

面向计算机科学与技术专业规范系列教材




---

# 形式语言与 自动机

陈有祺 编著

---



*Formal Languages and Automata*

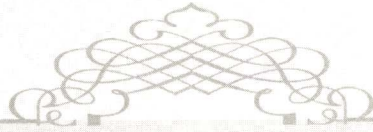


机械工业出版社  
China Machine Press



TP301  
CYQ

面向计算机科学与技术专业规范系列教材



# 形式语言与 自动机

陈有祺 编著



Formal Language and Automata



机械工业出版社  
China Machine Press

本书以四类形式语言(短语结构语言、上下文有关语言、上下文无关语言、正则语言)和四种自动机(有穷自动机、下推自动机、图灵机、线性有界自动机)为主线,讨论了形式语言与自动机方面的主要理论成果和应用实例。书中每一章的最后都配有大量不同难度的习题,有助于读者掌握本书内容。

本书采用通俗的语言和形象化的方法来表达概念和定理,逻辑严谨、思维缜密,可作为高等院校计算机及相关专业“形式语言与自动机”课程的教材。

版权所有,侵权必究。

本书法律顾问 北京市展达律师事务所

### 图书在版编目(CIP)数据

形式语言与自动机/陈有祺编著. —北京:机械工业出版社,2008.9  
(面向计算机科学与技术专业规范系列教材)

ISBN 978-7-111-23776-1

I. 形… II. 陈… 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. TP301

中国版本图书馆 CIP 数据核字(2008)第 038412 号

机械工业出版社(北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑:迟振春

三河市明辉印装有限公司印刷·新华书店北京发行所发行

2008 年 9 月第 1 版第 1 次印刷

185mm×260mm·15.25 印张

标准书号:ISBN 978-7-111-23776-1

定价:29.00 元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换  
本社购书热线:(010)68326294

# 出版者的话

机械工业出版社华章公司是国内重要的教育出版公司，培生教育集团(拥有 Addison Wesley、Prentice Hall 等品牌)是全球知名的教育出版集团，双方在过去长达十余年的合作中秉承“全球采集内容，服务教育事业”的理念，遴选、移译了国外大量的在计算机科学界享誉盛名的专家名著与名校教材，其中包括 Donald E. Knuth、Alfred V. Aho、Jeffrey D. Ullman、John E. Hopcroft、Dennis Ritchie 等大师名家的经典作品(收录在大理石封面的“计算机科学丛书”中)，这些作品对国内计算机教育及科研事业的发展起到了积极的促进作用。

随着国内计算机科学与技术专业学科建设的不断完善、教学研究的蓬勃发展，以及教材改革的逐渐深化，计算机科学与技术专业的优秀课程及教材不仅仅是“引进来”(版权引进)，而且需要“走出去”(版权输出)了。

近几年以来，教育部计算机科学与技术专业教学指导分委员会根据我国计算机专业教育的现状以及社会对人才的需求，发布了《高等院校计算机科学与技术专业发展战略研究报告暨专业规范(试行)》(以下简称《规范》)。为配合《规范》的实施推广，同时为落实中央“提高高等教育质量”的最新指导思想，在教育部计算机科学与技术专业教学指导分委员会的指导下，在国内知名高校众多教授的帮助下，我们出版了这套“面向计算机科学与技术专业规范系列教材”。

本套教材的作者在长达数十年的科研和教学经历中积累了大量的知识和经验，也奠定了他们在学术和教学领域的地位，教材的内容体现了他们的教学思想和教学理念，本套教材也是传承他们优秀教学成果的最好载体，是中国版的专家名著和名校教材，相信它们的出版对提高计算机科学与技术专业的教育水平和教学质量能够起到积极的作用。

华章与培生作为专业的出版团队，愿与高等院校的老师共同携手，在这套教材的出版上引进国际先进教材出版经验，在教学配套资源的建设上做出新的尝试，为促进中国计算机科学与技术专业教育事业的发展，为增进中国与世界文化的交流而努力。



