



普通高等教育“十一五”国家级规划教材

新世纪计算机基础教育丛书

丛书主编 谭浩强

计算机软件技术基础

(第二版)

徐士良 葛兵 编著



清华大学出版社

本书特色

- 系统性。以数据结构与算法为基础，以软件技术为线索，介绍软件开发中所需的软件知识和技术。
- 应用性。内容围绕解决软件开发中所遇到的软件技术问题展开。在介绍系统软件（如操作系统）时，为了便于读者理解，也适当介绍一些原理，但主要还是介绍实现系统软件中的技术，以便读者将这些技术用到应用软件的开发中。
- 可读性。深入浅出，通过实例引出基本概念，便于教学。

ISBN 978-7-302-14953-8



9 787302 149538 >

定价：29.00元

TP31/188

2007

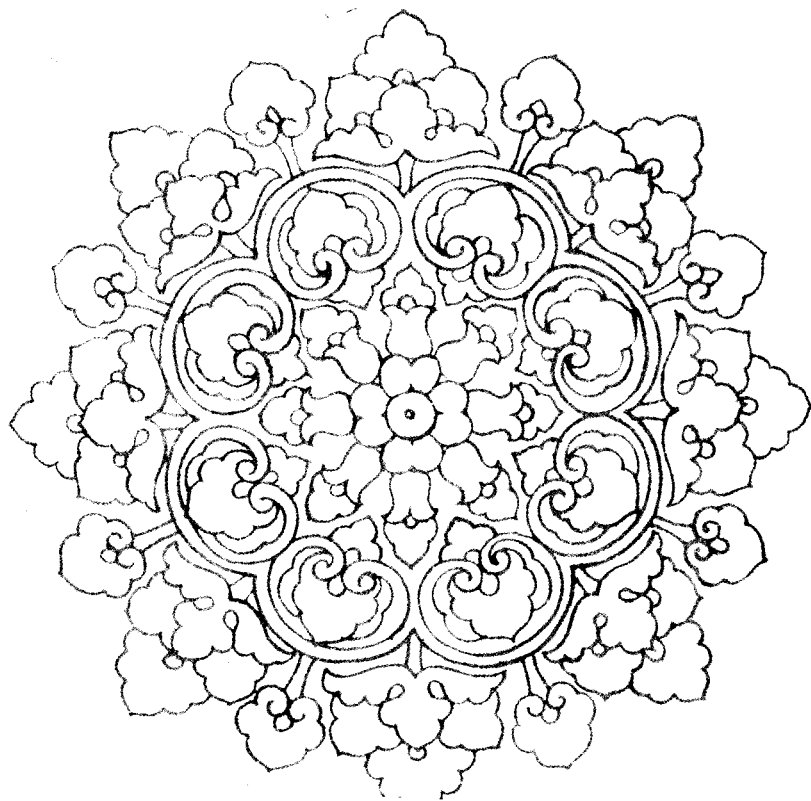
新世纪计算机基础教育丛书

丛书主编 谭浩强

计算机软件技术基础

(第二版)

徐士良 葛兵 编著



清华大学出版社
北京

内 容 简 介

本书是《计算机软件技术基础》的第二版,书中所有算法均采用 C++ 描述。全书共 6 章,介绍了计算机软件设计的基础知识、方法与实用技术。主要内容包括:算法,基本数据结构及其运算,查找与排序技术,资源管理技术,数据库技术,应用软件设计与开发技术。第二版主要修订了第一版的前 3 章,在每章后都配有一定数量的习题。

本书内容丰富,通俗易懂,实用性强,可作为非计算机专业的研究生、大学生的软件课程教材,也可作为广大从事计算机应用工作的科技人员的参考书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

计算机软件技术基础/徐士良,葛兵编著. —2 版. —北京:清华大学出版社,2007.7
(新世纪计算机基础教育丛书/谭浩强主编)

ISBN 978-7-302-14953-8

I. 计… II. ①徐… ②葛… III. 软件—高等学校—教材 IV. TP31

中国版本图书馆 CIP 数据核字(2007)第 044239 号

责任编辑:焦虹 薛阳

责任校对:时翠兰

责任印制:孟凡玉

出版发行:清华大学出版社

<http://www.tup.com.cn>

c-service@tup.tsinghua.edu.cn

社总机:010-62770175

投稿咨询:010-62772015

地 址:北京清华大学学研大厦 A 座

邮 编:100084

邮购热线:010-62786544

客户服务:010-62776969

印 刷 者:北京鑫海金澳胶印有限公司

装 订 者:三河市溧源装订厂

经 销:全国新华书店

开 本:185×260 印 张:23.5

字 数:539 千字

版 次:2007 年 7 月第 2 版

印 次:2007 年 7 月第 1 次印刷

印 数:1~5000

定 价:29.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:010-62770177 转 3103 产品编号:024025-01

