



“十二五”普通高等教育本科国家级规划教材
普通高等教育精品教材



21世纪大学本科
计算机专业系列教材

蒋宗礼 姜守旭 编著

形式语言与自动机理论 (第3版)

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

- 根据教育部“高等学校计算机科学与技术专业规范”组织编写
- 与美国 ACM 和 IEEE CS *Computing Curricula* 最新进展同步
- 国家级精品教材

清华大学出版社

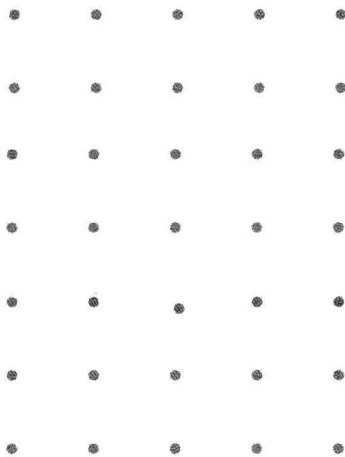


“十二五”普通高等教育本科国家级规划教材
普通高等教育精品教材

21世纪大学本科计算机专业系列教材

形式语言与自动机理论 (第3版)

蒋宗礼 姜守旭 编著



清华大学出版社

内 容 简 介

形式语言与自动机理论是计算机科学与技术专业的一门重要课程。本书是作者结合其近 30 年来在大学讲授该门课程的经验 and 体会,选择和组织有关内容撰写而成。基于计算机问题求解的需要讨论正则语言、上下文无关语言的文法、识别模型及其性质、图灵机的基本知识。其内容特点是抽象和形式化,既有严格的理论证明,又具有很强的构造性。叙述中特别注意引导读者分析与解决问题,以培养学生的形式化描述和抽象思维能力,使学生了解和初步掌握“问题、形式化、自动化(计算机化)”的解题思路。为了便于学生对内容的掌握,附录 A 还给出了建议的教学设计。

本书配套出版有《形式语言与自动机理论教学参考书(第 3 版)》,归纳各章知识点,解读主要内容,解析典型习题。

本书适合作为计算机学科研究生和高年级本科生的教材,也可供相关专业的学生、教师和科研人员参考。

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 13701121933

图书在版编目(CIP)数据

形式语言与自动机理论 / 蒋宗礼,姜守旭编著. —3 版. —北京:清华大学出版社,2013.5

21 世纪大学本科计算机专业系列教材

ISBN 978-7-302-31802-6

I. ①形… II. ①蒋… ②姜… III. ①形式语言—高等学校—教材 ②自动机理论—高等学校—教材 IV. ①TP301

中国版本图书馆 CIP 数据核字(2013)第 063608 号

责任编辑:张瑞庆

封面设计:傅瑞学

责任校对:李建庄

责任印制:何 芊

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 装 者:北京鑫海金澳胶印有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:19 字 数:471 千字

版 次:2003 年 1 月第 1 版 2013 年 5 月第 3 版 印 次:2013 年 5 月第 1 次印刷

印 数:1~3000

定 价:36.00 元

产品编号:050871-01

21 世纪大学本科计算机专业系列教材编委会

主 任：李晓明

副 主 任：蒋宗礼 卢先和

委 员：(按姓氏笔画为序)

马华东 马殿富 王志英 王晓东 宁 洪

刘 辰 孙茂松 李仁发 李文新 杨 波

吴朝晖 何炎祥 宋方敏 张 莉 金 海

周兴社 孟祥旭 袁晓洁 钱乐秋 黄国兴

曾 明 廖明宏

秘 书：张瑞庆

本书责任编辑：宋方敏

第3版前言

FOREWORD

培养创新人才,对本科教育来讲,主要是夯实基础、训练思维、养成探索之习惯。所以,创新能力(innovation ability)的培养不能着眼于眼前,简单追求立竿见影,必须面向未来,寻求可持续发展。所以,要追求雄厚的基础(fundaments)、有效的思维(thinking)、勤奋的实践(practice)。这3点简单归纳为“厚基础、善思维、常实践”,可以用如下公式表示:

