普通高等教育“十一五”国家级规划教材 计算机系列教材

编译原理学习指导与习题解析

陈英 玉贵珍 主编
计卫星 主审

清华大学出版社

普通高等教育“十一五”国家级规划教材　计算机系列教材

陈英　王贵珍　主编
计卫星　李侃　陈朔鹰　编著

编译原理学习指导与习题解析

清华大学出版社
北京

内 容 简 介

本书是《编译原理》(陈英、陈朔鹰主编,清华大学出版社出版)的配套参考书。为便于结合教学使用,本书各章内容及名称与主教材一致,每章分为"学习要点指导"、"习题"和"习题参考答案与解析"三部分。"学习要点指导"部分对每章知识进行了归纳和总结,使之简明扼要、重点突出;"习题"部分不仅覆盖主教材的习题,还注重提炼精华,选编了近500道各种层次、各种类型的习题,设置了单项或多项选择题、填空题、判断题、简答题和解答题等多种题型,对有一定难度的题目加注了星号;"习题参考答案与解析"部分给出了全部习题的参考答案,给出习题分析、解答的步骤,并对习题所涵盖的重要知识点、难点和重点进行了说明。

本书可以作为计算机学科类专业及相关专业本科和研究生编译原理的学习用书,也适合作为课程考试和研究生考试辅导书及任课教师的教学参考书。

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

编译原理学习指导与习题解析/陈英,王贵珍主编;计卫星,李侃,陈翔鹰编著. —北京:清华大学出版社,2011.3
(计算机系列教材)
ISBN 978-7-302-24668-8

Ⅰ.①编… Ⅱ.①陈… ②王… ③计… ④李… ⑤陈… Ⅲ.①编译程序—程序设计—高等学校—教学参考资料 Ⅳ.①TP314

中国版本图书馆CIP数据核字(2011)第014754号

责任编辑:谢 琛 李 晔
责任校对:白 蕾
责任印制:杨 艳

出版发行:清华大学出版社 **地 址**:北京清华大学学研大厦A座
http://www.tup.com.cn **邮 编**:100084
社 总 机:010-62770175 **邮 购**:010-62786544
投稿与读者服务:010-62795954,jsjjc@tup.tsinghua.edu.cn
质 量 反 馈:010-62772015,zhiliang@tup.tsinghua.edu.cn
印 刷 者:三河市君旺印装厂
装 订 者:三河市新茂装订有限公司
经 销:全国新华书店
开 本:185×260 **印 张**:13 **字 数**:309千字
版 次:2011年3月第1版 **印 次**:2011年3月第1次印刷
印 数:1~3000
定 价:25.00元

产品编号:040844-01

编译原理课程具有理论性、综合性和实践性较强的特点。因此，系统地掌握这些原理、方法和技术，强化对核心概念和知识点的掌握，进一步提高求解问题的技巧与方法，融会贯通地掌握编译原理与技术，切实提高分析问题和解决问题的能力，是本书的目的所在。

本书是《编译原理》(陈英、陈朔鹰主编，清华大学出版社出版)的配套参考书。本书的组织与主教材的知识体系保持一致，每章分为“学习要点指导”、“习题”和“习题参考答案与解析”三部分。“学习要点指导”部分紧密结合教学目标和特点，紧扣教材的重点内容，对每章知识进行了归纳和总结，使之简明扼要、重点突出；“习题”部分不仅覆盖主教材的习题，还注重提炼精华，选编了近500道各种层次、各种类型的习题，设置了单项或多项选择题、填空题、判断题、简答题和解答题等多种题型，对有一定难度的题目加注了星号；“习题参考答案与解析”给出了全部习题的参考答案，详尽给出了解题的思路及过程，并对习题所涵盖的重要知识点、难点和重点进行了说明，方便学生复习和自学，便于启迪读者思维，提高解题能力。

