NANKAI UNIVERSITY PRESS

Formal Languages and Automata Theory

形式语言与自动机

陈有祺 编著

南开大学出版社

形式语言与自动机

Formal Languages and Automata Theory

陈有祺 编著

社

形式语言与自动机

陈有祺　编著

南 开 大 学 出 版 社 出 版
（天津八里台南开大学校内）
邮编 300071 电话 23508542
新华书店天津发行所发行
南开大学印刷厂印刷

1999 年 4 月第 1 版　　1999 年 4 月第 1 次印刷
开本：850×1168　1/32　　印张：9
字数：220 千　　印数：1-1000

ISBN 7-310-01199-6
TP·97　定价：12.00 元

内 容 提 要

本书系统地论述了形式语言、自动机理论和有关的判定问题。全书共分 10 章。第 1 章介绍一些预备知识，其余各章分别论述四类形式语言和四类自动机以及它们的关系。其中第 8 章集中讨论不可判定性，第 10 章介绍确定的上下文无关语言和 LR(k) 文法。在每一章的最后，都附有不同难度的习题。

本书可作为计算机类各专业本科高年级学生和硕士研究生的教材，也可以作为计算机应用领域专业技术人员提高理论水平的参考书。

Abstract

Formal Languages, automata theory and relative decision problems are discussed systematieally in this book. This book is divided into 10 chapters. In the first chapter, some preliminarie are introduced. In the other chapters, four types of formal languages and automata and their relations are discussed respectively. We focus to discuss undecidability in the chapter 8, and introduce deterministic context-free languages and LR(k) grammars in the chapter 10. Some exercises with different levels are appended at the end of each chapter.

This book is intended as a textbook for the senior, or graduate studends in the computer specialities, and is also intended a reference book for technicians in the computer application areas to heighten their theoritical level.

前　言

形式语言理论和自动机理论是理论计算机科学的重要分支，自 60 年代以来它们的发展极为迅速，已经在计算机科学的许多领域起着理论基础和方法学的作用，特别是在程序语言的设计、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。因此，世界各发达国家的计算机界对此都十分重视，各主要大学的计算机科学系都把形式语言与自动机方面的内容列为本科高年级学生和研究生的主要课程。

本书编者在形式语言与自动机方面进行教学和研究工作已有二十多年的历史，在总结经验的基础上，参考了国内外有关著作，包括国外最近出版的教科书，写成这本教材。全书共分十章，选材力求简炼。其中第 1 章为预备知识，介绍一些本书最常用的基本概念和方法。第 2 章按 Chomsky 的体系，介绍四类形式文法。第 3 章和第 4 章，论述有穷自动机、正规表达式和正规文法的关系，以及正规集的性质。第 5 章和第 6 章，论述上下文无关文法与下推自动机的关系，以及上下文无关语言的性质。第 7 章介绍图灵机的各种等价模型，以及它与 0 型文法的关系。第 8 章集中论述不可判定问题。第 9 章介绍线性有界自动机和上下文有关语言的关系，并对各语言类之间的关系作一总结。第 10 章介绍确定的下推自动机与 LR(k)文法，以及它们之间的关系。本书如果作为本科生选修课的教材，第 8 章可以不讲，第 10 章也可以不讲或少讲。

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，一方面要逻辑严谨，思维严密；另一方面也要尽量采

用通俗易懂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,以达到提高理论素养和灵活运用的目的。书中每章最后都列有不同难度的一批习题,读者若能认真完成这些练习,则对掌握本书内容是大有帮助的。

在本书的编写和出版过程中,得到南开大学出版社的大力支持,在此表示衷心的谢意。由于编者的水平所限,书中的错误和不确切之处在所难免,恳请各位同行专家和广大读者批评指正。

陈有祺

1998年8月于南开大学

Contents

目 录

目录

Contents

Chapter 1
Preliminaries

第 1 章　预备知识

本书内容属于理论计算机科学的范畴，所需的数学基础知识较多。本章将对后面所需的基础知识作一扼要介绍，有些内容可能是读者已经学过的，这里作为复习；有些内容是为本书的讲述专门安排的，希望读者熟练掌握。

1.1 集合及其运算

由于集合是数学中最基本的概念之一，故无法对它下一个确切的数学定义，正如几何中无法定义点、直线一样。因此，我们只能对集合进行一些描述性的说明。

我们说一群无重复的对象的全体称为**集合**，而这些对象称为集合的**元素**。由此可见，集合由元素组成。元素与集合一样，也是无法定义的。实际上，世界上的任何事物，都可以作为原子并且由

这些原子组成集合，只是在不同的讨论场合，要对集合及其所包含的元素进行具体的指定。如地球上全体人类构成一个集合，全体大写的 26 个拉丁字母($A,B,\cdots,Z$)构成一个集合，全体自然数构成一个集合，等等。

对于集合的概念虽然不能精确定义，但是对于具体的集合，应尽量用明确的、形式化的方法来描述。对于元素个数较少的集合，我们可采用列举法，即将集合元素全部列出并括在一对花括号之中。例如 $A=\{a,b,c,d\}$，表示集合 A 由 a,b,c,d 四个元素组成。对于元素个数较多的集合或者由无穷多个元素组成的集合，可用集合形成模式 $\{x \mid P(x)\}$ 来描述，其中 x 表示该集合的任一元素，$P(x)$ 是一个谓词，它对 x 进行限定。$\{x \mid P(x)\}$ 指明由满足 $P(x)$ 的一切 x 构成的集合。至于 $P(x)$ 的形式，可以用明确无误的自然语言，也可以用大家熟知的数学表示法。例如，$\{n \mid n$ 是偶数$\}$，或者 $\{n \mid n \bmod 2=0\}$，都指明了一个由所有偶数组成的集合。