现

代科学技术的飞速发展,改变了世界,也改变了人类的生活。作为新世纪的大学生,应当站在时代发展的前列,掌握现代科学技术知识,调整自己的知识结构和能力结构,以适应社会发展的要求。新世纪需要具有丰富的现代科学知识,能够独立完成面临的任务,充满活力,有创新意识的新型人才。

掌握计算机知识和应用,无疑是培养新型人才的一个重要环节。计算机技术已深入到人类生活的各个角落,与其他学科紧密结合,成为推动各学科飞速发展的有力的催化剂。无论学什么专业的学生,都必须具备计算机的基础知识和应用能力。计算机既是现代科学技术的结晶,又是大众化的工具。学习计算机知识,不仅能够掌握有关的知识,而且能培养人们的信息素养。它是高等学校全面素质教育中极为重要的一部分。

高校计算机基础教育应当遵循的理念是:面向应用需要,采用多种模式,启发自主学习,重视实践训练,加强创新意识,树立团队精神,培养信息素养。

计算机应用人才的队伍由两部分人组成:一部分是计算机专业出身的计算机专业人才,他们是计算机应用人才队伍中的骨干力量;另一部分是各行各业中应用计算机的人员。这后一部分人一般并非计算机专业毕业。他们人数众多,既熟悉自己所从事的专业,又掌握计算机的应用知识,善于用计算机作为工具解决本领域中的问题。他们是计算机应用人才队伍中的基本力量。事实上,大部分应用软件都是由非计算机专业出身的计算机应用人员研制的。他们具有的这个优势是其他人难以代替的。从这个事实可以看到在非计算机专业中深入进行计算机教育的必要性。

非计算机专业中的计算机教育,无论目的、内容、教学体系、教材、教学方法等各方面都与计算机专业有很大的不同,绝不能照搬计算机专业的教学模式和做法。全国高等院校计算机基础教育研究会自1984年成立以来,始终不渝地探索高校计算机基础教育的特点和规律。2004年,全国高等院校计算机基础教育研究会与清华大学出版社共同推出了《中国高等院校计算机基础教育课程体系2004》(简称CFC2004);2006年,又共同推出了《中国高等院校计算机基础教育课程体系2006》(简称CFC2006),由清华大学出版社正式出版发行。

1988年起,我们根据教学实际的需要,组织编写了《计算机基础教育丛书》,邀请有丰富教学经验的专家、学者先后编写了多种教材,由清华大学出版社出版。丛书出版后,迅速受到广大高校师生的欢迎,对高等学校

的计算机基础教育起到了积极的推动作用。广大读者反映这套教材定位准确,内容丰富,通俗易懂,符合大学生的特点。

1999年,根据新世纪的需要,我们又在原有基础上组织出版了《新世纪计算机基础教育丛书》。由于内容符合需要,质量较高,该套丛书已被许多高校选为教材。丛书总发行量达1000多万册,这在国内是罕见的。

最近,我们又对丛书做了进一步的修订,根据发展的需要,增加了新的书目和内容。本丛书有以下特点:

(1) 内容新颖。根据21世纪的需要,重新确定了丛书的内容,以符合计算机科学技术的发展和教学改革的要求。本丛书除保留了原丛书中经过实践考验且深受群众欢迎的优秀教材外,还编写了许多新的教材。在这些教材中反映了近年来迅速得到推广应用的一些计算机新技术,以后还将根据发展不断补充新的内容。

(2) 适合不同学校组织教学的需要。本丛书采用模块形式,提供了各种课程的教材,内容覆盖高校计算机基础教育的各个方面。丛书中既有理工类专业的教材,也有文科和经济类专业的教材;既有必修课的教材,也包括一些选修课的教材。各类学校都可以从中选择到合适的教材。

(3) 符合初学者的特点。本丛书针对初学者的特点,以应用为目的,以应用为出发点,强调实用性。本丛书的作者都是长期在第一线从事高校计算机基础教育的教师,对学生的基础、特点和认识规律有深入的研究,在教学实践中积累了丰富的经验。可以说,每一本教材都是他们长期教学经验的总结。在教材的写法上,既注意概念的严谨和清晰,又特别注意采用读者容易理解的方法阐明看似深奥难懂的问题,力求做到例题丰富,通俗易懂,便于自学。这一点是本丛书一个十分重要的特点。