本书是作者多年教学研究、教学实践和经验的汇集、提炼和整理。完成本书的责任编著和辅助编著的直接承担者是：陈英第1、4、5、8章；王贵珍第2、6、7章；李侃第6、7、2章；计卫星第8、5、4、3章；陈朔鹰第3、2、1章，商建云参加了本书的策划和第3章的编写，刘庆晖对本书部分章节进行了审核。本书的编写参考、借鉴了一些业内专家学者、同行的相关资料，以求取长补短相得益彰；另外，本书中许多习题的解析亦集成并体现了数届学生对编译原理课程的关注和习题训练中许多新颖的思路及反馈意见和建议；本书完成过程中，得到了清华大学出版社的鼎力协助，尤其是责任编辑高效的工作和非常专业的指导，作者在此一并深表致谢。

鉴于作者水平有限，本书稿虽经审慎校阅，仍难免存在疏误，敬请读者不吝赐教。

编　者

2011年2月于北京理工大学

《编译原理学习指导与习题解析》 目录

试读结束：需要全本请在线购买：　www.ertongbook.com

第 1 章　编译程序基本概念

1.1　学习要点指导

1. 编译程序的定义

把用某一种程序设计语言编写的源程序翻译成等价的另一种语言程序(目标程序)的程序，称之为编译程序。

编译程序定义的要点是：编译程序是一个程序。编译程序属于系统软件。编译程序的处理对象是源程序，处理结果是目标程序，编译程序的功能是实现从源程序到目标程序的等价变换。这里等价的含义是指源程序和目标程序的动态语义是一样的。

2. 源程序的编译和执行

一个源程序编写后要在计算机上运行，编译程序支持的执行过程分为两个阶段，即编译阶段和运行阶段，如图 1-1 所示。编译阶段对整个源程序进行分析，翻译成等价的目标程序，然后在运行子程序的支持下在目标机上运行。运行子程序是为了支持目标的运行而开发的程序，例如有系统提供的标准函数及其他目标程序所调用的程序等。

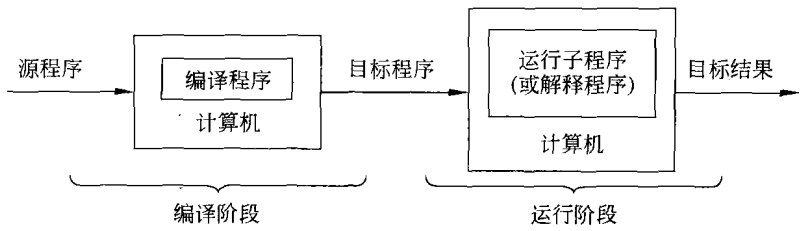
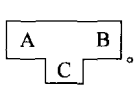


图 1-1　源程序的编译与运行

3. 编译程序的表示

一个编译程序可以用 3 种语言来刻画，即源语言、目标语言和宿主语言。对编译程序常用的表示形式是：函数表示、符号表示和 T 型图表示。

例如，若一个编译程序的源语言是 A 语言，目标语言是 B 语言，实现语言是 C 语言，则该编译程序对应的 3 种表示为：

函数表示：B=C(A)；符号表示：C_C^{AB}；T 型图表示： 

。

4. 编译程序的组成结构

1）经典结构

编译程序的处理过程复杂，且不同的编译程序实现方法千差万别，构造原理各异，但任何编译程序要完成的基本任务都是类似的，图 1-2 给出了编译程序总体结构的经典表示。

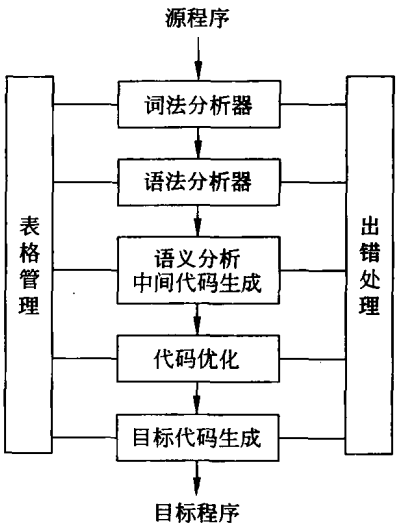


图 1-2 编译程序的总体结构

2）核心功能程序

编译程序的核心功能程序包括词法分析、语法分析、语义分析与中间代码生成、代码优化及目标代码生成。

词法分析的任务是对输入的符号串形式的源程序进行最初的加工处理。它依次扫描读入的源程序中的每个字符，根据源语言的词法规则识别出源程序中有独立意义的单词，用某种特定的数据结构对它的属性予以表示和标注。