华章教育



培生教育集团

# 面向计算机科学与技术本科专业规范系列教材

## 编委会

主任委员： 蒋宗礼

副主任委员： 王志英 钱乐秋

委 员：（以姓氏拼音为序）

陈道蓄	陈 明	何炎祥	傅育熙
黄刘生	贾云得	姜守旭	马殿富
李仁发	李晓明	刘 辰	齐 勇
孙吉贵	孙茂松	吴功宜	吴 跃
谢长生	于 戈	张 钢	周兴社

秘书组： 温莉芳 刘立卿 姚 蕾

本书责任编委： 蒋宗礼

# 序 言

近 20 年里, 计算机学科有了很大的发展, 人们普遍认为, “计算机科学”这个名字已经难以涵盖该学科的内容, 因此, 改称其为计算学科(Computing Discipline)。在我国本科教育中, 1996 年以前曾经有计算机软件专业和计算机及应用专业, 之后被合并为计算机科学与技术专业。2004 年以来, 教育部计算机科学与技术教学指导委员会根据我国计算机专业教育和计算学科的现状, 为更好地满足社会对计算机专业人才需求, 发布了《高等学校计算机科学与技术专业发展战略研究报告暨专业规范(试行)》(以下简称《规范》), 提出在计算机科学与技术专业名称之下, 构建计算机科学、计算机工程、软件工程和信息技术四大专业方向。《规范》中四大专业方向的分类, 在于鼓励办学单位根据自己的情况设定不同的培养方案, 以培养更具针对性和特色的计算机专业人才。

为配合《规范》的实施, 落实中央“提高高等教育质量”的精神, 我们规划了“面向计算机科学与技术本科专业规范系列教材”。本系列教材面向全新的计算学科, 针对我国高等院校逐步向新的计算机科学与技术专业课程体系过渡的趋势编写, 在知识选择、内容组织和教学方法等方面满足《规范》的要求, 并与国际接轨。本套教材具有以下几个特点:

(1) **体现《规范》的基本思想, 满足其课程要求。**为使教材符合我国高等院校的教学实际, 编委会根据《规范》的要求规划本套教材, 广泛征集在国内知名高校中从事一线教学和科研工作、经验丰富的优秀教师承担编写任务。

(2) **围绕“提高教育质量”的宗旨开发教材。**为了确保“精品”, 本系列教材的出版不走盲目扩大的路子, 每本教材的选题都将由编委会集体论证, 并由一名编委担任责任编委, 最大程度地保证这套教材的编写水准和出版质量。

(3) **教材内容的组织科学、合理, 体系得当。**本套教材的编写注重研究学科的新发展和新成果, 能够根据不同类型人才培养需求, 合理地进行内容取舍、组织和叙述, 还精心设计了配套的实验体系和练习体系。

(4) **教材风格鲜明。**本套教材按 4 个专业方向统一规划, 分批组织, 陆续出版。教材的编写体现了现代教育理念, 探讨先进的教学方法。

(5) **开展教材立体化建设。**根据需要配合主教材的建设适时开发实验教材、教师参考书、学生参考书、电子参考资料等教辅资源, 为教学实现多方位服务。

我们衷心希望本系列教材能够为我国高等院校计算机科学与技术等专业的教学作出贡献, 欢迎广大读者广为选用。

“面向计算机科学与技术专业规范系列教材”编委会

# 前言

形式语言和自动机理论是理论计算机科学的重要分支，自 20 世纪 60 年代以来发展极为迅速，已经在计算机科学的许多领域起着理论基础和方法论的作用，特别是对程序语言的设计、编译理论与技术、模式识别和自然语言理解等领域起了重要的促进作用。近年来，形式语言和自动机理论的应用范围不断扩大，无论是自然科学的许多领域，还是社会科学的某些领域，都在应用形式语言和自动机的理论成果和描述方式。因此，各发达国家的计算机科学界和教育界对此十分重视，各大学的计算机科学系都把形式语言和自动机方面的内容列为本科高年级学生和研究生的重要课程。美国每年的计算机专业 GRE 考试中，都有和本书内容有关的考题。我国从改革开放以来，各主要大学也都陆续开设了这方面的课程，而且形式语言和自动机理论越来越受到有关方面的重视，并有不断推广的趋势。