(4) 采用多样化的形式。除了教材这一基本形式外,有些教材还配有习题解答和上机指导,并提供电子教案。

总之,本丛书的指导思想是内容新颖、概念清晰、实用性强、通俗易懂、教材配套。简单概括为:“新颖、清晰、实用、通俗、配套”。我们经过多年实践形成的这一套行之有效的创作风格,相信会受到广大读者的欢迎。

本丛书多年来得到了各方面人士的指导、支持和帮助,尤其是得到了全国高等院校计算机基础教育研究会各位专家和各高校老师们的支持和帮助,我们在此表示由衷的感谢。

本书肯定有不足之处,竭诚希望得到广大读者的批评指正。

欢迎访问谭浩强网站: <http://www.tanhoqiang.com>

丛书主编

全国高等院校计算机基础教育研究会会长

谭浩强

第二版前言

Forword Forword Forword Forword

非

计算机专业的学生(包括广大科技人员)学习和掌握软件技术不可能像计算机专业的学生那样学习软件的各门课程,因此有必要将主要的软件技术和知识在一门课程中介绍,但又不能是拼盘式的组合。在国外根本没有这样的教材,国内这样的教材也不多,有的教材虽然名称叫“软件技术”,但实际上是一些新软件的使用。本教材介绍的是软件技术,而不是软件的使用。

本教材主要是针对非计算机专业的学生对于计算机软件技术的需要,将有关软件基础知识以及应用技术介绍给读者。其特点如下:

(1) 系统性。本教材并不是简单地将计算机专业的各门课程内容拼装在一起,而是根据软件应用技术的需要,将它们有机地结合在一起,为读者提供软件开发中所需要的软件知识和技术。全书以数据结构与算法为基础,以软件技术为线索,系统性强。

(2) 强调应用。本教材强调以应用为目的,全书中实例比较丰富。全书几乎是围绕解决软件开发中所遇到的软件技术问题来展开的。在介绍系统软件(如操作系统)时,为了便于读者理解,也适当介绍一些原理,但主要还是介绍实现系统软件中的技术,以便读者将这些技术用到应用软件的开发中。

(3) 可读性强。本书深入浅出,通过例子引出基本概念,便于读者接受。

本书第二版主要修订的是第一版的前3章。书中所有的算法均采用C++描述。由于C++语言既可用于面向过程的程序设计,又支持面向对象的程序设计,因此,作者在对算法进行描述时,尽量采用最合适的程序设计方法。例如,对于基本的数据结构(如顺序存储与链式存储的线性表、栈、队列等)采用面向对象的方法,将数据与运算封装成类,以便在其他应用程序中直接使用;而对于同一批数据进行同类操作的各种算法(如线性表的排序方法)采用面向过程的方法,将各种不同的算法用普通函数来描述。这样,书中的所有算法就可以直接在实际应用中方便地使用。

本书内容丰富,通俗易懂,实用性强,书中的所有算法程序(C++描

述)均上机调试通过。本书可作为非计算机专业的研究生或大学生的软件课程教材,也可作为广大从事计算机应用工作的科技人员的参考书。

由于作者水平有限,书中难免有不妥之处,恳请读者批评指正。

作 者

2007 年 1 月

新世纪

第一版前言

随着计算机技术的深入发展,计算机技术的应用已经渗透到各个领域,特别是计算机软件的设计与开发,已经不只是计算机专业人员的事情了。现在,越来越多的软件需要非计算机专业人员来设计与开发,很多系统软件与应用软件由非计算机专业人员来使用,并在此基础上进行二次开发。因此,普及计算机软件技术已经是大势所趋。

本书是在《软件应用技术基础》(该书由清华大学出版社出版,获电子工业部优秀教材一等奖)一书的基础上改写而成,使本书更适合于广大非计算机专业学习软件设计与开发的需要。作为应用计算机的科技人员,除了要掌握现有计算机软件的使用外,从实际应用出发,还必须要掌握软件设计与开发的基本知识和有关技术,如数据的组织、程序的组织、计算机资源的利用、数据处理技术等,以便得心应手地进行应用软件的设计与开发。