语法分析的任务是：在词法分析基础上，依据源语言的语法规则，对词法分析的结果进行语法检查，并识别出单词符号串所对应的语法范畴。

语义分析与中间代码生成的任务是：依据源语言的语义规则对语法分析所识别的语法范畴进行语义检查并分析其含义，翻译成与其等价的中间代码。

代码优化是为了改进目标代码的质量而在编译过程中进行的工作。代码优化可以在中间代码或目标代码级上进行，其实质是在不改变源程序语义的基础上对其进行加工变换，以期获得更高效的目标代码。而“高效”一般是指，对所产生的目标程序缩短其运行时间和节省存储空间。

目标代码生成的功能是：根据中间代码及编译过程中产生的各种表格的有关信息，最终生成所期望的目标代码程序。

3）公共程序

编译程序的公共程序一般包括表格管理与出错处理。

（1）表格与表格管理

编译程序在对源程序的处理过程中，为了记录源程序中的数据实体的有关信息和编译各阶段所产生的信息，以利于完成从源程序到目标程序的等价变换，需要创建和管理一系列表格。例如，编译程序需要记录源程序中变量的类型、数组的大小、函数的参数个数和类型等，这些信息的采集和记录，一般可以随着编译过程的需要建表、查表和填表，或修改表中某些数据，或从表中取得有关信息，支持编译的全过程。因此合理的设计和使用表格，构造高效的表格管理程序是编译程序设计和实现的重要任务。

（2）出错处理

编译程序对源程序中可能存在的错误（如词法错误、语法错误、语义错误等）进行检

查、分析和报告，并尽可能使编译继续进行。一个性能好、效率高的编译程序应该能够协助程序员及时准确地发现源程序中的错误，以提高调试程序的效率，方便用户修改程序，并能把错误限制在尽可能小的范围里。这方面的任务由编译程序的出错处理程序来完成。

4）遍

“遍”的概念是编译程序组织中的一个重要概念。“遍”是指对源程序或源程序的中间形式从头到尾扫描一遍，并做有关的分析加工，生成新的源程序的中间形式或生成目标程序，各遍之间通过临时文件相关联。

5. 重点概念及术语

（1）源语言、源程序：源语言是用来编写程序的语言，用源语言编写的程序称为源程序。

（2）目标语言、目标程序：目标程序是编译程序翻译源程序的结果，目标语言是用来编写目标程序的语言。目标语言实际可以是高级程序设计语言，可以是汇编语言，也可以是编译程序所运行的计算机的机器语言。但是，现在一般高级程序语言的编译程序的目标程序往往不是直接翻译到机器语言表示的目标程序。

（3）宿主语言：编译程序实现的语言或编写编译程序的语言称为宿主语言。

（4）宿主机（目标机）：编译程序所运行的计算机通常称为该编译程序的宿主机或目标机。

（5）汇编程序、解释程序：汇编程序和解释程序实际都属于编译程序，仅是从源语言类型或实现机制不同的角度对其的一种分类。

具体而言，汇编语言是计算机语言的符号形式。若源程序用汇编语言编写，经翻译生成的是机器语言表示的目标程序，该翻译程序称为“汇编程序”。而解释程序接收的是某一种语言的源程序（或经翻译生成的中间代码程序）直接在机器上解释执行的一类翻译程序。

1.2 习题

1-1 选择、填空题

（1）构造一个编译程序的三要素是_____，_____，_____。

（2）被编译的语言为 A 语言，编译的最终结果为 B 语言代码，编写编译程序的语言为 C 语言。那么，_____语言是源语言，_____语言是宿主语言，_____语言是目标语言。