本书作者从事形式语言与自动机方面的教学和研究工作已有 30 余年，在总结经验的基础上，广泛参考国内外有关著作(包括国外最新出版的教科书)，编写了这本教材。教材内容以四类形式语言和四种自动机为主线，讨论了形式语言与自动机方面的主要理论成果和应用实例。全书共分 12 章：第 1 章为预备知识，介绍一些本书常用的基本概念和推理方法。第 2 章引入文法的概念，按乔姆斯基体系介绍四类形式文法。第 3 章到第 5 章是一个单元，其中，第 3 章引入有穷自动机的基本概念，并详细介绍各种形式的有穷自动机和它们的应用；第 4 章介绍正则表达式，提出各种等价变换方法，并论述正则表达式和有穷自动机的关系；第 5 章对正则语言的各种表现形式加以总结，论述正则语言的各种重要性质，并重点讨论有穷自动机的极小化问题。第 6 章到第 8 章是一个单元，其中，第 6 章详细讨论乔姆斯基体系中最重要的一类文法——上下文无关文法；第 7 章引入下推自动机的概念，并证明下推自动机和上下文无关文法的等价性；第 8 章讨论上下文无关语言的性质，作为本单元的一个总结。第 9 章引入最重要的一类自动机——图灵机，给出它的基本模型和各种变形，其中用很多例子表明图灵机具有很强的识别能力和计算能力，并对图灵机与现代计算机的能力做了比较，最后证明图灵机和  $\Omega$  型文法的等价性。第 10 章集中论述不可判定问题，这一章理论性较强，有一定的难度。第 11 章介绍线性有界自动机，并证明它与上下文有关文法的等价性，最后对各语言类之间的关系做一总结。第 12 章进一步讨论确定的下推自动机和确定的上下文无关语言，给出  $LR(k)$  文法的定义，并论述它和确定的下推自动机之间的关系。

本书内容较为丰富，如要完全讲授至少需要 70 学时。建议目录中加“\*”号的章节不讲，第 10 章内容对研究生可以少讲，对本科生可以不讲。此外，其他章节内容也可适当略去一些。但是，无论如何都要保留四类语言和四种自动机以及它们的对应关系这个主线(即本书前 9 章的主要内容和第 11 章的部分内容)，否则就不能说学过“形式语言与自动机”这门课程。学习本课程之后，不仅能对学生学习某些后继课(如算法设计与分析、模式识别等)有直接帮助，而且更重要的是能提高他们的逻辑思维和抽象表达能力。

在写作方法上，作者力求做到深入浅出。在概念的引入和定理的证明上，一方面逻辑严谨、思维缜密；另一方面尽量采用通俗的语言和形象化的方法来表达，并配合一些能说明问

题的例子，使读者容易理解抽象理论的实质，以达到加强理论素养和提高灵活运用能力的目的。书中每一章的最后都配有大量不同难度的习题，有助于读者掌握本书内容，读者可在教师的指导下选择一部分难易适当的习题来做。

本书在立项、编写和出版的过程中，得到机械工业出版社华章公司的大力支持，在此表示衷心的感谢。还要对辛运伟教授致以深深的谢意，她对本书从大纲到具体内容都提出了许多宝贵的意见。由于作者的水平有限，书中难免有错误和不确切之处，恳请各位专家和广大读者批评指正。

陈有祺  
于南开园



# 教学建议

## 第1章 预备知识(2~4学时)

本章大部分内容属于复习性质。如果学生已经在离散数学等课程中学过有关集合论和图的内容,则这些内容简单复习一下即可。如果没有学过,则要用一定课时重点讲一下。至于1.4节的字母表、字符串和语言,是新的内容而且是后面必用的,一定要讲清楚。

## 第2章 文法的一般理论(6学时)

准确掌握文法的形式定义以及文法和语言的关系。通过若干例子的分析,逐渐熟练掌握如何从给出的文法得出它所产生的语言,如何从给出的语言构造产生它的文法。准确掌握乔姆斯基对文法的分类标准。正确理解文法等价的概念和文法类等价的概念。

## 第3章 有穷自动机(6~8学时)

准确掌握有穷自动机的形式定义以及它所接受的语言。熟悉并熟练运用有穷自动机的表达方式。通过若干例子的分析,逐渐熟练掌握如何从给出的语言构造接受它的有穷自动机。准确掌握有穷自动机的两种等价形式——非确定的有穷自动机和具有 $\epsilon$ 转移的有穷自动机的定义。了解有穷自动机的应用。至于具有输出动作的有穷自动机,在目录中已打了\*号,因为这部分内容在本书的主干体系之外,时间不够可以不讲。

## 第4章 正则表达式(4学时)

准确掌握正则表达式的定义和它代表的语言,逐渐熟练如何从给出的语言构造代表它的正则表达式。了解正则表达式和有穷自动机等价性的证明方法。熟悉正则表达式的各种等价变换方法。了解正则表达式的应用。