全书共分6章,每章后面都附有一定数量的习题。

第1章介绍算法。主要包括算法的基本概念,算法的基本设计方法,算法的复杂度分析等内容。

第2章介绍基本数据结构及其运算。主要包括数据结构的基本概念,线性表、栈、队列及其在顺序存储结构下的运算和应用,线性链表及其运算,数组,二叉树的概念、存储及其应用,图的存储及其遍历。

第3章介绍常用的查找与排序技术。主要包括基本的查找技术,哈希表技术,基本的排序技术,二叉排序树及其查找,多层索引树及其查找,拓扑分类。

第4章介绍资源管理技术。主要包括操作系统的功能与任务,多道程序设计,存储空间的组织。

第5章介绍数据库技术。主要包括数据库基本概念,关系代数,数据库设计,关系数据库语言 SQL。

第6章介绍应用软件设计与开发技术。主要包括软件工程概述,软件详细设计的表达,结构化分析与设计方法,测试与调试基本技术,软件开发新技术。

本书内容丰富,通俗易懂,实用性强,书中所有算法程序均上机调试

通过。本书可作为非计算机专业的研究生或大学生的软件课程教材,也可作为广大从事计算机应用工作的科技人员的参考书。

由于作者水平有限,书中难免有错误或不妥之处,恳请读者批评指正。

作 者

2001 年 9 月

目 录

Catalog Catalog Catalog Catalog



算 法

1.1	算法的基本概念	1
1.1.1	算法的基本特征	1
1.1.2	算法的基本要素	2
1.2	算法设计基本方法	3
1.3	算法的复杂度分析	9
1.3.1	算法的时间复杂度	9
1.3.2	算法的空间复杂度	11
	习题	12



基本数据结构及其运算

2.1	数据结构的基本概念	13
2.1.1	两个例子	13
2.1.2	什么是数据结构	16
2.1.3	数据结构的图形表示	19
2.2	线性表及其顺序存储结构	21
2.2.1	线性表及其运算	21
2.2.2	栈及其应用	30
2.2.3	队列及其应用	41
2.3	线性链表及其运算	52
2.3.1	线性链表的基本概念	52
2.3.2	线性链表的基本运算	66
2.3.3	循环链表	71
2.3.4	多项式的表示与运算	75
2.4	数组	83
2.4.1	数组的顺序存储结构	83
2.4.2	规则矩阵的压缩	84

2.4.3	一般稀疏矩阵的表示	87
2.5	树与二叉树	113
2.5.1	树的基本概念	113
2.5.2	二叉树及其基本性质	116
2.5.3	二叉树的遍历	119
2.5.4	二叉树的存储结构	120
2.5.5	穿线二叉树	126
2.5.6	表达式的线性化	138
2.6	图	140
2.6.1	图的基本概念	140
2.6.2	图的存储结构	141
2.6.3	图的遍历	145
2.6.4	图邻接表类	146
	习题	152



查找与排序技术

3.1	基本的查找技术	155
3.1.1	顺序查找	155
3.1.2	有序表的对分查找	155
3.1.3	分块查找	160
3.2	哈希表技术	161
3.2.1	哈希表的基本概念	161
3.2.2	几种常用的哈希表	164
3.3	基本的排序技术	182
3.3.1	冒泡排序与快速排序	183
3.3.2	简单插入排序与希尔排序	188
3.3.3	简单选择排序与堆排序	192
3.3.4	其他排序方法简介	196
3.4	二叉排序树及其查找	200
3.4.1	二叉排序树的基本概念	201
3.4.2	二叉排序树的插入	202
3.4.3	二叉排序树的删除	204
3.4.4	二叉排序树查找	206
3.5	多层索引树及其查找	208

3.5.1 B ⁻ 树	209
3.5.2 B ⁺ 树	219
3.6 拓扑分类	229
习题	232



资源管理技术

4.1 操作系统的概念	234
4.1.1 操作系统的功能与任务	234
4.1.2 操作系统的发展过程	235
4.1.3 操作系统的分类	238
4.2 多道程序设计	241
4.2.1 并发程序设计	241
4.2.2 进程	244
4.2.3 进程之间的通信	249
4.2.4 多道程序的组织	254
4.3 存储空间的组织	255
4.3.1 内存储器的管理技术	255
4.3.2 外存储器中文件的组织结构	259
习题	266



数据库技术

5.1 数据库基本概念	267
5.1.1 数据库技术与数据库系统	267
5.1.2 数据描述	272
5.1.3 数据模型	275
5.2 关系代数	278
5.3 数据库设计	285
5.3.1 数据库设计的基本概念	285
5.3.2 数据库设计的过程	286
5.3.3 数据字典	293
5.4 关系数据库语言 SQL	294
5.4.1 SQL 语言概述	294

