自动机理论、语言和计算导引

〔美〕J. E. 霍普克罗夫特
J. D. 厄 尔 曼 著

科学出版社

自动机理论、语言和计算导引

〔美〕J. E. 霍普克罗夫特
J. D. 厄尔曼 著

徐美瑞 译
洪加威 校

科学出版社
1986

内容简介

本书是讨论自动机理论、语言理论和计算理论（主要是计算复杂性理论）的专著.

全书共十四章. 第一章为预备知识；第二、三章讨论有穷自动机和正规集合；第四、五、六章讨论上下文无关语言和下推自动机；第七章讨论图灵机；第八章讨论不可判定性；第九章按Chomsky谱系对语言和自动机加以总结，同时介绍了上下文有关语言和线性有界自动机；第十章讨论上下文无关语言的一种特殊情形——确定的上下文无关语言；第十一章讨论抽象语言族；第十二、十三章讨论计算复杂性理论；第十四章介绍了几种其它的重要语言类.

本书可作为计算机科学系的研究生和高年级大学生有关课程的主要参考书，也可供从事计算机科学、计算机软件以及硬件方面工作的研究人员、工程技术人员和大专院校教师参考.

John E. Hopcroft, Jeffrey D. Ullman

INTRODUCTION TO AUTOMATA THEORY, LANGUAGES, AND COMPUTATION

Addison-Wesley Publishing Company, 1979

自动机理论、语言和计算导引

〔美〕 J. E. 霍普克罗夫特 J. D. 厄尔曼 著

徐美瑞 译

洪加威 校

责任编辑 那莉莉

科学出版社出版

北京朝阳门内大街137号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

1986年9月第一版 开本：850×1168 1/32

1986年9月第一次印刷 印张：18 3/4

印数：0001—4?750 字数：444,000

统一书号：15031·737

本社书号：4171·15—8

定价： 4.75 元

译 者 序

理论计算机科学是一门处于迅速发展中的学科，它对计算机硬件工程和软件工程的发展有重要影响．本书内容涉及理论计算机科学的基础，也包含该学科中某些核心分支．

如书名所示，本书大体包含三部分：自动机理论、语言理论和计算理论．第二、五、七、九和十四章，分别讨论了有穷自动机、下推自动机、图灵机、线性有界自动机及其它若干重要的自动机；第二、三、四、六、九、十和十一章讨论了正规语言、上下文无关语言、上下文有关语言、确定的上下文无关语言及抽象语言族；第七、八、十二和十三章讨论了计算理论、特别是计算复杂性理论．

本书在理论计算机科学领域中是一本有影响的著作，出版于1979年，总结了到当时为止，在自动机、语言和计算理论方面所获得的重要成果．对从事这方面研究的计算机科研人员来说，本书有重要的参考价值．每章后的文献注记，叙述了主要结果的来源．书末附有大量参考文献，这些参考文献对研究工作的开展也是有利的．

本书还是一本很好的教科书．美国许多大学采用它作为高年级大学生和研究生课程的教材．只要适当组织本书内容，就可以形成下列课程的主要素材：形式语言与自动机、自动机理论、语言理论、计算理论及计算复杂性理论．每章都附有习题，按难度用星号加以分类．少数典型习题还有题解，这些都颇受师生欢迎．

Hopcroft 和 Ullman 曾在 1969 年出版过一本书，名叫"Formal Languages and Their Relation to Automata"(有中译本）．由于理论计算机科学在七十年代得到蓬勃发展，新思

想、新方法、新成果大量涌现，作者不得不重新写了这本新书。本书内容在深度和广度上已远远超过前书，且在系统和结构上，作者也重新作了考虑，从以形式语言为中心、兼及自动机，改成现在的自动机理论、语言理论和计算理论三部分。

由于译者水平所限，译文中难免出现不妥之处，敬请读者批评指正。

译　　者

1984年10月

序　　言

十年前，作者曾经写了一本书，它包括形式语言、自动机理论和计算复杂性方面的已知内容。现回想起来， 在该书的237页中，仅有少许重要结果被遗漏了。在写一本关于这一题材的新书时，我们发现，这个领域已在许多新方向上都有了很大的发展，以致不可能形成一个全面的概括。与其试图使本书成为百科全书式的，不如在我们编辑材料时，进行严格的筛选，仅采用该领域理论发展中核心的论题，或对工程应用有重大意义的论题。