## 第5章 正则语言的性质(6~8学时)

本章是对正则语言的一个总结。通过正则文法和有穷自动机等价性的证明,将正则语言的几种等价的表现形式——正则文法、有穷自动机(包括非确定的有穷自动机和具有 $\epsilon$ 转移的有穷自动机)和正则表达式有机地联系起来。掌握正则语言泵引理的实质和它的应用。了解正则语言的各种封闭性质和判定算法。准确掌握有穷自动机最小化的理论和实际算法(这一部分有较大难度,对本科生可以酌情少讲)。

## 第6章 上下文无关文法(6学时)

掌握上下文无关文法语法分析的有关概念(语法分析树、最左推导与最右推导、文法的二义性、自顶向下分析和自底向上分析)。掌握上下文无关文法中无用符号、 $\epsilon$ -产生式和单一产生式的消除算法。准确掌握上下文无关文法两种范式的定义和构造算法。了解上下文无关文法的某些应用。

## 第7章 下推自动机(4学时)

准确掌握下推自动机的定义以及它所接受的语言。熟悉并熟练运用下推自动机的表达方式,逐渐熟练对于给定的语言构造出接受它的下推自动机。了解下推自动机和上下文无关文法等价关系的证明过程。了解确定的下推自动机的定义。

## 第 8 章 上下文无关语言的性质(4 学时)

掌握上下文无关语言泵引理(包括它的加强形式——Ogden 引理)的实质和它们的应用。了解上下文无关语言的各种封闭性质和判定算法。

## 第 9 章 图灵机导引(8 学时)

准确掌握图灵机基本模型的定义以及它所接受的语言。熟悉并熟练运用图灵机的表达方式。学习图灵机的各种程序设计方法,逐渐熟练对于给定的语言构造出接受它的图灵机。了解图灵机的各种变形和它们与基本模型等价性的证明。了解图灵机与现代计算机可以相互模拟的关系。掌握图灵机与 0 型文法等价关系的证明方法。

## 第 10 章 不可判定性(6~8 学时)

本章理论性较强,有一定难度。建议对本科生基本不讲,对研究生视学时情况可以少讲。其中 10.1 节没有什么困难,应当掌握。后面各节属于比较抽象的部分,但是研究生应当知道非递归可枚举集的存在和一些主要的不可判定问题(如停机问题、成员问题等)。至于莱斯定理和关于上下文无关语言的各种不可判定问题,时间不够可以不讲。最后一节波斯特对应问题在目录中已打上 \* 号,虽然很有名且有一定的重要性,但一般可以不讲。

## 第 11 章 线性有界自动机和上下文有关文法(4 学时)

准确掌握线性有界自动机的定义以及它所接受的语言。掌握线性有界自动机和上下文有关文法等价关系的证明方法。了解上下文有关语言的性质及其与递归语言的关系。了解各语言类之间的关系。这最后一部分内容带有总结性质,当然很重要,但是若在第 10 章基本未讲、本章的上下文有关语言和递归语言的关系又没有详细讲解的情况下,各语言类之间的关系只能给出结论,不能做严格证明。

## 第 12 章 确定的上下文无关语言和 LR(k)文法(4~8 学时)

在目录中本章已打上 \* 号,这意味着全章内容一般可以不讲。原因是:一方面学校教学计划不可能为本课程安排太多的学时;另一方面本章内容虽然也很重要,但不在本书的主干体系之中。如果有时间的话可以讲前两节,这样就把确定的上下文无关语言的有关内容讲清楚了,也是对上一章各语言类之间的关系做一个必要的补充。

## 习题和习题课(2~4 学时)

本书各章最后都附有习题,任课教师可以根据情况,给学生留一些基本的和中等难度的习题作为课外作业。如有时间,也可以安排 1 至 2 次习题课。在习题课上可以由教师讲解以前课外作业中存在的带有普遍性的问题,也可以安排稍难一些的习题让学生在课上做出解答,然后由教师指导进行讨论,最后得出不仅正确而且较好的答案。

如有条件,在讲完第 9 章之后可以安排一次上机作业(不算学时),内容可选第 9 章习题 9.9 作为练习,这样可以对图灵机有更深入的理解。

# 面向计算机科学与技术专业规范系列教材

---

- ☆离散数学及其应用
  - 数据结构与算法
- ☆C程序设计高级教程
  - 操作系统原理与实践
  - 数据库系统
  - 计算机网络
- ☆计算机网络与互联网
  - 编译原理与实践
- ☆计算机组成
  - 计算机体系结构
  - 嵌入式系统
- 软件工程
- ☆面向对象分析与设计
- ☆软件测试
- ☆软件需求工程
  - 软件项目管理
  - 人工智能
- ☆形式语言与自动机
  - 电路与系统
  - 数字电路与逻辑设计
  - 大规模集成电路设计
  - Web系统与技术
  - 信息安全