5.4.2	数据库定义	298
5.4.3	数据查询	303
5.4.4	数据修改	312
5.4.5	数据控制	315
5.4.6	嵌入式 SQL	318
习题		323



应用软件设计与开发技术

6.1	软件工程概述	325
6.1.1	软件工程的观念	325
6.1.2	软件生命周期	326
6.1.3	软件支援环境	328
6.2	软件详细设计的表达	329
6.2.1	程序流程图	330
6.2.2	NS 图	331
6.2.3	问题分析图 PAD	332
6.2.4	判定表	332
6.2.5	过程设计语言	333
6.3	结构化分析与设计方法	334
6.3.1	应用软件开发的原则和方法	334
6.3.2	结构化分析方法	336
6.3.3	结构化设计方法	340
6.4	测试与调试基本技术	346
6.4.1	测试	346
6.4.2	调试	353
6.5	软件开发新技术	355
6.5.1	原型方法	355
6.5.2	瀑布模型	356
6.5.3	面向对象技术	357
习题		358

参考文献	359
------------	-----

第1章 算 法

1.1 算法的基本概念

什么是算法呢？概括地说，算法是指解题方案的准确而完整的描述。

对于一个问题，如果可以通过一个计算机程序，在有限的存储空间内运行有限长的时间而得到正确的结果，则称这个问题是算法可解的。但算法不等于程序，也不等于计算方法。当然，程序也可以作为算法的一种描述，但程序通常还需考虑很多与方法和分析无关的细节问题，这是因为在编写程序时要受到计算机系统运行环境的限制。通常，程序的编制不可能优于算法的设计。

1.1.1 算法的基本特征

作为一个算法，一般应具有以下几个基本特征：

(1) 能行性(effectiveness)。算法的能行性包括以下两个方面：

① 算法中的每一个步骤必须能够实现。如在算法中不允许执行分母为0的操作，在实数范围内不能求一个负数的平方根等。

② 算法执行的结果要能够达到预期的目的。

针对实际问题设计的算法，人们总是希望能够得到满意的结果。但一个算法又总是在某个特定的计算工具上执行的，因此，算法在执行过程中往往要受到计算工具的限制，使执行结果产生偏差。例如，在进行数值计算时，如果某计算工具具有7位有效数字（如程序设计语言中的单精度运算），则在计算下列3个量

$$A=10^{12}, \quad B=1, \quad C=-10^{12}$$

的和时，如果采用不同的运算顺序，就会得到不同的结果，即

$$A+B+C=10^{12}+1+(-10^{12})=0$$

$$A+C+B=10^{12}+(-10^{12})+1=1$$

而在数学上， $A+B+C$ 与 $A+C+B$ 是完全等价的。因此，算法与计算公式是有差别的。在设计一个算法时，必须要考虑它的能行性，否则是不会得到满意结果的。

(2) 确定性(definiteness)。算法的确定性是指算法中的每一个步骤都必须是有明确定义的，不允许有模棱两可的解释，也不允许有多义性。这一性质也反映了算法与数学公式的明显差别。在解决实际问题时，可能会出现这样的情况：针对某种特殊问题，数学公式是正确的，但按此数学公式设计的计算过程可能会使计算机系统无所适从。这是因为根据数学公式设计的计算过程只考虑了正常使用情况，而当出现异常情况时，此计算过程就不能适应了。

(3) 有穷性(finiteness)。算法的有穷性是指算法必须能在有限的时间内做完，即算

法必须能在执行有限个步骤之后终止。数学中的无穷级数在实际计算时只能取有限项,即计算无穷级数值的过程只能是有穷的。因此,一个数的无穷级数表示只是一个计算公式,而根据精度要求确定的计算过程才是有穷的算法。

算法的有穷性还应包括合理的执行时间的含义。因为,如果一个算法需要执行千万年,显然失去了实用价值。

(4) 拥有足够的情报。一个算法是否有效还取决于为算法所提供的情报是否足够。通常,算法中的各种运算总是要施加到各个运算对象上,而这些运算对象又可能具有某种初始状态,这是算法执行的起点或是依据。因此,一个算法执行的结果总是与输入的初始数据有关,不同的输入将会有不同的结果输出。当输入不够或输入错误时,算法本身也就无法执行或导致执行有错。一般来说,当算法拥有足够的情报时,此算法才是有效的,而当提供的情报不够时,算法并不有效。