（3）下面对编译原理的有关概念描述正确的是_____。

A. 目标语言只能是机器语言　　B. 编译程序处理的对象是源语言

C. Lex 是语法分析自动生成器　　D. 解释程序属于编译程序

（4）_____不是编译程序的组成部分。

A. 词法分析程序　　B. 代码生成程序

C. 设备管理程序　　D. 语法分析程序

（5）下面对编译程序分为“遍”描述正确的是_____。

A. 分“遍”可以使编译程序结构清晰　　B. 可以提高程序的执行效率

C. 可以提高机器的执行效率　　D. 可以增加对内存容量的要求

(6) 编译程序各阶段的工作都涉及到_____,_____。

A. 表格管理　　B. 语法分析　　C. 出错处理　　D. 代码优化

(7) 编译程序的生成方式可以是_____,_____,_____。

A. 自编译　　B. 高级程序设计语言编写

C. 完全自动生成　　D. 汇编语言缩写

(8) 设有表达式 a * b－c,将其中 a * b 识别为表达式的编译阶段是_____。

A. 词法分析　　B. 语法分析　　C. 语义分析　　D. 代码生成

(9) 设一个编译器接收的源语言 A,目标语言为 B,宿主语言为 C,则该编译器的符号表示是_____。

(10) 下面对编译程序分"遍"应考虑的因素描述不正确的是_____。

A. 源语言的特征和约束　　B. 代码优化的因素

C. 编译程序的功能　　D. 目标代码的选择

1-2 判断题

(1) 解释执行与编译执行的根本区别在于解释程序对源程序没有真正进行翻译。(　　)

(2) 宿主语言是目标机的目标语言。(　　)

(3) 具有优化功能的编译器可以组织为一遍扫描的编译器。(　　)

(4) 编译程序是将用某一种程序设计语言编写的源程序翻译成等价的另一种语言程序(目标程序)。(　　)

(5) 编译程序是应用软件。(　　)

(6) 编译程序的基本组成中,词法分析、语法分析和语义分析应该是有序的。(　　)

(7) "遍"是指对源程序的从头到尾扫描。(　　)

(8) 用高级语言书写的源程序都必须通过翻译,产生目标代码后才能运行。(　　)

(9) 含有优化功能的编译程序执行效率高。(　　)

(10) 解释程序和编译程序的不同在于,解释程序根据语法翻译成目标代码并立即执行之,而编译程序需产生中间代码及优化。(　　)

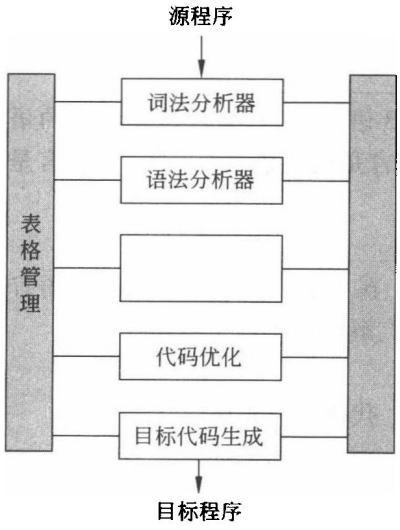


图 1-3　编译程序的总体结构

1-3 简答题

(1) 什么是编译程序?

(2) 源程序的编译执行和解释执行的主要区别是什么?

(3) 典型的编译程序在逻辑功能上由哪几部分组成?各部分的功能是什么?

1-4 图 1-3 示出了编译程序的组成结构,请填写完整。

1-5 编译程序实现的途径有哪些?

1-6* 用 T 型图描述将 A 机器上已经存在的 L 语言的编译程序移植到 B 机器上的过程。

1-7* 为不同目标机编写相同源语言的编译器

时，其设计变化最大的是后端，为什么？

1-8* 简述编译程序中“出错处理”程序的作用。

1-9* 为不同目标机编写相同源语言的编译器时，其设计变化最大的是后端，为什么？

1-10* 简述编译程序中“出错处理”程序的作用。

1.3 习题参考答案与解析

1-1 【解答】

(1) 源语言，目标语言，编译方法、技术及工具

【分析】 构造一个编译程序应从源语言、目标语言和编译方法、技术及工具三个方面入手。

① 源语言

这是编译程序处理的对象。要深刻理解所翻译的源语言的结构、词法、语法和语义规则，以及有关的约束和特点。

② 目标语言与目标机

这是编译程序处理的结果和运行环境。若选用机器语言作为目标语言，更需深入了解目标机的软件、硬件的有关资源、环境及特点。

③ 编译方法、技术及工具

这是生成编译程序的关键。应考虑与既定的源语言、目标语言相符合的编译方法与实现技术，使得编译程序设计合理，构造方便，同时兼顾考虑时间、空间上的高效率及实现的可能性和代价等诸多因素，并应尽可能地考虑使用先进、方便的生成编译程序的工具。

