高等学校教材

形式语言

李敏生 编著

北京理工大学出版社

形式语言

李敏生　编著

北京理工大学出版社

内 容 简 介

本书是由工科电子类计算机教材编审小组征稿，推荐出版的。本书讨论了形式语言的基本概念、基本理论和基本方法以及它们在计算机技术中的某些应用。本书较详尽地论述了3型文法与有限自动机、2型文法与下推机，并简要地介绍了0型、1型文法与相应的自动机，形成了较完整的理论体系。本书不追求过分形式化地讨论，强调基本概念的直观背景和主要定理证明的思路分析。本书配有较多的例题和一定数量的习题，并附有部分习题的解法提示。读者对象为计算机专业本科生、研究生和有关科技人员。

形式语言

李敏生 编著

*

北京理工大学出版社出版

新华书店北京发行所发行 各地新华书店经售

国防科工委印刷厂印刷

*

850×1168毫米 32开本 10.25印张 264千字

1989年6月第一版 1989年6月第一次印刷

ISBN7-81013-112-5/TP·10

印数：1—2300册 定价：2.65元

出 版 说 明

根据国务院关于高等学校教材工作分工的规定，我部承担了全国高等学校、中等专业学校工科电子类专业教材的编审、出版的组织工作。由于各有关院校及参与编审工作的广大教师共同努力，有关出版社的紧密配合，从1978年至1985年，已编审、出版了两轮教材，正在陆续供给高等学校和中等专业学校教学使用。

为了使工科电子类专业教材能更好地适应“三个面向”的需要，贯彻“努力提高教材质量，逐步实现教材多样化，增加不同品种、不同层次、不同学术观点、不同风格、不同改革试验的教材”的精神，我部所属的七个高等学校教材编审委员会和两个中等专业学校教材编审委员会，在总结前两轮教材工作的基础上，结合教育形势的发展和教学改革的需要，制订了1986～1990年的“七五”(第三轮)教材编审出版规划。列入规划的教材、实验教材、教学参考书等近400种选题。这批教材的评选推荐和编写工作由各编委会直接组织进行。

这批教材的书稿，是从通过教学实践、师生反映较好的讲议中经院校推荐，由编审委员会（小组）评选择优产生出来的。广大编审者、各编审委员会和有关出版社为保证教材的出版和提高教材的质量，作出了不懈的努力。

限于水平和经验，这批教材的编审、出版工作还会有缺点和不足之处，希望使用教材的单位，广大教师和同学积极提出批评建议，共同为不断提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

前　言

本教材系按电子工业部的工科电子类专业教材1986—1990年编审出版规划，由工科电子类教材编审委员会计算机编审小组征稿，推荐出版，责任编委是南京工学院王能斌教授。

本教材由北京工业大学丘玉圃教授主审。

形式语言的内容十分丰富，有关的理论与方法发展迅速。作为教材，本书只讨论了形式语言的基本概念、基本理论和基本方法以及它们在计算机技术中的某些应用。

本书共七章。第一章是预备知识。第二章至第五章及第七章是讨论Chomsky四类文法所产生的语言和这些语言的识别装置。在这几章中，较详尽地论述了在计算机科学中有重大意义的3型文法与有限自动机、2型文法与下推机，并简要地介绍了0型、1型文法与相应的自动机。形成了较完整的理论体系。第六章是前几章的理论与方法在编译方面的一个应用。为避免过分形式化，本书强调基本概念的直观背景，主要定理证明的思路分析和各部分内容的内在联系，把形式语言的实际应用穿插在各章节中，并配有较多的例题和习题解法提示，使比较抽象的概念、理论易于理解与接受。

本书可作为计算机专业的高年级学生和研究生选修课的教材。本课程的参考学时数为40学时。

本书在编写过程中得到了许多同志的鼓励和帮助，王能斌教授、丘玉圃教授仔细阅读了全部书稿，提出了许多中肯的宝贵意见，纠正了一些错误，在此表示诚挚的感谢。我还要感谢88级研究生陈方和张翠琴，他们作了本书的全部习题。

由于编者水平有限，书中难免还存在一些缺点和错误，殷切希望广大读者批评指正。

编者　1988.8

目 录

第一章 预备知识

本章的目的是为以后各章的讨论提供一些预备知识。其内容包括两部分：一是关系(§1-1—§1-3)；二是半群(§1-4—§1-5)。在关系这一部分中，我们讨论两个集合之间的关系(即二元关系)及其性质；在半群这一部分中，先讨论映射与运算，之后再介绍半群及其性质。