$$I=F+T+P$$

首先是“厚基础”,包括知识基础和能力基础。对计算机类专业人才来说,重要的理论基础主要来自于理论课程的学习。认真深入地读几本基础性的书,深入理解其中的内容,使自己的思想水平上升到一个新的高度,是非常必要的。为了达到学习知识以提升能力的目的,就要在学习知识的同时,注重对其中蕴含的思想和方法的学习,培养主动探索意识与精神。其次是“善思维”。古人云:“学而不思则罔,思而不学则殆。”要想将书中的知识转化成自己的知识和能力,就必须在认真读书的过程中勤奋地思考。在培养创新思维能力的过程中建立创新意识,形成创新能力。最后,“常实践”是手段。在实践中去加深理解,实践探索。“动手能力”不能是狭义的,它不仅仅简单地来自于下工厂、进企业、进实验室的活动,更不是简单地“编程序”。作为一名科技工作者,“动手”的关键在于“动脑”。

就计算学科而言,离开了理论的指导,就很难有高水平的实践。作者认为,“理论,可以使人‘站到巨人的肩膀上’,并拥有一个‘智慧的脑’”;“实践,需要用智慧的脑,练就一双灵巧的手,去开创一个新世界”。不应该将理论和实践教学割裂开,要有意识地将它们融在一起,这样会收到事半功倍的效果。这就是说,既要“动手”又要“动脑”,要用高水平的动脑,去“指挥”高水平的动手,也就是“理性实践”。而且,不同的专业、不同的课程需要不同形式的实践。就本课程而言,认真地读书,思考一些问题,做一些各种难度的练习,就是一种常规的实践。在这个过程中领悟大师们的思维,从而达到训练思维、提升思维水平的目的,不断强化自己探索未知的意识,提升探索的能力。

这些能力导向教育的思想如何体现在教材中?如何引导读者去发现问题、分析问题、解决问题?如何使得这些引导既深入又简单?它们一直是作者努力探讨的问题。在本书的写作中,除了叙述基本的知识内容外,还努力进行着问题的分析,从而使这些分析在本书中占有很大的篇幅。建议读者不要简单地背定义、定理,要深入地理解,达到能够用自己的语言表达它们的程度。特别要注意认真地阅读分析部分,其中的某一句话可能会使读者产生“恍然大悟”之感,而某一句话可能会引导读者思考更深入的问题。希望读者能够仔细地阅读这些内容,相信会有更多的收获。

本套书自2003年1月出版以来,其第1版在2004年获北京市高等教育教学成果一等奖,2005年被评为北京市精品教材。该套书的第2版是普通高等教育“十一五”国家级规划教材,2008年被评为国家级普通高等教育精品教材。本版作为普通高等教育“十二五”国家级规划教材出版。作者看到,10年来,该教材一直受到读者的欢迎和鼓励,开设此课程的学校很多将其选为教材,使得该套教材成为国内同类教材中发行量和影响力最大的精品教材。另外,清华大学出版社对本套教材的建设,给予了很大的支持,特别是本书的责任编辑张瑞庆编审发挥了重要作用。在此,我们一并表示真诚的感谢。我们相信,随着计算机专业教育的发展,在大家的支持下,该课程在高水平人才的培养中将会进一步发挥作用。

对书中的错误,请读者不吝赐教。联系地址:jiangzl@bjut.edu.cn。

作者

2013年2月

第2版前言

FOREWORD

“离散数学”和“形式语言与自动机理论”是计算机科学与技术专业本科两大专业基础理论课程,这两门课程不仅为学生提供本专业的基础知识,更肩负着培养学生计算思维能力的重要任务。在形式语言与自动机理论中,按照类研究问题的描述,并研究这些描述之间的关系、变换等,从而将问题的求解从实例计算推进到类计算和模型计算,这正是计算学科所追求的,也是本学科的工作者解决问题的着眼点和方法,以及所构建的系统的最大特点和优势。随着计算学科本科生和研究生教育的不断发展,这种优势将进一步凸现。