(2) A(语言) C(语言) B(语言)

(3) D

【分析】 对选项 A：目标语言可以是目标机上已有的其他语言。

对选项 B：应为源程序。

对选项 C：Lex 是词法分析自动生成器。

(4) C

(5) A

(6) A、C

【分析】 表格管理与出错处理是编译程序的公共程序。

(7) A、B、D

【分析】 编译程序的生成方式是指实现编译程序的途径。目前来讲，一个高级程序设计语言的编译程序完全自动生成尚受到技术的制约，例如，语义形式化描述问题，与特定目标机有关的代码优化和翻译问题等。

(8) B

【分析】 语法分析完成对源程序中句子或短语(也称为语法范畴)的识别，表达式 a * b 是一个语法范畴。

试读结束：需要全本请在线购买： www.ertongbook.com

(9) C_C^{AB}

(10) C

1-2 【解答】

(1) 错。

(2) 错。

【分析】 可以是目标机上已经存在的其他高、中级语言，不一定是机器语言。

(3) 错。

【分析】 一般代码优化需要对被编译的程序实施控制流、数据流分析的前期技术准备，然后进行优化实施。因此“一遍”的编译器难于完成。

(4) 错。

(5) 错。

(6) 对。

(7) 错。

(8) 对。

(9) 错。

【分析】 优化是针对被编译的程序的运行效率而言的，不是编译程序自身的运行效率而言的。

(10) 错。

【分析】 参考题 1-3(2)题的解答。

1-3 (1)**【解答】** 把用某一种程序设计语言编写的源程序翻译成等价的另一种语言程序(目标程序)的程序，称之为编译程序。

(2)**【解答】** 一般编译程序从对源程序执行途径的角度不同，可分为解释执行和编译执行。

所谓解释执行是借助于解释程序完成，即按源程序语句运行时的动态结构，直接逐句地边分析边翻译并执行。像自然语言翻译中的口译，随时进行翻译。所谓编译执行，是将源程序先翻译成一个等价的目标程序，然后再运行此目标程序，故编译执行分为编译阶段和运行阶段。

两种执行方式的主要区别是：编译执行是由编译程序生成一个与源程序等价的目标程序，它可以完全取代源程序，目标程序可运行任意多次，不必依赖编译程序。正像自然语言翻译中的笔译，一次翻译可多次阅读。而解释执行不生成目标程序，对源程序的每次执行都伴随着重新翻译的工作，而且不能摆脱翻译程序。

(3)**【解答】** 典型的编译程序在逻辑功能上由词法分析、语法分析、语义分析与中间代码生成、代码优化及目标代码生成五部分组成。各部分的简要功能是：

① 词法分析的任务是对输入的符号串形式的源程序进行最初的加工处理。它依次扫描读入的源程序中的每个字符，根据源语言的词法规则识别出源程序中有独立意义的单词，用某种特定的数据结构对它的属性予以表示和标注。

② 语法分析的任务是：在词法分析基础上，依据源语言的语法规则，对词法分析的结果进行语法检查，并识别出单词符号串所对应的语法范畴。

③ 语义分析与中间代码生成的任务是：依据源语言的语义规则对语法分析所识别的语法范畴进行语义检查并分析其含义，翻译成与其等价的中间代码。

④ 代码优化是为了改进目标代码的质量而在编译过程中进行的工作。代码优化可以在中间代码或目标代码级上进行，其实质是在不改变源程序语义的基础上对其进行加工变换，以期获得更高效的目标代码。而“高效”一般是指，对所产生的目标程序缩短其运行时间和节省存储空间。