注：☆为近期出版图书

---

# 教师服务登记表

尊敬的老师：

您好！感谢您购买我们出版的 \_\_\_\_\_ 教材。

机械工业出版社华章公司本着为服务高等教育的出版原则，为进一步加强与高校教师的联系与沟通，更好地为高校教师服务，特制此表，请您填妥后发回给我们，我们将定期向您寄送华章公司最新的图书出版信息。为您的教材、论著或译著的出版提供可能的帮助。欢迎您对我们的教材和服务提出宝贵的意见，感谢您的大力支持与帮助！

个人资料（请用正楷完整填写）

教师姓名		<input type="checkbox"/> 先生 <input type="checkbox"/> 女士	出生年月		职务		职称： <input type="checkbox"/> 教授 <input type="checkbox"/> 副教授 <input type="checkbox"/> 讲师 <input type="checkbox"/> 助教 <input type="checkbox"/> 其他
学校				学院			
联系电话	办公：			联系地址及邮编			
	宅电：						
	移动：			E-mail			
学历		毕业院校			国外进修及讲学经历		
研究领域							
主讲课程			现用教材名		作者及出版社	共同授课教师	教材满意度
课程： <input type="checkbox"/> 专 <input type="checkbox"/> 本 <input type="checkbox"/> 研 人数：      学期： <input type="checkbox"/> 春 <input type="checkbox"/> 秋							<input type="checkbox"/> 满意 <input type="checkbox"/> 一般 <input type="checkbox"/> 不满意 <input type="checkbox"/> 希望更换
课程： <input type="checkbox"/> 专 <input type="checkbox"/> 本 <input type="checkbox"/> 研 人数：      学期： <input type="checkbox"/> 春 <input type="checkbox"/> 秋							<input type="checkbox"/> 满意 <input type="checkbox"/> 一般 <input type="checkbox"/> 不满意 <input type="checkbox"/> 希望更换
样书申请							
已出版著作				已出版译作			
是否愿意从事翻译/著作工作 <input type="checkbox"/> 是 <input type="checkbox"/> 否				方向			
意见和建议							

填妥后请选择以下任何一种方式将此表返回：（如方便请赐名片）

地 址：北京市西城区百万庄南街1号 华章公司营销中心 邮编：100037

电 话：(010) 68353079 88378995 传 真：(010) 68995260

E-mail: hzedu@hzbook.com marketing@hzbook.com 图书详情可登录<http://www.hzbook.com>网站查询

# 目 录

出版者的话

序 言

前 言

教学建议

第 1 章 预备知识 .....	1
1.1 定理及其证明方法 .....	1
1.1.1 演绎法 .....	2
1.1.2 反证法 .....	3
1.1.3 归纳法 .....	5
1.2 集合及其基本运算 .....	7
1.2.1 集合基础知识 .....	7
1.2.2 集合的基本运算 .....	8
1.2.3 关系与映射 .....	9
1.3 图和树简介 .....	11
1.3.1 图的基本概念 .....	12
1.3.2 图的矩阵表示 .....	13
1.3.3 树的基本知识 .....	15
1.4 字母表、字符串和语言 .....	16
习题 .....	18
第 2 章 文法的一般理论 .....	21
2.1 问题的提出 .....	21
2.2 形式文法与形式语言 .....	23
2.3 文法的乔姆斯基分类 .....	30
习题 .....	40
第 3 章 有穷自动机 .....	42
3.1 非形式化描述 .....	42
3.2 有穷自动机的基本定义 .....	44
3.3 非确定的有穷自动机 .....	48
3.4 具有 $\epsilon$ 转移的有穷自动机 .....	52
3.5 有穷自动机的应用 .....	55
3.5.1 在文本中查找字符串 .....	55
3.5.2 用于文本搜索的非确定 的有穷自动机 .....	55
3.5.3 识别关键字集合的 DFA .....	56
3.6 具有输出的有穷自动机 .....	58
习题 .....	60
第 4 章 正则表达式 .....	63
4.1 正则表达式的定义 .....	63
4.2 正则表达式和有穷自动机的 关系 .....	65
4.3 正则表达式的等价变换 .....	70
4.3.1 交换律与结合律 .....	70
4.3.2 单位元与零元 .....	71
4.3.3 分配律 .....	71
4.3.4 与“*”构造有关的 定律 .....	72
4.3.5 发现正则表达式定律的 一般方法 .....	73
4.4 正则表达式的应用 .....	75
4.4.1 UNIX 中的正则 表达式 .....	75
4.4.2 词法分析 .....	76
4.4.3 查找文本中的模式 .....	77
习题 .....	78
第 5 章 正则语言的性质 .....	80
5.1 正则文法和有穷自动机的 关系 .....	80
5.2 正则语言的泵引理 .....	82
5.3 正则语言的封闭性 .....	85
5.4 正则语言的判定算法 .....	89
5.5 有穷自动机的最小化 .....	91
习题 .....	99
第 6 章 上下文无关文法 .....	101
6.1 上下文无关文法的语法分析 .....	101
6.2 上下文无关文法的化简 .....	107
6.3 上下文无关文法的范式 .....	111
6.4 上下文无关文法的应用 .....	115
6.4.1 用上下文无关文法描述 语言 .....	115
6.4.2 语法分析器生成工具 YACC .....	116