熟悉上述内容的读者可略去本章，直接学习第二章。

§1-1 关系及其表示法

在现实生活中，人们常用到“关系”这个词，如人与人之间的关系，事与事之间的关系，人与事之间的关系，数与数之间的关系，等等。上述诸关系都可以抽象为集合与集合之间的关系。为了讨论集合与集合之间的关系，先介绍一些基本概念。

集合及其运算大家已熟悉，就不再讨论了。

一、集合的乘积与划分

设A、B是两个非空集合，若$a\in A$，$b\in B$，则称(a,b)是一个有序对。若a_1、$a_2\in A$，b_1、$b_2\in B$，当且仅当$a_1=a_2$、$b_1=b_2$时，则称有序对(a_1,b_1)、(a_2,b_2)相等，记作$(a_1,b_1)=(a_2,b_2)$。否则，称它们不等，记作$(a_1,b_1)\neq(a_2,b_2)$。显然，在一般情况下，$(a,b)\neq(b,a)$。

定义 1　集合的乘积

设A、B是两个非空集合，集合A和B的乘积是有序对(a,b)所组成的集合，其中$a\in A$、$b\in B$，记作$A\times B$，即

$$A\times B=\{(a,b)\mid a\in A,b\in B\}。$$

一般地，可考虑n个非空集合A_1、A_2、…A_n的乘积，用$A_1\times A_2\times\cdots\times A_n$表示。

例 1 设集合$A=\{1,2,3\}$，集合$B=\{r,s\}$，写出$A\times B$和$B\times A$。

解 $A\times B$的元素可以安排在一个如表1-1的数组形式的表中

表1-1

A \ B ($A\times B$)	r	s
1	$(1,r)$	$(1,s)$
2	$(2,r)$	$(2,s)$
3	$(3,r)$	$(3,s)$

这样，我们就有

$A\times B=\{(1,r),\ (1,s),\ (2,r),\ (2,s),\ (3,r),\ (3,s)\}$。

同样有

$B\times A=\{(r,1),\ (s,1),\ (r,2),\ (s,2),\ (r,3),\ (s,3)\}$。

显然$A\times B\neq B\times A$。

定义 2 集合的划分

非空集合S的划分是S的非空子集A_1、A_2、……所组成的集类，它满足下述条件

(1) $A_i\cap A_j=\phi$，$i\neq j$；

(2) 对于每个元素$s\in S$，s属于S的某个子集A_i。

A_i称作S的划分的块。

例 2 设$S=\{a,b,c,d,e,f,g,h\}$，令

$A_1=\{a,b,c,d\}$，　　$A_2=\{a,c,e,f,g,h\}$，

$A_3=\{a,c,e,g\}$，　　$A_4=\{b,d\}$，

$A_5=\{f,h\}$。求哪些是S的划分。

解 $\{A_1, A_2\}$不是S的一个划分，因为$A_1 \cap A_2 \neq \phi$；$\{A_1, A_5\}$也不是S的一个划分，因为$e \notin A_1$，$e \notin A_5$；而$\{A_3, A_4, A_5\}$是S的一个划分，这时我们说S被划分为三个块。

例 3 考虑某一个工厂的所有职工组成的集合S。按工资的级别把S分成若干个子集，在同一个子集内的人的工资级别相同。这样，按工资级别分成的S的这些子集，就构成了S的一个划分。

二、关系的概念

在这一部分中，我们主要讨论两个集合之间的关系，即二元关系。

两个集合之间可以定义很多关系。

例如，A表示家庭中男人的集合，B表示家庭中女人的集合。我们可以定义由A到B的父女关系，设此关系用F表示，若$x \in A$，$y \in B$，且x通过关系F与y发生联系，则可记作xFy，这表示x是y的父亲，y是x的女儿。同样，可以定义由B到A的母子关系，由A到A的父子关系，等等。

又如，设所有实数组成的集合为R_1，由R_1到R_1有很多关系。例如，"小于"关系，通常用"$<$"表示，若x小于y，则x通过关系"$<$"与y发生联系，记作$x<y$。同样，可以定义由R_1到R_1的大于($>$)，小于等于($\leqslant$)，大于等于($\geqslant$)，等于($=$)，不等于($\neq$)诸关系。

两个集合之间的关系，可以用语言来描述，也可以用字母来表示。在用字母表示关系时，两个集合之间的不同的关系要用不同的字母。

如何给两个集合之间的关系下一个严格的定义呢？为了探讨这个问题，让我们分析一个简单的例子。

设集合$A=\{1,2,3,4\}$，由A到A的关系R是小于关系。用R表示小于关系如下

$$1R2，1R3，1R4，2R3，2R4，3R4。$$

上述小于关系也可以表示为如下形式

$R=\{(1,2),\ (1,3),\ (1,4),\ (2,3),\ (2,4),\ (3,4)\}$。

显然，由A到A的小于关系R上述两种表示方法是等价的。从后一种表示方法中，我们看到：由A到A的关系R是$A\times A$的一个子集。这启发我们，用两个集合乘积的子集来定义两个集合之间的关系。