从2002年到2003年,作者以近20年教学积累和相应的教案为基础,辅以10余年对教育教学的思考撰写成了《形式语言与自动机理论》和配套的《形式语言与自动机理论教学参考书》。这套书出版后,受到了广大读者的欢迎,被选作本科生和研究生教材,成为国内相应课程发行最广的教材,一些相识的和不相识的读者对该书给予了充分的肯定。2004年这套书被评为北京市精品教材,并且荣获2004年北京市高等教育教学成果一等奖,2006年被评为教育部普通高等教育“十一五”国家级规划教材。

在这次修订中,又融进了近几年来作者在从事相关课程教学以及计算机科学与技术专业优秀教材建设中所获得的经验和体会。例如,进一步强调教材是写给读者的,不是写给自己的。而且强调教材的写作特征,认为在这些读者中,首先是写给学生,要考虑他们是初学者,不同类型的学生有不同的关注点,更需要强调用词和描述的准确性、一致性,语言表达的清晰性和叙述的完整性,杜绝陌生名词的突然出现和使用;其次是教师,面对他们,要考虑对现代教育思想的体现和课程的容量;第三是普通的读者,需要通俗易懂,可以提供一些问题的查阅。考虑到作为理论基础性课程教学所确定的内容的稳定性和相应课程关于人才培养的实际需要,本次修订没有追求对知识点及其讲述顺序的调整,主要是进一步提高其可读性、系统性、严密性,特别对一些不太容易掌握含义的地方做了更清楚、更确切的描述,进一步提高了可理解性。

我们必须承认,本课程的基本内容高度抽象,虽然在写作中按照理工科人才培养的实际需要,强调了构造、等价变化等设计形态的内容,但是与其他的课程相比,也难免有一些难度。如何更好地解决这些问题,我们渴望读者能够给出宝贵的建议。另外,书中难免有这样或那样的错误,也恳请读者不吝赐教。联系地址:jiangzl@bjut.edu.cn。

作者

2007年4月

第 1 版前言

FOREWORD

当我们用计算机进行问题的求解时,首先需要建立模型并用适当的数据进行问题表示,然后再用适当的算法通过对这些数据进行变换来获得问题的求解结果。因此,对问题进行抽象和形式化表示,然后进行处理是进行计算机问题求解的基本途径。形式语言与自动机理论给出了一类基本问题的基本描述与计算模型——抽象表示,并通过研究这些模型的性质及其变化方法来对这些问题进行研究。这些模型都是问题模型化的典范,给计算机问题求解提供了一种优美而坚实的基础,而且,也向人们展示了一种典型的方法和思想。另外,它还是研究算法及其理论的基础。

形式语言与自动机理论对计算机科学与技术工作者是非常重要的,它已经成为国际上计算机科学与技术专业本科生的一门重要课程。CC2001-CS 和 CCC2002 给出了明确的要求,里面不仅含有本学科最基本的知识内容,更涉及本学科方法论中所包含的三个学科形态。它们可以被用来引导学生站在更高的高度去看待问题,去粗存真,直击本质,从关键点上以“计算机”的方式解决问题。难怪作者在 1989 年到美国进修时被首先问到的两个问题之一就是“是否学过形式语言与自动机理论?”(另一个问题是“是否学习过算法设计与分析?”)。据统计,在每年 GRE 的考题中,大约有 8~15 道题是关于本课程内容的。

本书包括了 CC2001-CS 和 CCC2002 规定的全部相关知识单元的内容,并且完全满足 CC2001 建议的高级课程自动机理论的教学大纲的要求。它不仅是后续课“编译原理”的理论基础,而且还广泛地用于一些新兴的研究领域。与国外现有的教材比较,本书主要突出如下特点:①充分考虑国内教学计划的容量,进行内容的取舍和组织;②在培养读者的计算思维能力上做进一步的尝试;③尽量照顾国内读者的特点,并且按照国内的教学风格讨论问题。