6.4.3	标记语言	117	9.4.3	比较计算机与图灵机的 运行时间	172
6.4.4	XML 和文档类型 定义	118	9.5	图灵机与 0 型文法的关系	173
	习题	121		习题	176
<b>第 7 章</b>	<b>下推自动机</b>	124	<b>第 10 章</b>	<b>不可判定性</b>	178
7.1	下推自动机的定义	124	10.1	递归集和递归可枚举集的 性质	178
7.2	下推自动机接受的语言	125	10.2	通用图灵机和第一个不可判定 问题	180
7.3	下推自动机和上下文无关 文法的关系	130	10.3	归约方法和莱斯定理	183
7.4	确定的下推自动机	134	10.4	关于上下文无关语言的不可 判定问题	188
	习题	135	*10.5	波斯特对应问题的不可判定性 及其应用	191
<b>第 8 章</b>	<b>上下文无关语言的性质</b>	136		习题	197
8.1	上下文无关语言的泵引理	136	<b>第 11 章</b>	<b>线性有界自动机和上下文 有关文法</b>	199
8.2	上下文无关语言的封闭性	141	11.1	线性有界自动机	199
8.3	上下文无关语言的 判定算法	142	11.2	线性有界自动机和上下文有关 文法的关系	201
	习题	146	11.3	上下文有关语言的性质及其与 递归集的关系	202
<b>第 9 章</b>	<b>图灵机导引</b>	148	11.4	各语言类之间的关系	204
9.1	图灵机的基本模型	148		习题	205
9.2	图灵机的程序设计技术	154	<b>*第 12 章</b>	<b>确定的上下文无关语言和 LR(<math>k</math>)文法</b>	206
9.2.1	在状态中存储符号	154	12.1	确定的下推自动机的标准 形式	206
9.2.2	多道技术	155	12.2	确定的上下文无关语言的 性质	207
9.2.3	子程序技术	159	12.3	LR(0)文法	212
9.3	图灵机的变形	161	12.4	LR(0)文法与 DPDA 的 关系	216
9.3.1	双向无限带	161	12.5	LR( $k$ )文法	221
9.3.2	多带	162		习题	226
9.3.3	非确定的图灵机	165	<b>参考文献</b>		228
9.3.4	双栈机	166			
9.3.5	作为枚举器的 图灵机	167			
9.4	图灵机与计算机	169			
9.4.1	用计算机模拟 图灵机	169			
9.4.2	用图灵机模拟 计算机	170			



## 预备知识

本书属于理论计算机科学的范畴，需要较多的数学基础知识和严格的推理能力。因此，本章对后面各章所需的基础知识作一个简要的介绍，其中有些内容可能是读者已经学过的，这里就作为复习；有些内容是为本书的讲述专门安排的，希望读者认真掌握。

### 1.1 定理及其证明方法

本书的理论性较强，许多论断都写成了定理的形式，而定理的正确性是需要证明的。读者从中学学习初等数学开始，一直到大学学习高等数学，已学习了不少证明知识。但是，本节还是要简述究竟什么是证明，并且介绍常用的证明方法，其中许多方法都是本书中常用的。