在过去的十年中，有两个研究方向是十分重要的。一个是使用语言理论概念（如非确定性和复杂性谱系等）来证明某些实际问题的固有复杂性的下界，另一个是应用语言理论思想（如正规表达式和上下文无关文法等）来设计诸如编译程序和文本处理程序等软件。这两方面的发展都有助于形成本书的结构。

本书用法

两位作者都曾使用第一章到第八章作为大学高年级的教材，只是略去了其中第四章的关于固有多义性的内容和第八章的一部分内容。第七、八、十二和十三章形成了关于计算复杂性课程的核心。以第二章到第七章、第九章到第十一章和第十四章为基础，可以构成一个关于语言理论的高级教程。

习　　题

我们使用这样的约定：最难的问题被打上双星号，较难的问题被打上单星号。用S做记号的习题在该章末尾有解答。我们不打算提供一个答案集，而是有选择地解答了少数具有指导意义的习题。

致　谢

我们对下述各位表示感谢，感谢他们很有见地的评论和建议：A. Aho, N. Francez, J. Goldstine, J. Hartmanis, D. Maier, F. Springsteel 和 J. Valdes. 本书手稿是由康奈尔大学的 M. Olton 和 A. Roberts，以及普林斯顿大学的 G. Pecht 打印的。

J.E.H　（纽约　伊萨卡）
J.D.U　（新泽西　普林斯顿）
1979年3月

目　　录

第一章 预备知识

本章，我们将全面地考察一些基本的数学概念，这是理解本书内容所必需的。这些概念包括图、树、集合、关系、字符串、抽象语言和数学归纳法。我们还将对全书的内容进行简要的介绍并阐明本书的目的。对于具有所述数学基础知识的读者，可以跳到1.6节去了解关于本书目的的讨论。

1.1 字符串、字母表和语言

“符号”是一个抽象的实体，我们将不去形式地定义它，就象在几何学中对“点”和“线”不加定义一样。字母和数字是经常使用的一些符号的例子。**字符串**（或**字**）是并置起来的有穷符号序列。例如 a,b 和 c 是符号，$abcb$ 是字符串。字符串 w的**长度**（记作 $|w|$）是组成该字符串的符号的个数。例如，$abcb$ 的长度为4。空字符串记作$\in$，它由零个字符组成，于是$|\in|=0$.

字符串的**前缀**是该字符串领头的若干符号，**后缀**则是结尾的若干符号。例如，字符串 abc 具有前缀$\in,a,ab$ 和 abc；其后缀是 $\in,c,bc$和 abc。如果字符串的前缀（或后缀）不是字符串本身，则称之为**真前缀**（或**真后缀**）。

两个字符串的**连结**形成这样的字符串：先写第一个，接着写第二个，中间没有空格。例如，dog 和 house 的连结是 doghouse。用并置作为连结操作，亦即，如果w和x是字符串，那么wx 是这两个字符串的连结。空字符串是连结操作的单位元素，亦即，对于每个字符串 w，$\in w=w\in=w$.

字母表是符号的一个有穷集合。（形式）**语言**是一个符号选

自某个字母表的字符串的集合。空集ϕ和由空字符串组成的集合$\{\in\}$都是语言。注意，它们是不同的，后者有元素，而前者没有。字母表$\{0,1\}$上的回文（顺读和倒读都一样的字符串）的集合是一个无穷语言。这个语言中的某些元素是$\in$,0,1,00,11,010和1101011。注意，从技术上说，一个无穷符号集合上所有的回文不是一个语言，因为它的字符串不能由一个字母表完全地建立起来。

一个固定的字母表Σ上所有的字符串的集合是另一个语言。我们把这个语言记作Σ^*。例如，若$\Sigma=\{a\}$，则$\Sigma^*=\{\in,a,aa,aaa,\cdots\}$；若$\Sigma=\{0,1\}$，则$\Sigma^*=\{\in,0,1,00,01,10,11,000,\cdots\}$。

1.2 图和树

一个图由一个有穷顶点（或结点）集合V和一个称为边的顶点偶集合E组成，记作$G=(V,E)$。图1.1是一个这样的例子。这里$V=\{1,2,3,4,5\}$，而$E=\{(n,m)\mid n+m=4$或$n+m=7\}$。