计算机科学与技术学科要求学生具有形式化描述和抽象思维能力,掌握逻辑思维方法。我们称这种能力为“计算思维”能力,或者叫“计算机思维”能力。当然,一种能力的培养决不是一两门孤立的课程可以实现的,尤其是思维能力的培养,更是如此。它需要一系列的课程,并且通过长期的修养来完成。本课程是这个系列课程中的一门,关于这个系列课程的具体讨论我们将放到 1.4 节进行。本书内容的主要特点是抽象和形式化,既有严格的理论证明,又具有很强的构造性,包含一些基本模型、模型的建立与性质等。通过对本课程的学习,除了使学生掌握有关正则语言、上下文无关语言的文法、识别模型及其基本性质、图灵机的基本知识外,更重要的是还能培养学生的形式化描述和抽象思维能力,同时使学生了解和初步掌握“问题、形式化描述、自动化(计算机化)”的解题思路。这样,我们就扣上“什么能被有效地自动化”这一计算学科的主题。

哈尔滨工业大学从 1977 级本科生开始,一直坚持在本科教学计划中设置此课程。为了

给没有学过此课程的研究生提供机会,还从1982级工学硕士研究生开始,在其计算机科学与技术学科的硕士研究生的培养方案中安排了此门课程。与其他课程相配合,在对学生进行计算思维能力的培养上,取得了良好的效果。本书是作者根据其在该校进行10余年的形式语言与自动机理论课程教学的教案,并参考有关教材撰写而成的。促使作者将教案变成教材的另一个原因是,在国内的教材市场上,这类教材少之又少,根本无法与它在计算机学科的人才培养中的地位相匹配。另外,我们也希望将自己积累的经验和体会提供给大家参考。在本书中,我们希望通过对一些思想和方法的介绍,使读者在这门课程中享受其高度抽象和形式化所带来的美和乐趣。希望通过这些努力,能使这些看似抽象枯燥的内容活起来。许多都是我们自己的体会,其中也难免存在不完善的地方。为了帮助读者更好地学习,附录A提供了包括内容取舍、讲授要点等在内的教学设计。在每章的后面,我们都附有一定量的习题。这些习题用来深化对课程知识的理解,并为读者提供应用所学知识解决问题的机会,使读者亲身体验用相关方法和思想进行探索的乐趣。特别难的习题我们没有列出来,请感兴趣的读者查阅本书后面给出的参考文献。

虽然目前国内计算机科学与技术学科本科生的课程计划中,除了一些重点院校外,设置形式语言与自动机理论课程的学校还不是很普遍,甚至在一些学校的研究生的培养方案中也难以见到此课程。但是,我们相信,随着我国计算机学科教学的不断发展,条件的逐渐成熟,将会有越来越多的学校开设本课程。

本书共分10章。第1章绪论,带领读者回顾在离散数学中学过的本书将要用到的一些基础知识,包括集合及其表示,集合之间的关系,集合的运算,无穷集合,二元关系及其性质,等价关系与等价类,关系的合成,关系的闭包,无向图,有向图,树;另外,介绍形式语言及其相关的基本概念,为后续的章节做准备。第2章介绍文法,包括文法的直观意义与形式定义,推导,归约,文法产生的语言、句子、句型,文法的构造,乔姆斯基体系,左线性文法,右线性文法,空语句。第3章讨论有穷状态自动机,包括DFA作为对实际问题的抽象,直观物理模型,形式定义,DFA接受的句子、语言,状态转移图,构造方法,NFA与DFA的等价性,带空移动的NFA与NFA的等价性,正则文法与FA的等价性及其相互转换方法,基本问题的判定。第4章研究正则表达式,包括正则表达式的定义及其与FA的等价性证明。第5章讨论正则语言的性质,包括正则语言的泵引理的证明及其应用,正则语言的封闭性,Myhill-Nerode定理与FA的极小化,判定算法。第6章讲述上下文无关语言,包括文法二义性,派生与派生树,上下文无关文法的化简,乔姆斯基范式,格雷巴赫范式。第7章叙述下推自动机,包括下推自动机的基本定义,下推自动机用终态接受的语言和用空栈接受的语言,构造方法,下推自动机与上下文无关文法的等价性。第8章研究上下文无关语言的性质,包括上下文无关语言的泵引理、Ogden引理及其应用,上下文无关语言的封闭性,判定算法。第9章介绍图灵机,包括图灵机作为一个计算模型的基本定义,图灵机接受的语言,构造技术,通用图灵机,丘奇-图灵论题,图灵机的变形,可计算语言,不可判定性,P-NP问题。第10章介绍上下文有关语言,包括图灵机与短语结构文法的等价性,线性有界自动机的定义及其与上下文有关语言的关系。