定义 3　二元关系

设A、B是两个非空集合，由A到B的二元关系R是$A\times B$的子集，即$R\subseteq A\times B$，二元关系R是用下述方法确定的

当$(a,b)\in R$时，我们说a是通过R与b发生关系，记作aRb；

当$(a,b)\notin R$时，我们说a是通过R与b没有关系，记作$a\not R b$。

由A到B的二元关系，简称由A到B的关系。

特别地，当$B=A$时，我们说$R\subseteq A\times A$是A上的关系，不必说R是由A到A的关系。

设$R\subseteq A\times B$是由A到B的关系，构成R的有序对的第一个元素所组成的集合称作R的定义域，用$\mathrm{Dom}(R)$表示，即

$$\mathrm{Dom}(R)=\{a\mid(a,b)\in R\}。$$

构成R的有序对的第二个元素所组成的集合称作R的值域，用$\mathrm{Ran}(R)$表示，即

$$\mathrm{Ran}(R)=\{b\mid(a,b)\in R\}。$$

例 4　设$A=\{1,2,3\}$，$B=\{r,s\}$，令

$R=\{(1,r),\ (2,s),\ (3,r)\}$，求$\mathrm{Dom}(R)$，$\mathrm{Ran}(R)$。

解　R是由A到B的关系，且

$$\mathrm{Dom}(R)=\{1,2,3\}=A,$$

$$\mathrm{Ran}(R)=\{r,s\}=B。$$

例 5　设$A=B=R_1$，R_1是所有实数组成的集合，我们定义R_1上的关系如下

$$aRb，当且仅当\ a=b。$$

显然，$R=\{(a,a)\mid a\in R_1\}$。实际上，这个关系就是高等数学中

的函数 $y=x$。

例6 设A是某个计算机所有可能的输入组成的集合，B是这个计算机所有可能的输出组成的集合。定义由A到B的关系R如下

aRb，当且仅当b是由输入a所产生的输出。

例7 现有为五个城市服务的航线，下面的表给出了由城市C_i到城市C_j的票价。由表1-2可以看出由城市C_1到城市C_3的票价为100元，由城市C_4到城市C_2的票价为200元，等等。

表1-2

发站＼到站	C_1	C_2	C_3	C_4	C_5
C_1		140	100	150	200
C_2	190		200	160	220
C_3	110	180		190	250
C_4	190	200	120		150
C_5	200	100	200	150	

在五个城市组成的集合$A=\{C_1,C_2,C_3,C_4,C_5\}$上定义关系$R$如下

C_iRC_j，当且仅当由城市C_i到城市C_j的票价小于等于180元。求关系R。

解 关系R是$A\times A$的子集，由(C_i,C_j)组成，且满足由C_i到C_j的票价小于等于180元，由上面的表可知

$$R=\{(C_1,C_2),\ (C_1,C_3),\ (C_1,C_4),\ (C_2,C_4),\ (C_3,C_1),\ (C_3,C_2),\ (C_4,C_3),\ (C_4,C_5),\ (C_5,C_2),\ (C_5,C_4)\}$$

三、关系矩阵与方向图

1. 关系矩阵 设$A=\{a_1,a_2,\cdots,a_m\}$，$B=\{b_1,b_2,\cdots,b_n\}$，

R是由A到B的关系。我们可以用$m\times n$的矩阵M_R表示R，M_R的第i行第j列的元素m_{ij}定义如下

$$m_{ij}=\begin{cases}1 & 当(a_i,b_j)\in R时\\ 0 & 当(a_i,b_j)\notin R时\end{cases}$$

矩阵M_R称作R的**关系矩阵**。

例 8 求本节例 4 中所定义的关系R的关系矩阵。

解

$$M_R=\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 0\end{bmatrix}$$

例 9 设布尔矩阵（即元素是 0 或 1 的矩阵）

$$M=\begin{bmatrix}1 & 0 & 0 & 1\\ 0 & 1 & 1 & 0\\ 1 & 0 & 1 & 0\end{bmatrix}$$

求一个关系R，使$M_R=M$。

解 由于M是3×4的布尔矩阵，我们可设

$$A=\{a_1,a_2,a_3\},\qquad B=\{b_1,b_2,b_3,b_4\}。$$

定义由A到B的关系如下

$$(a_i,b_j)\in R，当且仅当\ m_{ij}=1。$$

其中m_{ij}为M的第i行第j列的元素。这样定义的关系R，其关系矩阵$M_R=M$。

由上面的两个例子可见：若A、B都是有限集，R是由A到B的一个关系，那么R可以用其关系矩阵M_R来表示；反之，给了一个布尔矩阵M，我们可以构造两个有限集A、B，使得由A到B的一个关系R的关系矩阵M_R，恰好是所给的布尔矩阵M。

