



“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 编著

形式语言与自动机理论教学参考书(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

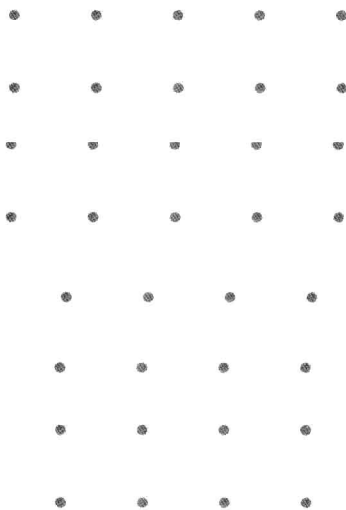


“十二五”普通高等教育本科国家级规划教材
普通高等教育精品教材

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 教学参考书 (第3版)

蒋宗礼 编著



清华大学出版社
北京

内 容 简 介

本书作为《形式语言与自动机理论(第3版)》(主教材)的配套教学辅导用书,按照主教材的结构编写而成。本书包括有关内容的讲解、学习要点、问题分析、求解思路和方法、注意事项。考虑到该课程习题求解具有相当的难度,以及给出全部习题解答又不利于学生学习,只给出了典型习题的解析。为了引导读者及时总结学习内容,按照小节给出知识点和主要内容解读,为读者学习和掌握主教材中的知识点和问题求解方法,体会问题求解的核心思想提供帮助,对教师和学生来说,阅读这些内容都是很有意义的。

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教学参考书/蒋宗礼编著. —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 978-7-302-31781-4

I. ①形… II. ①蒋… III. ①形式语言—高等学校—教学参考资料 ②自动机理论—高等学校—教学参考资料 IV. ①TP301

中国版本图书馆CIP数据核字(2013)第059889号

责任编辑:张瑞庆

封面设计:傅瑞学

责任校对:梁毅

责任印制:宋林

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:北京国马印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:13.75 字 数:342千字

版 次:2008年3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:26.00元

产品编号:050870-01

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(fundamentals)、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,这种优势将进一步的凸现。

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,一些相识的和不相识的读者对该书给予了充分的肯定,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。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

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,并且能够很好地掌握逻辑思维方法。我们称其为“计算思维”能力,或者叫做“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。它需要一系列课程,并且通过长期的学习和实践来完成。

形式语言与自动机理论不仅对问题及其求解提供了良好的形式化描述工具,更在通过适当的描述和解析而降低难度之后,成为对本科生和研究生进行“计算思维”能力培养的一门重要技术基础课程。

抽象和形式化是本书所涉及内容的主要特点。在这里,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、性质等。通过对本课程的学习,除了使学生掌握正则语言、下文无关语言的文法、识别模型及其基本性质、图灵机的基本知识以外,主要致力于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,我们就扣上了“什么能被有效地自动化”这一计算学科的主题。

当我们用计算机进行问题的求解时,需要实现对“问题”所在系统的状态及其变换的描述,要用适当的数据在计算机中表示该问题,并用适当的算法通过对这些数据的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。它们都是问题的数学模型化的典范,给计算机问题求解提供了一种优美而坚实的基础,而且,它也向人们展示了一种典型的方法和思想。另外,形式语言与自动机理论还是研究算法及其理论的基础。

形式语言与自动机理论对于一个计算机科学与技术工作者来说,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。ACM 和 IEEE CC2001 及《中国计算机科学与技术学科教程 2002》(简称为 CCC2002)给出了明确的要求。这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全部 3 个学科形态。它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,抓住问题及求解的关键点,以“计算机”的方式解决问题。

《形式语言与自动机理论》一书包括了 CC2001 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,《形式语言与自动机理论》一书主要突出了3个特点: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出进一步的尝试;③主要考虑国内读者的特点,并且按照国内的教学风格的要求讨论问题。本书将按照小节清晰地列举出相关知识点,给出主要内容的解读。通过这些,进一步地讨论讲解学习的要点、问题分析、求解思路和方法、注意事项等内容,为读者学习和掌握原书中的知识点和问题求解方法、体会问题求解的核心思想提供帮助。考虑到初学者在解答习题中将会遇到的主要困难,本书选择了一些典型习题,并且给出了解答及分析。