⑤ 目标代码生成的功能是：根据中间代码及编译过程中产生的各种表格的有关信息，最终生成所期望的目标代码程序。

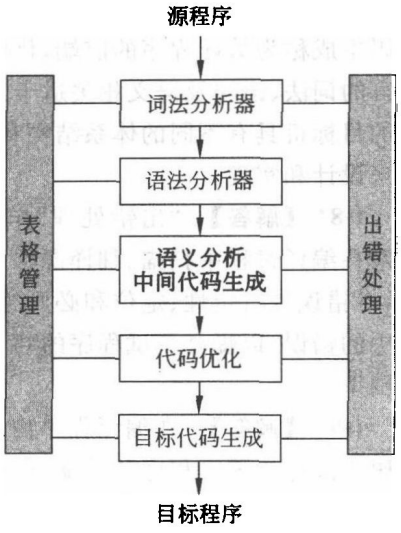


图 1-4 编译程序的总体结构

1-4 【解答】 图 1-3 填写完整后如图 1-4 所示。

1-5 【解答】 编译程序的实现途径即实现方式一般可以用高、中级程序设计语言编程实现，可以通过移植的方式实现，可以通过自编译的方式，还可以通过部分自动生成的方式实现。

1-6* 【解答】 用 T 型图描述的将 A 机器上已经存在的 L 语言的编译程序移植到 B 机器上的过程如下：

① 设已有 C_A^{LA}，用其已经实现的语言 L 为 B 机器的源语言 L 编写一个交叉编译器；即创建 C_L^{LB}，如图 1-5 所示。

② 生成机器 B 的目标代码，即创建 C_A^{LB}，如图 1-6 所示。

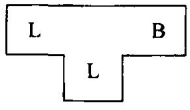


图 1-5 编译器 C_L^{LB} 的 T 型图

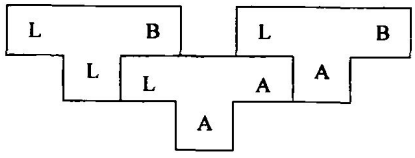


图 1-6 编译器 C_A^{LB} 的 T 型图

③ C_A^{LB} 通过 C_L^{LB} 来运行得到编译器 C_A^{LB}，编译器 C_A^{LB} 的 T 型图可以用如图 1-7 所示的叠放在一起的 T 型图表示。

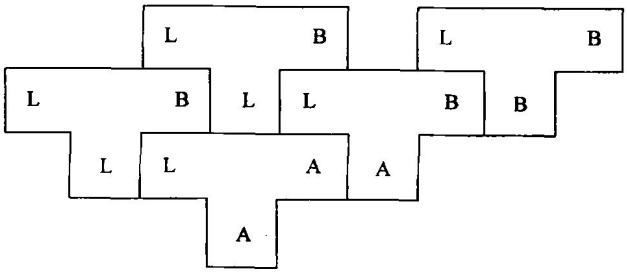


图 1-7 编译器 C_A^{LB} 叠放的 T 型图

1-7*【解答】 在编译程序构成的经典划分中，词法分析、语法分析及语义分析中间代码生成称为编译程序的前端，代码优化及代码生成称为后端。涉及前端的功能仅与源语言的词法、语法及语义相关适于自动生成。对后端实现的代码优化和代码生成，鉴于不同的目标机具有不同的体系结构和指令系统，代码优化和代码生成需要基于特定的目标机来设计和实现。

1-8*【解答】 “出错处理”作为编译程序的一个不可或缺的公共程序，其主要作用是对在编译过程中扫描、翻译源程序时根据文法规则和语义规则诊断源程序中存在的错误，对错误进行定性、定位和必要的容错处理。这样可以协助程序员及时准确地发现源程序中的错误，以提高调试程序的效率，方便用户修改程序，并能把错误限制在尽可能小的范围里。

1-9*【解答】 在编译程序构成的经典划分中，词法分析、语法分析及语义分析中间代码生成称为编译程序的前端，代码优化及代码生成称为后端。涉及前端的功能仅与源语言的词法、语法及语义相关适于自动生成。对后端实现的代码优化和代码生成，鉴于不同的目标机具有不同的体系结构和指令系统，代码优化和代码生成需要基于特定的目标机来设计和实现。