综上所述，我们得到如下结论：由有限集A到有限集B的任何一个关系R，都由其关系矩阵M_R唯一确定。

2．方向图 设$A=\{a_1,a_2,\cdots,a_m\}$，R是A上的关系。我们可以用一个图表示R：对A的每一个元素画一个小圆圈，每个圆圈称作一个顶点（或结点），若a_iRa_j，则从顶点a_i到顶点a_j画一条

边（有向线段）。这样得到的图称作关系 R 的**方向图**。

例10 设 $A=\{1,2,3,4\}$，A 上的关系 R 为

$R=\{(1,1),\ (1,2),\ (2,1),\ (2,2),(2,3),(2,4),\ (3,4),$
$(4,1)\}$。

求 R 的方向图。

解

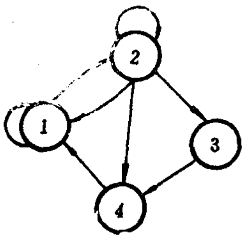

图1-1

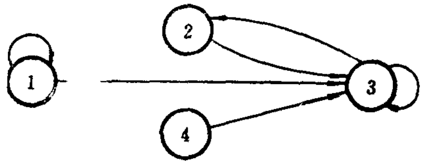

图1-2

例11 求关系 R，其方向图如图1-2所示。

解 根据方向图构成的规则知，上述方向图所表示的关系 R 如下

$R=\{(1,1),\ (1,3),\ (2,3),\ (3,2),\ (3,3),\ (4,3)\}$

由上面两个例子可以得到如下结论：有限集 A 上的任何一个关系 R，都可以由其方向图唯一确定。

§1-2 关系中的路

一、路的概念

定义1 长度为 n 的路

设 R 是集合 A 上的一个关系，由其方向图中的顶点 a 到顶点 b 长度为 n 的路，是一个有限序列 $l=a,x_1,x_2,\cdots,x_{n-1},b$，使得 aRx_1，$x_1Rx_2,\cdots,x_{n-1}Rb$，简称 l 为 R 中的路。a 为 l 的起始点，b 为 l 的终点。

注意，顶点a，x_1，x_2，…，x_{n-1}，b中可能有相同的。

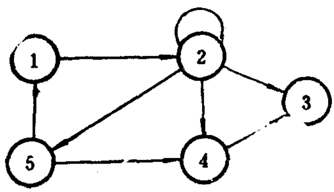

图1-3

例 1　考虑图 1-3 所示的方向图中的路。

其中 $l_1=1,2,5,4,3$ 是由顶点 1 到顶点 3 长度为 4 的路；$l_2=1,2,5,1$是由顶点 1 到它本身长度为 3 的路；$l_3=2,2$ 是由顶点 2 到它本身长度为 1 的路。

始、终点相同的路称作**回路**（或**环路**）。在上例中，l_2 和l_3分别是长度为 3 和 1 的回路。

设 n 是正整数，R是集合A上的关系。在A上可以定义关系R^n，xR^ny表示在R中由x到y有长度为n的路；在A上还可以定义关系R^∞，$xR^\infty y$表示在R中由x到y有一些路，这些路的长度为任意正整数，也就是说，在R中只要由x到y有路，就可以用 $xR^\infty y$ 表示。R^∞也称作R的**连接关系**。

例 2　设A是中国有航班的城市组成的集合，A上的关系R定义如下

xRy，当且仅当由x市到y市有直接航班。于是，xR^ny表示由x市到y市的航班，中间恰好有$n-1$个停留站；$xR^\infty y$表示由x市可乘飞机到达y市。

例 3　设 $A=\{1,2,3,4,5,6\}$，A上的关系R的方向图如图1-4所示，图1-5是A上的关系R^2的方向图。

在图1-5中，两个顶点之间有边，当且仅当它们有R^2关系，也就是说，在图1-4中，相应的两个顶点有长度为 2 的路。

由图1-4得到图1-5的依据是：

因为$1R2$和$2R2$，所以有$1R^22$；因为$1R2$和$2R3$，所以有$1R^23$；等等。

例 4　设 $A=\{a,b,c,d,e\}$，A上的关系R如下