形式语言与自动机理论课程的教学,最大的问题有两个,一是内容非常抽象,这就导致阅读起来比较枯燥,而且它的作用主要是在潜移默化中体现的,难以让学生看出其“用处”,似乎让人感到学习这门课程是在“自讨辛苦”,而且这种“辛苦”没有太大的意义,不如学习Java语言等编程更容易,更实用;其二是这些内容具有较大的难度,难以找到体验感性认识的具体实例,这就导致读者难以发现相关知识点的来龙去脉,以达到深入领会之目的。要想解决这两个问题,必须掌握问题求解的思想和方法,并且通过对它们的研究,来领略这门课程在高度抽象和形式化下的优美和乐趣,使这些看似抽象枯燥的内容活起来。实际上,许多内容都可能是读者自己的体会,哪怕这些体会是不完善的,甚至是过于理想化(理性)的,与历史不十分符合的。但是,它们一定是更理性的“思路”和“想法”。实际上,这正是人们在科学研究中所努力追求的。

虽然目前在国内的计算机科学与技术学科的本科生的课程教学计划中,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还未开设此课程,但是,随着我国的计算机科学与技术学科教学的不断发展和条件的逐渐成熟,将会有越来越多的学校开设本课程。

本书共分12章。为了便于阅读,从第1章到第10章完全与原书结构相对应。第1章回顾在离散数学中学过的本书将要用到的一些基础知识,为后续的章节做好准备。由于主要是为了复习,所以这里给出的是知识点和应该注意的事项。除最后一节的“形式语言及其相关的基本概念”为新内容外,其他内容都可以由学生自学。第2章到第8章是教学的重点内容,主要讨论正则语言和上下文无关语言的文法、自动机描述及其性质。第9章对计算进行介绍,包括一般计算模型图灵机的概念、构造方法、修改,与计算相关的不可判定性、P-NP等问题。第10章介绍上下文有关语言。在这些章节中,以知识点、主要内容解读、典型习题解析的形式,对讨论的内容进行归纳和解析。第11章对本书的内容按类进行全面总结。在第12章中安排了教学设计,从总体上讨论本课程的讲授等问题。

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

2003年6月

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	3
1.1.3	集合的运算	4
1.2	关系	5
1.2.1	二元关系	5
1.2.2	递归定义与归纳证明	6
1.2.3	关系的闭包	6
1.3	图	7
1.3.1	无向图	7
1.3.2	有向图	8
1.3.3	树	8
1.4	语言	9
1.4.1	什么是语言	9
1.4.2	形式语言与自动机理论的产生与作用	10
1.4.3	基本概念	11
1.5	小结	14
1.6	典型习题解析	14
第 2 章	文法	20
2.1	启示	21
2.2	形式定义	22
2.3	文法的构造	26
2.4	文法的乔姆斯基体系	29
2.5	空语句	32
2.6	小结	33
2.7	典型习题解析	33

第3章 有穷状态自动机	45
3.1 语言的识别	46
3.2 有穷状态自动机	46
3.3 不确定的有穷状态自动机	51
3.3.1 作为对 DFA 的修改	51
3.3.2 NFA 的形式定义	51
3.3.3 NFA 与 DFA 等价	52
3.4 带空移动的有穷状态自动机	55
3.5 FA 是正则语言的识别器	57
3.5.1 FA 与右线性文法	57
3.5.2 FA 与左线性文法	58
3.6 FA 的一些变形	60
3.6.1 双向有穷状态自动机	60
3.6.2 带输出的 FA	61
3.7 小结	62
3.8 典型习题解析	63
第4章 正则表达式	70
4.1 启示	70
4.2 正则表达式的形式定义	71
4.3 正则表达式与 FA 等价	73
4.3.1 正则表达式到 FA 的等价变换	73
4.3.2 正则语言可以用正则表达式表示	75
4.4 正则语言等价模型的总结	77
4.5 小结	78
4.6 典型习题解析	79
第5章 正则语言的性质	83
5.1 正则语言的泵引理	84
5.2 正则语言的封闭性	85
5.3 Myhill-Nerode 定理与 DFA 的极小化	89
5.3.1 Myhill-Nerode 定理	89
5.3.2 DFA 的极小化	93
5.4 关于正则语言的判定算法	95
5.5 小结	96
5.6 典型习题解析	96