由于作者水平有限,书中的错误和不当之处在所难免,敬请读者批评指正。

作者

2002年12月



普通高等教育“十一五”国家级规划教材 21世纪大学本科计算机专业系列教材

近期出版书目

- 计算概论(第2版)
- 计算概论——程序设计阅读题解
- 计算机导论(第3版)
- 计算机导论教学指导与习题解答
- 计算机伦理学
- 程序设计导引及在线实践
- 程序设计基础(第2版)
- 程序设计基础习题解析与实验指导
- 程序设计基础(C语言)
- 程序设计基础(C语言)实验指导
- 离散数学(第3版)
- 离散数学学习题解答与学习指导(第3版)
- 数据结构(STL框架)
- 算法设计与分析
- 算法设计与分析(第2版)
- 算法设计与分析习题解答(第2版)
- C++程序设计(第2版)
- Java程序设计
- 面向对象程序设计(第2版)
- 形式语言与自动机理论(第3版)
- 形式语言与自动机理论教学参考书(第3版)
- 数字电子技术基础
- 数字逻辑
- FPGA数字逻辑设计
- 计算机组成原理(第3版)
- 计算机组成原理教师用书(第3版)
- 计算机组成原理学习指导与习题解析(第3版)
- 微机原理与接口技术
- 微型计算机系统与接口(第2版)
- 计算机组成与系统结构
- 计算机组成与体系结构习题解答与教学指导
- 计算机组成与体系结构(第2版)
- 计算机系统结构教程
- 计算机系统结构学习指导与题解
- 计算机系统结构实践教程
- 计算机操作系统(第2版)
- 计算机操作系统学习指导与习题解答
- 编译原理
- 软件工程
- 计算机图形学
- 计算机网络(第3版)
- 计算机网络教师用书(第3版)
- 计算机网络实验指导书(第3版)
- 计算机网络习题解析与同步练习
- 计算机网络软件编程指导书
- 人工智能
- 多媒体技术原理及应用(第2版)
- 计算机网络工程(第2版)
- 计算机网络工程实验教程
- 信息安全原理及应用

第 1 章 绪论	1
1.1 集合的基础知识	2
1.1.1 集合及其表示.....	2
1.1.2 集合之间的关系.....	4
1.1.3 集合的运算.....	5
1.2 关系.....	10
1.2.1 二元关系	10
1.2.2 等价关系与等价类	11
1.2.3 关系的合成	12
1.2.4 递归定义与归纳证明	12
1.2.5 关系的闭包	15
1.3 图.....	15
1.3.1 无向图	16
1.3.2 有向图	17
1.3.3 树	19
1.4 语言.....	20
1.4.1 什么是语言	20
1.4.2 形式语言与自动机理论的产生与作用	21
1.4.3 基本概念	23
1.5 小结.....	28
习题	29
第 2 章 文法	35
2.1 启示.....	36
2.2 形式定义.....	38
2.3 文法的构造.....	46
2.4 文法的乔姆斯基体系.....	54
2.5 空语句.....	64
2.6 小结.....	66
习题	66