1-10*【解答】 “出错处理”作为编译程序的一个不可或缺的公共程序，其主要作用是对在编译过程中扫描、翻译源程序时根据文法规则和语义规则诊断源程序中存在的错误，对错误定性、定位和必要的容错处理。这样可以协助程序员及时准确地发现源程序中的错误，以提高调试程序的效率，方便用户修改程序，并能把错误限制在尽可能小的范围里。

第 2 章　形式语言与自动机理论基础

2.1　学习要点指导

2.1.1　文法和语言的形式定义

语言的基本成分与运算

1）字符和字符串

（1）字符和字符串的概念

字符：字母表中的一个字母。

字符串：Σ 是字母表，Σ 上的字符串的集合 Σ^* 可递归定义如下：

① ε（ε 是由 Σ 中的 0 个字符组成的符号，称为空串）$\in\Sigma^*$。

② 如果 $\omega\in\Sigma^*$ 且 $a\in\Sigma$，那么 $\omega a\in\Sigma^*$。

③ $\omega\in\Sigma^*$ 当且仅当 ω 由有限步①和②产生。

Σ^* 中的元素称为字符串，也可叫做符号串、串或句子。

（2）运算

① 字符串的基本运算：长度、子串、前/后缀、真前/后缀、连接、幂。

长度：符号串 ω 中含有字母表中符号的个数，表示为 $|\omega|$。

前缀：从符号串的尾部删去 0 个或若干个符号之后剩余的部分。

后缀：从符号串的首部删去 0 个或若干个符号之后剩余的部分。

子串：从一个符号串中删去它的一个前缀和一个后缀之后剩余的部分。

真前缀（真后缀）：不是自身的前缀（后缀）。

连接：将符号串 v 直接拼接在符号串 ω 之后，记作 ωv。

幂：把符号串 ω 自身连接 n 次得到符号串，记作 ω^n。

② 字符串集合的运算：乘积、幂、正闭包/自反闭包。

乘积：$AB=\{\omega v|(\omega\in A)\wedge(v\in B)\}$

幂：A 与自身的乘积。$A^n=A^{n-1}A=AA\cdots A$（n 个 A）

正闭包：$A^+=A^1\cup A^2\cup\cdots\cup A^n\cup\cdots$

自反闭包：$A^*=\{\varepsilon\}\cup A^1\cup A^2\cup\cdots\cup A^n\cup\cdots=\varepsilon\cup A^+$

2）文法和语言的形式定义

（1）文法的形式定义

一部文法 G 是一个四元组 $G=(V_N, V_T, S, P)$

V_N：非空有限的非终结符号集，其中的元素称为非终结符，或称为语法变量，代表了一个语法范畴，表示一类具有某种性质的符号。

V_T：非空有限的终结符号集，其中的元素称为终结符，其代表了组成语言的不可再分的基本符号集。V_T 即字母表 Σ。

设 V 是文法 G 的符号集，则有

$$V = V_T \cup V_N, \quad \text{并且} \quad V_T \cap V_N = \Phi$$

S：文法的开始符号或识别符号，亦称公理，$S \in V_N$。S 代表语言最终要得到的语法范畴。

P：产生式集。所谓产生式就是按一定格式书写的定义语法范畴的文法规则，它是一部文法的实体。产生式的形式为：

$\alpha \rightarrow \beta$ 或 $\alpha ::= \beta$（$\alpha \in V^+$，且 α 中至少包含 V_N 中的一个元素，$\beta \in V^*$）

其中 α 称为产生式的左部，β 称为产生式的右部或称为 α 的候选式。

(2) 语言的形式定义

$$L(G) = \{\alpha \mid \alpha \in V_T^* \wedge S => \alpha, S \text{ 是文法 } G \text{ 的开始符号}\}$$

(3) 基本概念及术语

直接推导、直接归约、最左(右)推导、句型、句子、规范推导(句型、归约)、递归、左递归、右递归、直接递归、文法的等价。

直接推导：若 $\alpha \rightarrow \beta \in P$，则称 $\delta\alpha\gamma$ 直接推导出 $\delta\beta\gamma$，记作 $\delta\alpha\gamma \Rightarrow \delta\beta\gamma$。

直接归约：若 $\alpha \rightarrow \beta \in P$，则称 $\delta\beta\gamma$ 直接归约到 $\delta\alpha\gamma$。