综上所述,所谓算法,是一组严谨地定义运算顺序的规则,并且每一个规则都是有效的,且是明确的,此顺序将在有限的次数下终止。

1.1.2 算法的基本要素

一个算法通常由两种基本要素组成:一是对数据对象的运算和操作,二是算法的控制结构。

(1) 算法中对数据的运算和操作。每个算法实际上是按解题要求从环境能进行的所有操作中选择合适的操作所组成的一组指令序列。因此,计算机算法就是计算机能处理的操作所组成的指令序列。

通常,计算机可以执行的基本操作是以指令的形式描述的。一个计算机系统能执行的所有指令的集合称为该计算机系统的指令系统。计算机程序就是按解题要求从计算机指令系统中选择合适的指令所组成的指令序列。在一般的计算机系统中,基本的运算和操作有以下4类:

- ① 算术运算,主要包括加、减、乘、除等运算。
- ② 逻辑运算,主要包括“与”、“或”、“非”等运算。
- ③ 关系运算,主要包括“大于”、“小于”、“等于”、“不等于”等运算。
- ④ 数据传输,主要包括赋值、输入、输出等操作。

前面提到,计算机程序也可以作为算法的一种描述,但由于在编制计算机程序时通常要考虑很多与方法和分析无关的细节问题(如语法规则),因此,在设计算法的一开始,通常并不直接用计算机程序来描述算法,而是用别的描述工具(如流程图、专门的算法描述语言,甚至用自然语言)来描述算法。但不管用哪种工具来描述算法,算法的设计一般都应从上述4种基本功能操作考虑,按解题要求从这些基本操作中选择合适的操作组成解题的操作序列。算法的主要特征着重于算法的动态执行,它区别于传统的着重于静态描述或按演绎方式求解问题的过程。传统的演绎数学是以公理系统为基础的,问题的求解过程是通过有限次推演来完成的,每次推演都将对问题作进一步的描述,如此不断推演,直到直接将解描述出来为止。而计算机算法则是使用一些最基本的操作,通过对已知条件一步一步的加工和变换,从而实现解目标。这两种方法的解题思路是不同的。

(2) 算法的控制结构。一个算法的功能不仅取决于所选用的操作,而且还与各操作之间的执行顺序有关。算法中各操作之间的执行顺序称为算法的控制结构。

算法的控制结构给出了算法的基本框架,它不仅决定了算法中各操作的执行顺序,而且也直接反映了算法的设计是否符合结构化原则。描述算法的工具通常有传统流程图、N-S 结构化流程图、算法描述语言等。一个算法一般都可以用顺序、选择、循环 3 种基本控制结构组合而成。

1.2 算法设计基本方法

本节介绍工程上常用的几种算法设计方法,在实际应用时,各种方法之间往往存在着一定的联系。

1. 列举法

列举法的基本思想是,根据提出的问题,列举所有可能的情况,并用问题中给定的条件检验哪些是需要的,哪些是不需要的。因此,列举法常用于解决“是否存在”或“有多少种可能”等类型的问题,例如求解不定方程的问题。

列举法的特点是算法比较简单。但当列举的可能情况较多时,执行列举算法的工作量将会很大。因此,在用列举法设计算法时,使方案优化,尽量减少运算工作量,是应该重点注意的。通常,在设计列举算法时,只要对实际问题进行详细的分析,将与问题有关的知识条理化、完备化、系统化,从中找出规律;或对所有可能的情况进行分类,引出一些有用的信息,是可以大大减少列举量的。

下面举例说明利用列举算法解决问题时如何对算法进行优化。

例 1.1 设每只母鸡值 3 元,每只公鸡值 2 元,两只小鸡值 1 元。现要用 100 元钱买 100 只鸡,设计买鸡方案。

假设买母鸡 i 只,公鸡 j 只,小鸡 k 只。根据题意,粗略的列举算法如下:

```
procedure baiji
for i=0 to 100 do
for j=0 to 100 do
for k=0 to 100 do
{ m=i+j+k
n=3i+2j+0.5k
if ((m=100)and(n=100)) then
output i,j,k
}
return
```

在这个算法中,共嵌套有 3 层循环,每层循环各需要循环 101 次,因此,总循环次数为 101^3 。但只要对问题进行分析,发现还可以对这个算法进行优化,减少大量不必要的循环次数。

首先,考虑到母鸡为 3 元一只,因此,母鸡最多只能买 33 只,即算法中的外循环没有必要从 0 到 100,而只需要从 0 到 33 就可以了。