第3章 有穷状态自动机	70
3.1 语言的识别	70
3.2 有穷状态自动机	72
3.3 不确定的有穷状态自动机	83
3.3.1 作为对 DFA 的修改	83
3.3.2 NFA 的形式定义	84
3.3.3 NFA 与 DFA 等价	86
3.4 带空移动的有穷状态自动机	90
3.5 FA 是正则语言的识别器	94
3.5.1 FA 与右线性文法	94
3.5.2 FA 与左线性文法	98
3.6 FA 的一些变形	99
3.6.1 双向有穷状态自动机	100
3.6.2 带输出的 FA	101
3.7 小结	102
习题	103
第4章 正则表达式	108
4.1 启示	108
4.2 正则表达式的形式定义	109
4.3 正则表达式与 FA 等价	111
4.3.1 正则表达式到 FA 的等价变换	111
4.3.2 正则语言可以用正则表达式表示	119
4.4 正则语言等价模型的总结	124
4.5 小结	126
习题	126
第5章 正则语言的性质	129
5.1 正则语言的泵引理	129
5.2 正则语言的封闭性	134
5.3 Myhill-Nerode 定理与 DFA 的极小化	140
5.3.1 Myhill-Nerode 定理	140
5.3.2 DFA 的极小化	148
5.4 关于正则语言的判定算法	156
5.5 小结	157
习题	158
第6章 上下文无关语言	160
6.1 上下文无关文法	160
6.1.1 上下文无关文法的派生树	161
6.1.2 二义性	166

6.1.3 自顶向下的分析和自底向上的分析	169
6.2 上下文无关文法的化简	171
6.2.1 去无用符号	172
6.2.2 去 ϵ -产生式	175
6.2.3 去单一产生式组	178
6.3 乔姆斯基范式	181
6.4 格雷巴赫范式	184
6.5 自嵌套文法	189
6.6 小结	190
习题	190
第7章 下推自动机	194
7.1 基本定义	194
7.2 PDA 与 CFG 等价	200
7.2.1 PDA 用空栈接受和用终止状态接受等价	200
7.2.2 PDA 与 CFG 等价	203
7.3 小结	212
习题	212
第8章 上下文无关语言的性质	215
8.1 上下文无关语言的泵引理	215
8.2 上下文无关语言的封闭性	221
8.3 上下文无关语言的判定算法	226
8.3.1 L 空否的判定	226
8.3.2 L 是否有穷的判定	227
8.3.3 x 是否为 L 的句子的判定	228
8.4 小结	230
习题	230
第9章 图灵机	231
9.1 基本概念	232
9.1.1 基本图灵机	232
9.1.2 图灵机作为非负整函数的计算模型	239
9.1.3 图灵机的构造	241
9.2 图灵机的变形	247
9.2.1 双向无穷带图灵机	247
9.2.2 多带图灵机	250
9.2.3 不确定的图灵机	252
9.2.4 多维图灵机	253
9.2.5 其他图灵机	255
9.3 通用图灵机	257
9.4 几个相关的概念	259

9.4.1 可计算性	259
9.4.2 P 与 NP 相关问题	259
9.5 小结	260
习题	260
第 10 章 上下文有关语言	263
10.1 图灵机与短语结构文法的等价性	263
10.2 线性有界自动机及其与上下文有关文法的等价性	266
10.3 小结	267
习题	267
附录 A 教学设计	269
A.1 课程内容体系	269
A.1.1 基本描述	269
A.1.2 教学定位	269
A.1.3 知识点与学时分配	270
A.2 课程的讲授	271
A.2.1 重点与难点	271
A.2.2 讲授中应注意的方法等问题	275
A.3 作业	276
A.3.1 指导思想	276
A.3.2 关于大作业和实验	276
A.4 考试与成绩记载	276
A.4.1 成绩评定	276
A.4.2 考题设计	276
附录 B 缩写符号	278
词汇索引	280
参考文献	287

第 1 章

绪 论

计算机科学与技术学科是以数学和电子学科为基础发展起来的,该学科主要包含三方面的内容:一方面是研究计算机领域中的一些普遍规律,描述计算的基本概念与模型,其重点是描述现象,解释规律;另一方面是包括计算机硬件、软件(系统软件和应用软件)在内的计算系统设计和实现的工程技术;第三方面是如何有效地应用计算机技术解决实际问题。因此,人们称这个学科为计算学科,目前在我国称为计算机科学与技术学科,简称计算机学科。该学科的研究与实践包括“理论”、“实验”和“计算”三大科学研究范型。对该学科的人才培养来说,理论和实践教学都非常重要。而其中的理论是基础。按照计算机科学与技术学科方法论的“抽象”、“理论”、“设计”3个过程(形态)来看,实际工作通过理论得到升华,而在理论指导下的设计(实现)才可能是理性的、高水平的。