最左(右)推导：推导过程中，总是对字符串中最左(右)边的非终结符进行替换。

句型：若 $S \Rightarrow \alpha$ ($\alpha \in V^*$)，则称 α 为 $G(S)$的句型。

句子：若 $S \Rightarrow \alpha$ ($\alpha \in V_T^*$)，则称 α 为 $G(S)$的句子。

规范推导：最右推导也称为规范推导。

规范句型：仅用规范推导得到的句型。

规范归约：规范推导的逆序。

递归：存在形如 $A \Rightarrow \alpha A \beta$ 的递归推导。

左递归：递归推导 $A \Rightarrow \alpha A \beta$ 中，$\alpha = \varepsilon$。

右递归：递归推导 $A \Rightarrow \alpha A \beta$ 中，$\beta = \varepsilon$。

直接递归：存在形如 $A \rightarrow \alpha A \beta$ 的递归产生式。

文法的等价：若 $L(G_1) = L(G_2)$，则称文法 G_1 和 G_2 是等价的。

3) 文法的表示方法

文法的表示方法有三种：BNF、EBNF 和语法图。

4) 分析树与二义性

分析树：分析树的根结点是文法的开始符号，结点间的父子关系为产生式规则，即若父结点标识为 A，子结点从左到右依次标识为 $a_1, a_2, \cdots, a_n$，则在文法中存在一产生式 $A \rightarrow a_1 a_2 \cdots a_n$。

文法的二义性：至少存在一个句子，对应两棵(或两棵以上)不同的分析树。

5) 文法和语言的类型

(1) 0 型文法，也称短语结构文法。

0 型文法描述的语言为 0 型语言，可由图灵机来识别。

(2) 1 型文法,也称上下文有关文法。

1 型文法描述的语言为 1 型语言,可由线性有界自动机来识别。

(3) 2 型文法,也称上下文无关文法。

2 型文法描述的语言为 2 型语言,可由非确定的下推自动机来识别。

(4) 3 型文法,也称正则文法或线性文法。

3 型文法描述的语言为 3 型语言或正则语言,可由确定的有限自动机来识别。

2.1.2 语言的识别——有限自动机 FA

1. 确定的有限自动机 DFA

1) 定义及三种表示方法

一个确定的有限自动机 M(DFA M) 是一个五元组

$$M=(Q,\Sigma,f,q_0,Z)$$

其中:Q:状态的有限集合,每个元素称为一个状态。

Σ:输入字符的有限集合(或有穷字母表)。每个元素是一个输入字符。

f:状态转换函数,是一个从 $Q\times\Sigma$ 到 Q 的映射。

q_0:M 的唯一初态(也称开始状态),$q_0\in Q$。

Z:M 的终态集(或接受状态集),$Z\subseteq Q$。

有三种表示方法:形式描述、状态图和状态表。

2) 识别机制:语言与 DFA

设 $M=(Q,\Sigma,f,q_0,Z)$是一确定的有限自动机,$\omega=w_1w_2\cdots w_n$ 是字母表 Σ 上的一个字符串,如果存在 Q 中的状态序列 $p_0,p_1,\cdots,p_n$,满足下列条件:

① $p_0=q_0$;

② $f(p_i,w_{i+1})=p_{i+1},i=0,1,\cdots,n-1$;

③ $p_n\in Z$。

即有 $f(q_0,\omega)\in Z$,则称 M 接受(识别)ω。否则称 M 拒绝(不识别)ω。

2. 非确定的有限自动机 NFA

1) 定义及三种表示方法

一个非确定的有限自动机 M(NFA M) 是一个五元组

$$M=(Q,\Sigma,f,q_0,Z)$$

其中:Q:状态的有限集合。

Σ:输入字符的有限集合(有穷字母表)。

f:状态转换函数,从 $Q\times(\Sigma\cup\{\varepsilon\})\to 2^Q$ 的映射。

q_0:开始状态,$q_0\in Q$。

Z:终态集(接受状态集),$Z\subseteq Q$。

表示方法:形式描述、状态图和状态表。

试读结束:需要全本请在线购买: www.ertongbook.com