第 6 章 上下文无关语言	103
6.1 上下文无关文法	104
6.1.1 上下文无关文法的派生树	104
6.1.2 二义性	107
6.1.3 自顶向下的分析和自底向上的分析	109
6.2 上下文无关文法的化简	110
6.2.1 去无用符号	110
6.2.2 去 ϵ -产生式	113
6.2.3 去单一产生式	116
6.3 乔姆斯基范式	117
6.4 格雷巴赫范式	119
6.5 自嵌套文法	122
6.6 小结	123
6.7 典型习题解析	123
第 7 章 下推自动机	127
7.1 基本定义	128
7.2 PDA 与 CFG 等价	130
7.2.1 PDA 用空栈接受和用终止状态接受等价	130
7.2.2 PDA 与 CFG 等价	131
7.3 小结	133
7.4 典型习题解析	134
第 8 章 上下文无关语言的性质	140
8.1 上下文无关语言的泵引理	141
8.2 上下文无关语言的封闭性	144
8.3 上下文无关语言的判定算法	146
8.3.1 L 空否的判定	147
8.3.2 L 是否有穷的判定	147
8.3.3 x 是否为 L 的句子的判定	148
8.4 小结	150
8.5 典型习题解析	150
第 9 章 图灵机	152
9.1 基本概念	153
9.1.1 基本图灵机	153
9.1.2 图灵机作为非负整函数的计算模型	157
9.1.3 图灵机的构造	157
9.2 图灵机的变形	160

9.2.1	双向无穷带图灵机	160
9.2.2	多带图灵机	163
9.2.3	不确定的图灵机	164
9.2.4	多维图灵机	165
9.2.5	其他图灵机	166
9.3	通用图灵机	168
9.4	几个相关的概念	170
9.4.1	可计算性	170
9.4.2	P与NP相关问题	170
9.5	小结	171
9.6	典型习题解析	171
第10章	上下文有关语言	183
10.1	图灵机与短语结构文法的等价性	183
10.2	线性有界自动机及其与上下文有关文法的等价性	186
10.3	小结	187
10.4	典型习题解析	187
第11章	内容归纳	190
11.1	文法与语言	190
11.2	正则语言	190
11.3	上下文无关语言	191
11.4	图灵机	192
第12章	教学设计	194
12.1	概述	194
12.2	课程内容体系	195
12.2.1	课程的基本描述	195
12.2.2	教学定位	195
12.2.3	知识点与学时分配	196
12.3	讲授提示	198
12.3.1	重点与难点	198
12.3.2	讲授中应注意的方法等问题	201
12.4	习题与实验	202
12.4.1	指导思想	202
12.4.2	关于大作业和实验	202
12.5	考试与成绩记载	202
12.5.1	成绩评定	202
12.5.2	考题设计	203
参考文献		204

第 1 章

绪 论

计算机科学与技术学科,简称计算机学科,系统地研究信息描述及其变换算法,包括它们的理论、分析、效率、实现和应用。学科的根本问题是:什么能且如何被(有效地)自动计算。经过多年的发展,计算机学科已经发展成为计算学科(computing discipline)。该学科既研究计算领域中的一些普遍规律,描述计算的基本概念与模型,又研究包括计算机硬件、软件(系统软件和应用软件)在内的计算系统设计与实现的工程技术。理论和实践在该学科占有重要地位,其中的理论扮演着重要基础的角色。这可以从计算学科(计算机科学与技术学科)方法论中找到依据。另外,深入分析不难发现,即使像形式语言与自动机理论这样的内容,也同时具有抽象、理论、设计 3 个形态的内容。对不同类型学生的教学,可以通过强调不同形态的内容来达到教学目的。

众所周知,建立物理符号系统并对其实施变换是计算机学科进行问题的描述和求解的重要手段。“可行性”所要求的“形式化”及其“离散变换特征”使得数学成为重要工具。尤其是离散数学和计算模型无论从方法还是从工具等方面,更表现出它在计算学科中的直接应用。

虽然形式语言与自动机理论的论述只是用到集合、关系、图等基本概念,但是却不需要对这些基本概念进行过多的解释。因此,从知识的联系的角度来看,集合论和图论不一定要作为本课程的先修课。但是,从理解和掌握本课程的内容来讲,应该是在学习过集合论和图论,具有一定的知识基础和思维能力基础后,再开始本书内容的学习才是比较有利的。考虑到集合论和图论通常都被划入离散数学,所以,在本科生的教学计划中,形式语言与自动机理论被作为离散数学的后续课程。而如果是在研究生阶段学习形式语言与自动机理论,通常也假定学生具有离散数学的基本知识。为了平稳地过渡,本章首先简要回顾在离散数学中学过的部分基本概念和方法,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、有向图和树。这一部分内容分布在 1.1 节到 1.3 节。建议快速浏览这 3 节内容,以熟悉相应的表达方式。如果要介绍这一部分的内容,需要另外增加 4~6 个学时。