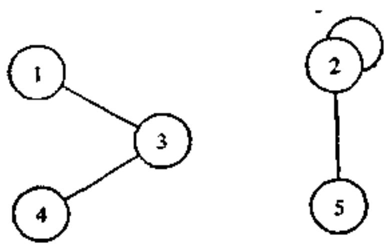

图1.1 图的例子

图中的一条路径是一个顶点序列$v_1,v_2,\cdots,v_k$，$k\geqslant1$，使得对于每个$i(1\leqslant i<k)$，(v_i,v_{i+1})都是一条边。这条路径的长度是$k-1$。例如，1,3,4是图1.1中的图的路径，5本身也是一条路径。如果$v_1=v_k$，则路径是一个循环。

有向图

一个有向图是由一个有穷顶点集合V和一个称为弧的有序顶点偶集合E组成的，记作 $G=(V,E)$. 我们用 $v\rightarrow w$ 表示由v到w的弧。一个有向图的例子如图1.2所示。

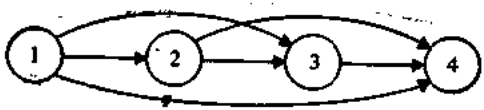

图1.2 有向图$(\{1,2,3,4\},\{i\rightarrow j\,|\,i<j\})$

有向图中的一条路径是一个顶点序列 $v_1,v_2,\cdots,v_k,k\geqslant 1$，使得对于每个 $i(1\leqslant i<k),v_i\rightarrow v_{i+1}$ 都是一条弧。我们还称该路径由 v_1 到 v_k. 因此，$1\rightarrow 2\rightarrow 3\rightarrow 4$是图1.2中有向图的由1到4的路径. 如果 $v\rightarrow w$ 是一条弧，我们说 v 是 w 的前趋，而 w是v 的后继。

树

一棵树（严格地说，一棵有序有向树）是一个具有下列性质的有向图：

1）有一个顶点，称作根，它没有前趋，由它出发到每一个顶点都有一条路径.

2）除根以外，每个顶点都恰有一个前趋。

3）每个顶点的后继都被"自左"排序.

我们将这样来画树，即根在顶上，所有的弧都指向下面. 因此弧上的箭头已无需用来表明方向了，故它们将不表示出来。每个顶点的后继按自左至右的顺序画出. 图1.3示出一棵树的例子，它是英语句子"The quick brown fox jumped over the lazy dog"的图解. 在这个例子中，顶点都没有命名，但有一个"标记"，它们或者是"字"，或者是语言的成分。

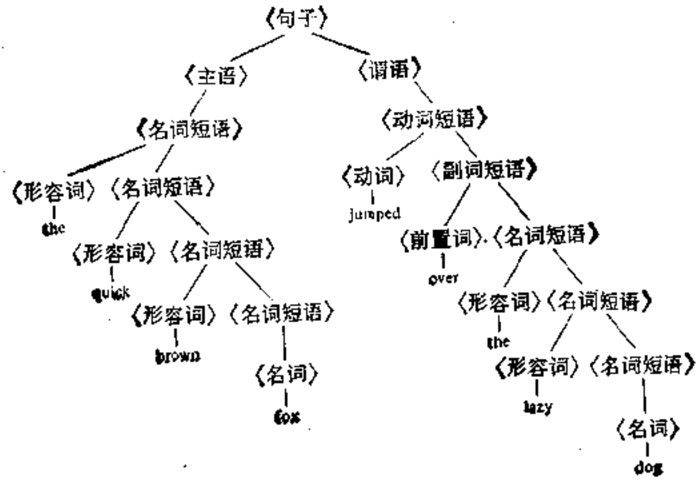

图1.3　一棵树

对于树，有一种特殊的术语，它与一般图中使用的术语不一样。顶点的后继称作**儿子**，前趋称作**父亲**。若有一条路径自v_1到v_2，那么v_1称作v_2的**先辈**，而v_2称作v_1的**后裔**。注意，不排除$v_1=v_2$的情况，任何顶点都是自己的先辈和后裔。没有儿子的顶点称为**叶**，其它顶点称为**内部顶点**。例如，在图1.3中，标着<动词>的顶点是标着<动词短语>顶点的儿子，后者则是前者的父亲。标着“dog”的顶点是它自身的后裔，也是标着<动词短语>的顶点和标着<句子>的顶点以及其它六个顶点的后裔。用英语的字来标记的顶点都是叶。那些用包在角括号中的语言成分标记的顶点则是内部顶点。