简单地讲,计算机科学与技术学科通过建立模型并在计算机上模拟物理过程来进行科学调查和研究,它系统地研究信息描述和变换算法,主要包括信息描述和变换算法的理论、分析、效率、实现和应用。

随着计算机科学与技术学科的发展,以及人们对该学科的认识的不断深化,我们认为,该学科的根本问题是:什么能且如何被(有效地)自动计算。

问题的计算机求解建立在高度抽象的基础上,问题的符号表示及处理过程的机械化等固有特性决定了数学是计算机科学与技术学科的重要基础之一,数学及其形式化描述以及严密的表达和计算,使之成为计算机科学与技术学科的重要工具。建立物理符号系统并对其实施变换是计算机科学与技术学科进行问题描述和求解的重要手段。学科所要求的计算机问题求解的“可行性”限定了从问题抽象开始到根据适当理论的指导进行设计(实现)的科学实践过程,“可行性”所需要的“形式化”后呈现出的符号及其变换的“离散特征”,实质上决定了计算机科学与技术学科进行问题求解的重要特征,从而决定了离散数学对本学科的重要性。

本章简要回顾离散数学中的部分基本概念和方法,以使读者能够顺利地进行本书的学习。如果读者对这部分内容比较熟悉,建议快速地浏览 1.1 节~1.3 节,从而熟悉本书使用的符号和对问题的叙述方式。

另外,建议读者能较好地完成本章后面所列的习题,尤其是一些构造性题目,以及关于语言的所有题目,它们对后续内容的学习十分重要。

本章的主要内容有两部分:一是有关集合、关系、图、证明方法的基本知识;二是讲述形式语言及其相关的基本概念。

1.1 集合的基础知识

集合论(set theory)是德国数学家康托(Georg Cantor)于1874年创立的,至今已经历了两个阶段:1908年以前称为朴素集合论,又称为康托集合论。在朴素集合论中,存在着严重的集合悖论问题。在朴素集合论刚出现时,人们认为找到了数学的基础,而集合悖论的发现使得人们无比沮丧,觉得数学失去了重要的基础,甚至有人认为以严密为重要特征的数学是不可靠的。为了避免集合悖论,哲墨罗(E. Zermelo)于1908年提出了第一个集合论公理系统,后经富兰科尔(A. A. Fraenkel)和斯库利姆(A. T. Skolem)改进和补充,形成了ZF公理系统。同年,B. 罗素也给出了关于集合型的层次理论——类型论。

无论如何,人们还是公认集合论在数学中占有非常重要的地位,它的基本概念已经渗透到许多领域。计算机科学与技术学科出现后,也将集合论作为其重要基础之一,其大多数基本概念和理论都采用了以集合论为基础的术语来描述。

1.1.1 集合及其表示

集合是集合论中最原始的概念。这里只能给出它的非形式描述,以说明它的意义:一定范围内确定的,并且彼此可以区分的对象汇集在一起形成的整体叫作集合(set),简称为集。简单地说,集合是具有某种性质的对象的全体。构成集合的每一个对象称为这个集合的一个成员,它们可以是具体的东西,也可以是抽象的概念。通常称集合的成员为该集合的元素(element)。

例 1-1 集合的实际例子。

- (1) 北京市的所有交通工具汇集在一起构成一个集合。
- (2) 北京市的所有公共汽车是一个集合。
- (3) 中国所有高等院校构成一个集合。
- (4) 某高校的所有院系构成一个集合。
- (5) 全体自然数构成一个集合;全体有理数构成一个集合;全体实数构成一个集合。
- (6) 一个学校的所有班级的全体是一个集合;一个班的学生的全体是一个集合;一个学生的所有用品的全体是一个集合。

显然,对象和集合之间有这样一种关系:该对象要么是该集合的一个元素,要么不是该集合的元素,二者必居其一。一个集合中的元素可能都在另一个集合中,也可能部分在另一个集合中。一个对象可以是某一个集合的元素,它本身也可以是一个集合。