为简单起见，由有穷多个元素组成的集合称为**有穷集**，由无穷多个元素组成的集合称为**无穷集**。另外，一个常用的记号是 $\in$，若 a 是集合 A 的一个元素，则记作 $a \in A$；否则记作 $a \bar{\in} A$。我们还约定用 ϕ 表示**空集**，即不包含任何元素的集合。

在本节我们重点叙述一下集合运算和集合关系的有关概念。

定义 1.1 对于两个集合 A,B，若集合 A 的元素都是集合 B 的元素，则称集合 A 包含于集合 B 中(或称集合 B 包含集合 A)，记作 $A \subseteq B$，或 $B \supseteq A$，并且称 A 是 B 的**子集**。若 $A \subseteq B$，且 B 中至少有一个元素不属于 A，则称 A 真包含于 B 中(或称 B 真包含 A)，记作 $A \subset B$，或 $B \supset A$，此时称 A 是 B 的**真子集**。两个集合 A 和 B 相等，当且仅当 $A \subseteq B$ 且 $B \subseteq A$。

常用所谓的文(Venn)氏图来描述集合之间的关系。图 1.1 是 $A \subseteq B$ 的文氏图。

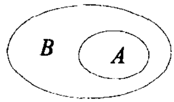

图 1.1　集合包含关系的文氏图，$A \subseteq B$

定义 1.2　集合 A 和 B 的并(或称作和)，记作 $A \cup B$，它是由 A 的所有元素和 B 的所有元素合并在一起组成的集合，即 $A \cup B = \{x \mid x \in A$ 或 $x \in B\}$。

集合 A 和 B 的交，记作 $A \cap B$，它是由 A 和 B 的所有公共元素组成的集合，即 $A \cap B = \{x \mid x \in A$ 且 $x \in B\}$。

集合 A 和 B 的差，记作 $A - B$，它是由属于 A 而不属于 B 的所有元素组成的集合，即 $A - B = \{x \mid x \in A$ 且 $x \notin B\}$。

若集合 $B \subseteq A$，我们也称 $A - B$ 为 B(关于 A)的补，记作 $\overline{B}(A)$。此时 A 称为 B 的全集。在全集已指明的情况下，也可直接说集合 B 的补，记作 $\overline{B}$。

关于集合运算的文氏图如图 1.2 所示。

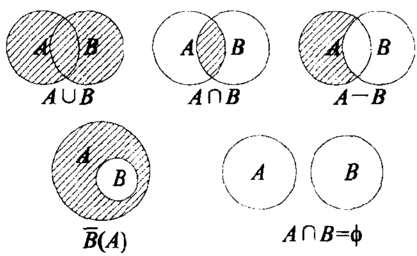

图 1.2　几种集合运算的文氏图

定义 1.3　设 I 是某些标号的集合，使得对 I 中的每个 i，A_i 是一个已知的集合，于是我们将 $\{x \mid$ 存在 $i \in I$，使得 $x \in A_i\}$ 写成 $\bigcup_{i \in I} A_i$。由于 I 可能是无穷的，所以这定义可以看成是两个集合的并的推广。

也可以由谓词 $P(i)$ 来定义 I，这时我们将 $\bigcup\limits_{i\in I}A_i$ 改写成 $\bigcup\limits_{P(i)}A_i$。例如，$\bigcup\limits_{i>2}A_i$ 意味着 $A_3\cup A_4\cup A_5\cdots$。

定义 1.4 设 A 为集合，A 的幂集为 A 的所有子集的集合，记作 2^A，即 $2^A=\{B\mid B\subseteq A\}$。

例 1.1 设 $A=\{a,b\}$，则 A 的幂集

$$2^A=\{\phi,\{a\},\{b\},\{a,b\}\}$$

注意，2^A 的元素都是集合，由于空集 ϕ 包含于任何集合中，故任何元素的幂集都包含 ϕ。

当 A 为有穷集时，设 A 的元素个数为 m，则 2^A 的元素个数（A 的所有子集的个数）为 2^m，这就是称 2^A 为 A 的幂集并且记为 2^A 的原因。当 A 为无穷集时，仍然用 2^A 表示 A 的幂集，当然它的元素个数也是无穷的。

定义 1.5 集合 A 和 B 的笛卡儿乘积用 $A\times B$ 表示，它是集合

$$\{(a,b)\mid a\in A \text{ 且 } b\in B\}$$

$A\times B$ 的元素均呈 (a,b) 的形式，称为有序对。在这些有序对中，总是 A 的元素在前，B 的元素在后，所以 $A\times B$ 与 $B\times A$ 一般是不相等的。

例 1.2 设 $A=\{1,2\}$，$B=\{a,b,c\}$，则

$$A\times B=\{(1,a),(1,b),(1,c),(2,a),(2,b),(2,c)\}.$$

定义 1.6 设 A 和 B 为集合，则由 A 到 B 的关系是 $A\times B$ 的任何子集。若 $A=B$，则称为 A 上的关系。若 R 为由 A 到 B 的关系，当 (a,b) 在 R 内时，可写成 aRb。

例 1.3 设 A 为整数集，则 A 上的关系“$<$”是集合 $\{(a,b)\mid a,b\in A, \text{且 } a<b\}$。

定义 1.7 设 R 是集合 A 上的关系。我们说：

(1)如果对 A 中任一元素 a，都有 aRa，则称 R 是自反的。