第二部分是关于形式语言及相关基本概念,包括字母表、字母及其特性、句子、出现、句子的长度、空语句、句子的前、后缀、语言及其运算。这一部分是本章的重点,属于本课程的正式内容。讲授这一部分的内容需要 2 个学时。

本章后面列举了大量的习题,主要用于使学生对所要求的内容进一步巩固和复习。在这些习题中,希望读者能够完成一些构造性题和证明题。关于语言的题目,应该尽可能多地完成。因为它们都涉及到最基本的训练。

1.1 集合的基础知识

无论是朴素集合论(set theory),还是公理化集合论,都是整个数学的基础。计算机科学与技术领域中的大多数基本概念和理论都采用与集合论有关的术语来描述。

1.1.1 集合及其表示

1. 知识点

(1) 集合:一定范围内的、确定的、并且彼此可以区分的对象汇集在一起形成的整体称为集合(set),简称为集。

(2) 元素:集合的成员为该集合的元素(element)。

(3) a 是集合 A 的一个元素:如果 a 是集合 A 的一个元素,则记为 $a \in A$,且称 a 属于 A ,或者 A 含有 a ;否则记为 $a \notin A$,且称 a 不属于 A ,或者 A 不含 a 。 $a \in A$ 读作 a 属于 A ; $a \notin A$ 读作 a 不属于 A 。

(4) 集合描述形式。

① 列举法(listing):将所有的元素逐一地列举在大括号 $\{\}$ 中,在能使读者立即看出规律时,某些元素可用省略号表示。

② 命题法(proposition):其基本形式为 $\{x|P(x)\}$,其中 P 为谓词,表示此集合包括所有使 P 为真的 x 。

(5) 多重集合:一个元素可以在同一个集合里重复出现。

(6) 基数: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)。集合 A 的基数又叫做集合 A 的势,一般用 $|A|$ 表示。对有穷集来说,它的基数就是它所包含的元素的个数。

(7) 集合的分类。

① 由有限个元素构成的集合叫做有限集(finite set),又称为有穷集。由无穷多个元素组成的集合叫做无穷集(infinite set)。

② 如果 $|A|=0$,则称 A 为空集(null set),一般用 \emptyset 表示。

③ 无穷集可以分成可数集(countable infinite set 或 countable set)和不可数集(uncountable set)。与自然数集对等的集合称为可数集。

(8) 整数集、有理数集是可数的,实数集是不可数的。实数集的不可数性质可以用著名的对角线法(diagonalization)进行证明。

2. 注意事项

本节回忆集合及其表示的基本内容,不用进一步扩展,而且在回忆中可随时以实际例子加以说明。注意以下表示集合及其元素的习惯。

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

\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。

1.1.2 集合之间的关系

1. 知识点

(1) P_1 是 P_2 的充要条件记为 $P_1 \Leftrightarrow P_2$, 或者 P_1 iff P_2 。

(2) 全称量词和存在量词。

“ $\forall x$ ”表示“对(论域中)所有的 x ”, “ $\exists x$ ”表示“(论域中)存在一个 x ”。

(3) 子集。

如果集合 A 中的每个元素都是集合 B 的元素, 则称集合 A 是集合 B 的子集(subset), 集合 B 是集合 A 的包集(container)。记作 $A \subseteq B$, 也可记作 $B \supseteq A$ 。 $A \subseteq B$ 读作集合 A 包含在集合 B 中; $B \supseteq A$ 读作集合 B 包含集合 A 。

如果集合 A 是集合 B 的子集 $A \subseteq B$, 且 $\exists x \in B$, 但 $x \notin A$, 则称 A 是 B 的真子集(proper subset), 记作 $A \subset B$ 。

(4) 集合相等。

如果集合 A, B 含有的元素完全相同, 则称集合 A 与集合 B 相等(equivalence), 记作 $A = B$ 。

2. 注意事项

(1) 对于集合的如下结论, 可以在回忆上述基本概念时穿插在其中考虑, 不用特意去证明, 在提到这些结论时, 可以以思考题的方式引导学生去思考, 并鼓励这方面知识不扎实的同学在课后自行努力完成几个证明。

对任意集合 A, B, C :