通常用大写的英文字母 A, B, C, \dots 和大写的希腊字母 $\Gamma, \Sigma, \emptyset, \dots$ 表示集合,用小写字母 a, b, c, \dots 表示集合的元素。例如,一般地,

\mathbf{N} ——表示全体自然数集合

\mathbf{Q} ——表示全体有理数集合

\mathbf{R} ——表示全体实数集合

Σ ——表示字母的集合

关于集合和元素,用如下记法:

如果 a 是集合 A 的一个元素,则记为 $a \in A$,读作 a 属于 A ,或者 A 含有 a ;否则记为 a

$\notin A$, 读作 a 不属于 A , 或者 A 不含 a 。

例 1-2 集合与元素。

(1) $6 \in \mathbf{N}, 1 \in \mathbf{N}, 1.5 \notin \mathbf{N}, 0.81 \notin \mathbf{N}$ 。

(2) $4 \in \mathbf{N}, 1.5 \notin \mathbf{N}, 4 \in \mathbf{Q}, 1.5 \in \mathbf{Q}, 4 \in \mathbf{R}, 1.5 \in \mathbf{R}$ 。

(3) 设 U 是中国所有高等院校组成的集合, 则有

清华大学 $\in U$, 北京大学 $\in U$, 哈尔滨工业大学 $\in U$, 北京工业大学 $\in U$

(4) 设北京工业大学表示该校所有院系组成的集合, 则有

计算机学院 \in 北京工业大学

集合可以用两种形式加以描述:

第一种形式称为**列举法**(listing): 将所有的元素逐一地列举在大括号 $\{ \}$ 中, 读者能立即看出规律时, 某些元素可用省略号表示。

例 1-3 集合的列举表示。

(1) $\{1, 3, 6, 9, 10\}$ 。

(2) $\{a, b, c, \dots, z\}$ 。

(3) $\{\text{本科生}, \text{硕士研究生}, \text{博士研究生}, \text{进修生}\}$ 。

(4) $\{1, 2, 3, 4, 5, \dots\}$ 。

(5) $\{0, 5, 10, 15, \dots, 200\}$ 。

值得注意的是, 在使用列举法时, 集合中元素出现的先后顺序是没有意义的。例如, $\{1, 3, 6, 9, 10\}$ 与 $\{6, 3, 10, 9, 1\}$ 表示的是同一个集合。

集合的第二种表示形式称为**命题法**(proposition), 其基本形式为

$$\{x \mid P(x)\}$$

其中, P 为谓词, 表示此集合包括所有使 P 为真的 x 。

例 1-4 集合的命题表示。

(1) $\{x \mid 0 \leq x \leq 200 \text{ 且 } (\exists n \in \mathbf{N} (n \cdot 5 = x))\}$ 。

(2) $\{x \mid 3x^2 + 8x + 4 = 0\}$ 。

(3) $\{x \mid x \in [0, 1]\}$, 按照适当约定, 该集合就是 $[0, 1]$ 。

(4) $\{x \mid x \in \{\text{本科生}, \text{硕士研究生}, \text{博士研究生}, \text{进修生}\}\}$, 也就是 $\{\text{本科生}, \text{硕士研究生}, \text{博士研究生}, \text{进修生}\}$ 。

在有的集合中, 一个元素可以重复出现, 这种集合称为多重集合。本书不考虑多重集合的问题, 所提到的集合均不允许元素重复出现。

由有限个元素构成的集合叫作**有限集**(finite set), 又称为**有穷集**。由无穷多个元素构成的集合叫作**无穷集**(infinite set)。这是一个直观的描述, 作为一个思考题, 读者可以根据定义 1-1 分别给有穷集和无穷集一个比较严格的定义。

定义 1-1 如果集合 A, B 之间有一个一一对应, 则称它们具有相同的**基数**(cardinality), 通常用 $|A|$ 表示集合 A 的基数。

集合的基数又叫作集合的**势**。

对有穷集来说, 它的基数就是它所包含的元素的个数。

例 1-5 有穷集合的基数。

(1) $|\{x \mid 0 \leq x \leq 200 \text{ 且 } (\exists n \in \mathbf{N} (n \cdot 5 = x))\}| = 41$ 。