和含地址范围的街道网络及待查询地址的属性值。

(3) 资源分配

资源分配网络模型由中心点(分配中心)及其状态属性和网络组成。分配有两种方式，一种是由分配中心向四周输出，另一种是由四周向中心集中。

资源分配模型可用来计算中心地的等时区、等交通距离区、等费用距离区等。可用来进行城镇中心、商业中心或港口等地的吸引范围分析，以用来寻找区域中最近的商业中心，进行各种区划和港口腹地的模

拟等。

16.3DTM 分析

DEM 和 DTM 主要用于描述地面起伏状况，可以用于提取各种地形参数，如坡度、坡向、粗糙度等，并进行通视分析、流域结构生成等应用分析。因此，DEM 在各个领域中被广泛使用。

DEM 可以有多种表达方法，包括网格、等高线、三角网等。

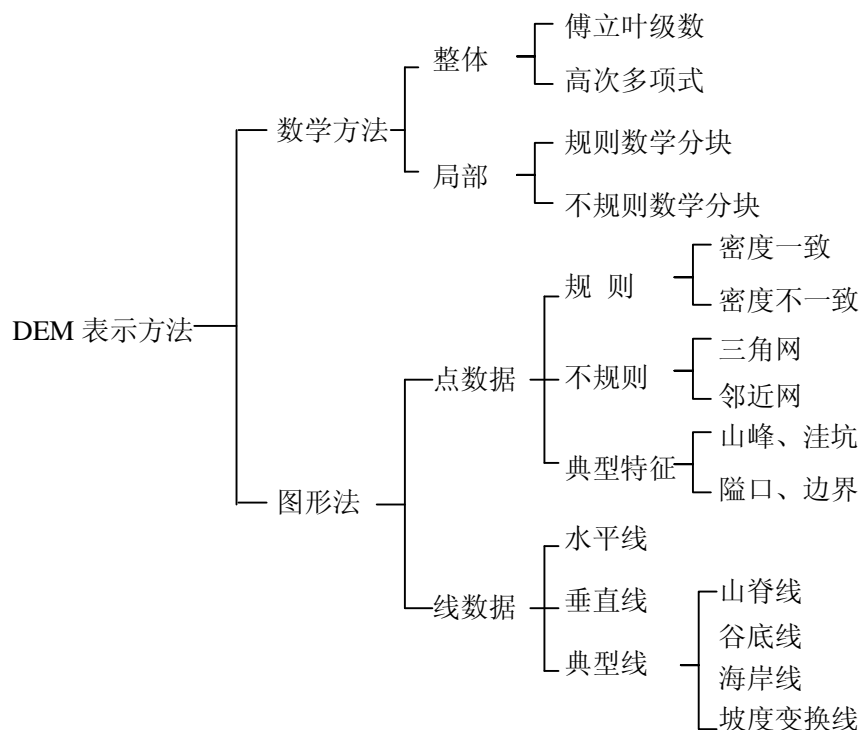
16.3.1 DTM 和 DEM

从数学的角度，高程模型是高程 z 关于平面坐标 x, y 两个自变量的连续函数，数字高程模型 (DEM) 只是它的一个有限的离散表示。高程模型最常见的表达是相对于海平面的海拔高度，或某个参考平面的相对高度，所以高程模型又叫地形模型。实际上地形模型不仅包含高程属性，还包含其它的地表形态属性，如坡度、坡向等。

数字地形模型是地表形态属性信息的数字表达，是带有空间位置特征和地形属性特征的数字描述。数字地形模型中地形属性为高程时称为数字高程模型 (Digital Elevation Model, 简称 DEM)。高程是地理空间中的第三维坐标。由于传统的地理信息系统的数据结构都是二维的，数字高程模型的建立是一个必要的补充。DEM 通常用地表规则网格单元构成的高程矩阵表示，广义的 DEM 还包括等高线、三角网等所有表达地面高程的数字表示。

16.3.2 DEM 的表示法

一个地区的地表高程的变化可以采用多种方法表达，用数学定义的表面或点、线、影像都可用来表示 DEM。



在地理信息系统中，DEM 最主要的三种表示模型是：规则格网模型，等高线模型和不规则三角网模型。



16.3.6 DEM 的分析和应用

- 1 地形曲面拟合
- 2 通视分析
- 3 流域特征地貌提取与地形自动分割
- 4 等高线追踪
- 5 DEM 计算地形属性：1) 坡度、坡向；2) 表面积

- ✓ 在数字地形图数据库中存贮高程数据；
- ✓ 解决道路设计和其他民用及军用工程中的一些与高程有关的问题；
- ✓ 三维地形显示及风景设计和规划；
- ✓ 剖面视觉分析；
- ✓ 道路规划、大坝选址等；
- ✓ 不同地形之间的静态分析和比较；
- ✓ 产生坡度图、坡向、及坡度剖面图，辅助地貌分析（淹没分析、土方计算等）或建立侵蚀图；工程量计算
- ✓ 作为专题信息的显示背景或将地形数据与专题数据如土壤、土地利用或植被等进行叠加；作为背景值和遥感影像等图层叠加，以制作各种专题图；
- ✓ 为景观的图像模拟和景观处理提供数据；
- ✓ 通过将高程替换为其他连续变化的属性，DEM能表示传播时间、费用、人口、污染程度、地下水深等信息

附录 参考问答题

简答题

1. 地图投影与 GIS 的关系？
2. 遥感技术在 3S 中的作用有哪些？
3. 我国 GIS 平台厂家有哪些，其产品品牌是？
4. GIS 元数据的基本概念与作用？
5. GIS 在电子政府建设中的应用？
6. 常见的国内外的数据库管理系统有哪些？
7. 简述目前国内外常用的 GIS 平台软件有哪些，分别是由哪些公司研发？
8. 什么是计算机病毒，计算机病毒有哪些特点？
9. GIS 平台与 CAD 的关系与区别？
10. 地理信息系统的组成与基本功能是什么？
11. 拓扑关系有什么作用？怎样生成拓扑关系？
12. 什么是地图投影，地图投影的意义？
13. RS 在国土资源调查中的作用用？
14. 我国当前常用的地图投影方式有那些？

综合题目

1. 计算机制图的特点及数字制图的一般流程？
2. 假如你是一名 GIS 相关专业的学生，请问你如何向你的家人或者朋友介绍你的专业，让他们明白 GIS 是什么，GIS 能做什么，GIS 的前景如何？
3. 如果你是一名国产 GIS 商业平台企业的市场人员，在某一次招标活动中，你应该怎样说服招标单位使用国产平台；如果你是一名国外 GIS 商业平台企业的市场人员，你又应该怎样说服招标单位使用国外平台？
4. 叙述一种有代表性的 GIS 商业化软件的基本情况和主要功能？
5. 阐述 3S 的结合使用，可用案例说明？
6. 说明 GIS 技术在我国的应用现状，并针对某一行业或领域深入阐述 GIS 的作用？
7. 结合某一 GIS 平台软件，叙述数字制图的基本流程？