1.3　归纳证明

本书中许多定理是用数学归纳法证明的。假定我们有一个关于非负整数n的命题$P(n)$。一个经常选用的例子是取$P(n)$为：

$$\sum_{i=0}^{n} i^2 = \frac{n(n+1)(2n+1)}{6} \tag{1.1}$$

由数学归纳法的原理，P(n)是由

(a) P(0)，

(b) 对于 $n \geqslant 1$，P($n-1$)推出 P(n)。在归纳证明中，条件(a)称为基始，条件(b)称为归纳步骤。(b)的左部 P($n-1$)称为归纳假设。

例1.1 我们用归纳法证明（1.1）式。在（1.1）式中用 0 替换n，可看到两边都为 0，于是我们证实了 P(0)。为了证明(b)，在（1.1）式中用 $n-1$ 替换 n，并力求由这个结果证明（1.1）式。即必须证明，对于 $n \geqslant 1$，

$$\sum_{i=0}^{n-1} i^2 = \frac{(n-1)n(2n-1)}{6} \text{推出} \sum_{i=0}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

因为

$$\sum_{i=0}^{n} i^2 = \sum_{i=0}^{n-1} i^2 + n^2$$

又因为已给定

$$\sum_{i=0}^{n-1} i^2 = \frac{(n-1)n(2n-1)}{6}$$

我们只须证明

$$\frac{(n-1)n(2n-1)}{6} + n^2 = \frac{n(n+1)(2n+1)}{6}$$

这个等式可通过简单的代数计算得到，因此证明了（1.1)式。

1.4 集 合

假定读者熟悉集合的概念，即集合是由一群无重复的客体（它们被称为集合的元素）组成。有穷集合可以通过在括号中列

出其元素的办法来加以规定。例如，用$\{0, 1\}$来表示符号0和1的字母表。我们也可以用集合形成模式

$$\{x \mid P(x)\} \tag{1.2}$$

或

$$\{A\text{ 中的}x \mid P(x)\} \tag{1.3}$$

来规定集合。(1.2) 式读作“使得 $P(x)$ 为真的那些 x 的集合”，这里 $P(x)$ 是关于 x 的某个命题。(1.3) 式是“集合 A 中使得 $P(x)$ 为真的那些 x 的集合”。(1.3) 式等价于$\{x \mid P(x)\text{且 } x \text{ 在 } A\text{中}\}$。例如，

$$\{i \mid i\text{ 是一个整数且存在一个整数 } j\text{，使得 } i=2j\}$$

就是规定偶数的一种方法。

若A的每个元素都是B的元素，则可写成$A\subseteq B$，且称A被包含在B中。$A\supseteq B$与$B\subseteq A$同义。如果$A\subseteq B$且$A\neq B$，即A的每个元素在B中，而有某些B中的元素不在A中，那么写成$A\subsetneq B$。集合A和B相等，如果它们有相同的元素。$A=B$，当且仅当$A\subseteq B$且$B\subseteq A$。

集合的运算

在集合上定义的常见的运算是：

(1) $A\cup B$(A和B的并)是

$\{x \mid x\text{在}A\text{中或 } x \text{ 在 } B\text{中}\}$

(2) $A\cap B$ (A和B的交)是

$\{x \mid x\text{在}A\text{中且}x\text{在}B\text{中}\}$

(3) $A-B$(A和B的差)是

$\{x \mid x\text{在 }A\text{中但 } x \text{ 不在 } B\text{中}\}$

(4) $A\times B$(A和B的笛卡尔积)是有序偶(a,b)的集合，其中 a 在A中，而 b 在B中。

(5) 2^A(A的幂集)是A的所有子集的集合。

例1.2 设 $A=\{1,2\}$，$B=\{2,3\}$，则

$$A\cup B=\{1,2,3\},\ A\cap B=\{2\},\ A-B=\{1\}$$