① $A = B$ iff $A \subseteq B$ 且 $B \subseteq A$ 。

② 如果 $A \subseteq B$, 则 $|A| \leq |B|$ 。

③ 如果 $A \subset B$, 则 $|A| < |B|$ 。

④ 如果 A 是有穷集, 且 $A \subset B$, 则 $|B| > |A|$ 。

⑤ 如果 $A \subseteq B$, 则对 $\forall x \in A$, 有 $x \in B$ 。

⑥ 如果 $A \subset B$, 则对 $\forall x \in A$, 有 $x \in B$ 并且 $\exists x \in B$, 但 $x \notin A$ 。

⑦ 如果 $A \subseteq B$ 且 $B \subseteq C$, 则 $A \subseteq C$ 。

⑧ 如果 $A \subseteq B$ 且 $B \subset C$, 或者 $A \subset B$ 且 $B \subset C$, 或者 $A \subset B$ 且 $B \subseteq C$, 则 $A \subset C$ 。

⑨ 如果 $A = B$, 则 $|A| = |B|$ 。

(2) 通过充要条件、存在量词、全称量词的使用, 告诉学生尽量用符号、式子去表达和叙述问题, 培养学生形式化表达问题的能力。

1.1.3 集合的运算

1. 知识点

(1) 并

将集合 A 的元素和 B 的元素放在一起构成的集合称为 A 与 B 的并(union), 记作 $A \cup B$ 。 $A \cup B = \{a \mid a \in A \text{ 或者 } a \in B\}$ 。

“ \cup ”为并运算符, $A \cup B$ 读作 A 并 B 。

设 A_1, A_2, \dots, A_n 是 n 个集合, 则它们的并 $A_1 \cup A_2 \cup \dots \cup A_n = \{a \mid \exists i, 1 \leq i \leq n, \text{ 使得 } a \in A_i\}$;

设 $A_1, A_2, \dots, A_n, \dots$ 是一个集合的无穷序列, 则它们的并 $A_1 \cup A_2 \cup \dots \cup A_n \cup \dots = \{a \mid \exists i, i \in \mathbb{N}, \text{ 使得 } a \in A_i\}$, 也可记为 $\bigcup_{i=1}^{\infty} A_i$ 。

当一个集合的元素都是集合时, 可以称为集族。设 S 是一个集族, 则 S 中的所有元素的并为:

$$\bigcup_{A \in S} A = \{a \mid \exists A \in S, a \in A\}$$

(2) 交

集合 A 和 B 中都有的所有元素放在一起构成的集合称为 A 与 B 的交(intersection), 记作 $A \cap B$ 。 $A \cap B = \{a \mid a \in A \text{ 且 } a \in B\}$ 。

“ \cap ”为交运算符, $A \cap B$ 读作 A 交 B 。

如果 $A \cap B = \emptyset$, 则称 A 与 B 不相交。

(3) 差

由属于 A , 但不属于 B 的所有元素组成的集合称为 A 与 B 的差(difference), 记作 $A - B$ 。 $A - B = \{a \mid a \in A \text{ 且 } a \notin B\}$ 。

“ $-$ ”为减(差)运算符, $A - B$ 读作 A 减 B 。

(4) 对称差

由属于 A 但不属于 B , 以及属于 B 但不属于 A 的所有元素组成的集合称为 A 与 B 的对称差(symmetric difference), 记作 $A \oplus B$ 。 $A \oplus B = \{a \mid a \in A \text{ 且 } a \notin B \text{ 或者 } a \notin A \text{ 且 } a \in B\}$ 。

“ \oplus ”为对称差运算符, $A \oplus B$ 读作 A 对称减 B 。

$$A \oplus B = (A \cup B) - (A \cap B) = (A - B) \cup (B - A)。$$

(5) 笛卡儿积

A 与 B 的笛卡儿积(cartesian product)是一个集合, 该集合是由所有这样的有序对 (a, b) 组成的: 其中, $a \in A, b \in B$, 记作 $A \times B$ 。 $A \times B = \{(a, b) \mid a \in A \text{ 且 } b \in B\}$ 。

“ \times ”为集合的笛卡儿乘运算符, $A \times B$ 读作 A 叉乘 B 。

(6) 幂集

A 的幂集(power set) $2^A = \{B \mid B \subseteq A\}$ 。

(7) 补集

补集又称为余集, 它是基于某个论域而言的。论域 U 中的、不在 A 中的所有元素组成的