什么是证明呢？最严格意义下的证明应当是由一个形式系统来刻画的，这个形式系统包括以下基本成分：

- (1) 一组基本符号；
- (2) 一组形成规则；
- (3) 一组公理；
- (4) 一组推理规则。

基本符号包含常量符号、变量符号和运算符号等。从这些基本符号出发，按照形成规则构造各种语句。有些语句的合法性是人们通过长期生活实践所公认的，这些语句就是系统的公理。此外，系统还需要一些推理规则，通过系统的公理和这些推理规则，能得出新的合法语句。

在一个系统中，当我们要证明某个语句为真(合法)时，需要通过一个语句序列，它们是：

- (1) 每个语句或者是公理，或者能借推理规则由一个或多个前面的语句推导出来。
- (2) 此序列的最后一个语句就是所要证明的语句。

能够加以证明的语句，称为该形式系统的定理。显然，形式系统中的每个公理都是该系统的定理。

下面我们对公理和推理规则稍加解释。例如，某形式系统中有公理  $P$ 、 $Q$  和  $R$ ，还有以下推理规则：

- (1)  $P, Q \vdash S$  (意思是当  $P$  和  $Q$  皆为真时， $S$  为真)。
- (2)  $Q, R \vdash T$  (意思是当  $Q$  和  $R$  皆为真时， $T$  为真)。

(3) $R, S, T \vdash W$ (意思是当  $R, S, T$  皆为真时,  $W$  为真)。

若想证明  $W$  是定理, 则推理过程为:

(1) $P$  为真(因为  $P$  是公理)

(2) $Q$  为真(因为  $Q$  是公理)

(3) $R$  为真(因为  $R$  是公理)

(4) $S$  为真(应用规则(1))

(5) $T$  为真(应用规则(2))

(6) $W$  为真(应用规则(3))

这样, 就证明了  $W$  是定理。

虽然某些数学分支可以建立自己的一套形式系统, 并在此系统内用形式推理的方法给出各种定理, 但是实际上要把该领域的公理和推理规则都写出来是很不容易的。而且, 用这种方法证明定理的过程冗长而枯燥。只有在某些特殊的领域(如欧几里得几何学), 用计算机进行自动定理证明时使用形式推理系统。在一般的数学著作中, 仍然使用非形式化的证明方法, 即在证明过程中用自然语言叙述, 用一般的逻辑推理方法进行推理。其中用到的事实有的是从前面的定义中来, 有的是前面已经证明了的结论, 有的是读者早已学过的知识(也许是中小学课本中的内容), 等等。而且, 当一些事实为大家所熟知时, 可以略去不提; 当一些命题非常明显时(如  $y \geq 1$  则  $y^2 \geq 1$ ), 则不必分析证明。这就大大简化了证明的过程, 且不失证明的严格性。

### 1.1.1 演绎法

上面讲到的形式化证明就是典型的演绎法, 非形式化的演绎法也是根据已知事实和定理的假设, 使用逻辑推理的方法逐步得出结论。下面以定理形式给出几个例子。

**定理 1.1** 如果  $x \geq 4$ , 则  $2^x \geq x^2$ 。

为了便于证明, 现将结论变成等价的形式:

$$2^x - x^2 \geq 0 \quad (1-1)$$

对(1-1)式左边进行微商后, 得出

$$2^x \ln 2 - 2x \quad (1-2)$$

对(1-2)式再进行微商后, 得出

$$2^x (\ln 2)^2 - 2 \quad (1-3)$$

因为  $\ln 2 = 0.6971$ , 所以当  $x \geq 4$  时, (1-3)式明显地恒为正值, 即(1-2)式中的函数在  $x \geq 4$  时为递增函数。而当  $x = 4$  时, 它是大于 0 的, 故当  $x \geq 4$  时, (1-2)式恒大于 0。同理, (1-1)式的左边当  $x \geq 4$  时为递增函数, 且当  $x = 4$  时, 它的值为  $2^4 - 4^2 = 0$ , 故当  $x \geq 4$  时, (1-1)式成立。最后, 将(1-1)式两边加上  $x^2$ , 即可得出结论。□

下一个例子体现了一个简单定理的完整演绎证明, 其中用到定理 1.1 的结果。

**定理 1.2** 如果  $x$  是 4 个正整数的平方和, 则  $2^x \geq x^2$ 。

**证明** 本定理的结论和定理 1.1 的结论相同, 只是前提不同, 因此首先考虑利用定理 1.1 的结果。如果我们能从本定理的前提推出定理 1.1 的前提, 问题就解决了。事实上,  $x$  是 4 个正整数的平方和, 则  $x$  至少是 4。当然, 这个简单的推理还可以分解成几个更明显的推理步骤。为了表现演绎推理的严格性, 我们把本定理的证明过程写成下列推理步骤:

命题	理由
(1) $x = a^2 + b^2 + c^2 + d^2$	定理的假设
(2) $a \geq 1, b \geq 1, c \geq 1, d \geq 1$	定理的假设
(3) $a^2 \geq 1, b^2 \geq 1, c^2 \geq 1, d^2 \geq 1$	由(2)和算术性质得出
(4) $x \geq 4$	由(1)、(3)和算术性质得出
(5) $2^x \geq x^2$	由(4)和定理 1.1 得出

在步骤(1)中, 将定理的假设“ $x$  是 4 个正整数的平方和”作为已知。这里给其中 4 个未命名的量(正整数)分别取名为  $a, b, c, d$ , 是为了与下面命题的衔接。在步骤(2)中, 继续以命题形式写出定理假设的其余内容, 即  $a, b, c, d$  均大于或等于 1(因为它们是正整数)。在步骤(3)中, 用到命题(2)作为已知, 同时根据最基本的算术性质: 如果某个数至少是 1, 则这个数的平方也至少是 1。步骤(4)用到命题(1)和(3)作为已知。命题(1)说,  $x$  是所讨论的 4 个数的平方和; 命题(3)说, 每个平方数至少是 1。根据众所周知的不等式的代入性质, 就得出:  $x$  至少是  $1+1+1+1$ , 即  $x \geq 4$ 。步骤(5)是最后一步, 它首先用到命题(4), 这已经是定理 1.1 的前提, 再用到定理 1.1 前提和结论的关系, 得出本定理的结论。至此, 定理 1.2 得以证明。 □

以上两个定理都是“如果……则”型的。下面再举一个“当且仅当”型推理的例子。“当且仅当”型实际上是两个方向的“如果……则”型。例如, 命题“ $A$  当且仅当  $B$ ”就是“如果  $A$  则  $B$ ”和“如果  $B$  则  $A$ ”两个命题的合成, 前者称为“仅当”部分( $A$  仅当  $B$ ), 后者称为“当”部分( $A$  当  $B$ )。在证明中都是要证这两个方面, 看下面简单的例子。

**定理 1.3** 设  $x$  是实数, 则  $\lfloor x \rfloor = \lceil x \rceil$  当且仅当  $x$  是整数(这里  $\lfloor x \rfloor$  代表实数  $x$  的向下取整, 即小于或等于  $x$  的最大整数;  $\lceil x \rceil$  代表实数  $x$  的向上取整, 即大于或等于  $x$  的最小整数)。

**证明** 先证“仅当”部分。设  $\lfloor x \rfloor = \lceil x \rceil$ , 要证  $x$  是整数。利用向下取整和向上取整的定义, 知道  $\lfloor x \rfloor \leq x$  和  $\lceil x \rceil \geq x$ 。但是, 已知  $\lfloor x \rfloor = \lceil x \rceil$ , 因此, 在第一个不等式中用向上取整代替向下取整, 就得到  $\lceil x \rceil \leq x$ 。由于  $\lceil x \rceil \leq x$  和  $\lceil x \rceil \geq x$  同时成立, 所以  $\lceil x \rceil = x$ , 即得出  $x$  是整数的结论。

再证“当”部分。现在假设  $x$  是整数, 要证  $\lfloor x \rfloor = \lceil x \rceil$ 。这是很明显的。根据向下取整和向上取整的定义, 对于整数  $x$ ,  $\lfloor x \rfloor$  和  $\lceil x \rceil$  都等于  $x$ , 自然  $\lfloor x \rfloor = \lceil x \rceil$ 。至此, 定理得以证明。 □

### 1.1.2 反证法

反证法也称归谬法, 是数学中常用的一种证明方法。下面举例说明反证法在某些场合是有很有效的, 甚至是很必要的。

**定理 1.4** 设有牛、羊、猪三种动物共 10 头, 则这三种动物中至少有一种动物不少于 4 头。

**证明** 假如采用直接的证明方法, 即把总共 10 头的这三种动物中每种动物出现的可能头数都列举出来, 就能一眼看出结论是否成立。现在把这些情况列表如下:

牛	10	9	9	8	8	8	7	7	7	7	...	4	...	0
羊	0	0	1	0	1	2	0	1	2	3	...	2	...	10
猪	0	1	0	2	1	0	3	2	1	0	...	4	...	0