



面向21世纪计算机专业本科系列教材

编译原理

Compiler Principles

(第二版)

何炎祥 主编



华中科技大学出版社
<http://press.hust.edu.cn>

面向 21 世纪计算机专业本科系列教材

编译原理

(第二版)

主编 何炎祥

编者 何炎祥 李晓燕 王汉飞

华中科技大学出版社

图书在版编目(CIP)数据

编译原理(第二版)/何炎祥 主编

武汉:华中科技大学出版社,2005年8月

ISBN 7-5609-3441-2

I. 编…

II. ①何… ②李… ③王…

III. 编译原理-高等学校-教材

IV. TP314

编译原理(第二版)

何炎祥 主编

策划编辑:沈旭日

责任编辑:徐正达

责任校对:刘 飞

封面设计:潘 群

责任监印:熊庆玉

出版发行:华中科技大学出版社

武昌喻家山 邮编:430074 电话:(027)87557437

录 排:华中科技大学惠友文印中心

印 刷:湖北新华印务有限公司

开本:787×960 1/16

印张:21.25

字数:380 000

版次:2005年8月第2版

印次:2005年8月第6次印刷

定价:27.00元

ISBN 7-5609-3441-2/TP·575

(本书若有印装质量问题,请向出版社发行部调换)

内 容 简 介

本书主要介绍设计和构造编译程序的基本原理和方法,内容包括适应高级程序设计语言翻译的形式语言理论和自动机理论、常用的词法分析方法、各种经典的语法分析技术、语法制导翻译方法、存储组织与管理方法、造查表方法、代码优化和代码生成方法、编译自动化和并行编译程序,以及词法分析器生成工具 LEX 和语法分析器生成工具 YACC 等。本书特别注重理论与实践的沟通,基本概念清晰,循序渐进,深入浅出。各章附有难度不一的习题。

本书可作为高等院校计算机专业的教材,也可供有关教师、研究生和科技工作者学习和参考。

面向 21 世纪计算机教材出版指导委员会

主任 陈火旺 沈绪榜 邹寿彬
委员 (以姓氏笔画为序)
 王长胤 韦 敏 卢开澄 卢正鼎
 齐 欢 刘太林 何炎祥 张 峰
 杨传谱 苏锦祥
秘书 沈旭日

面向 21 世纪计算机专业本科系列教材编委会

主任 何炎祥 卢正鼎
委员 (以姓氏笔画为序)
 卢炎生 肖德宝 陈 珉 张 峰
 贺贵明 姜新祺 熊前兴
秘书 沈旭日



总 序

自 1946 年世界上第一台电子数字计算机 ENIAC 诞生以来,计算机硬件系统经过了电子管、晶体管、小规模集成电路和大规模集成电路等几个阶段,正遵循着摩尔定律高速地发展:1998 年,速度最快的个人 PC 微处理器是 Intel 450MHz 的 Xeon, 1999 年速度最快的已达 800MHz;1997 年 2.1GB 的磁盘容量已经很不错了,1999 年则已突破 10GB……软件方面,无论是操作系统、数据库系统,还是编程语言、应用软件,更是频繁地更新换代,令人眼花缭乱。

与此同时,作为计算机与通信技术结合的产物——计算机网络得到了迅速发展,特别是 Internet 技术的广泛应用,使得计算机网络的规模越来越大,网上主机数目一直保持每 3 年增长 10 倍的速率,Internet 上的数据流量则保持着平均每半年就翻一番的增长速率,信息网络已交叉纵横整个世界,将偌大的世界连成了一个“地球村”。

计算机技术日新月异的进步,对现有的计算机专业的教学模式提出了挑战,同时也带来了前所未有的机遇。深化面向 21 世纪的教学改革,寻求一条行之有效的途径,培养跨世纪的高素质的科技人才,已是当务之急。如果说教学内容、课程体系的改革是教学改革的重点和难点,那么,教材建设则是其不可或缺的重要组成部分。华中理工大学出版社敏锐地抓住了这一点,在其倡议和组织下,我们经过研究、讨论和对教学经验进行总结,规划了这套“面向 21 世纪计算机教材”。为了满足各级各类学校人才培养的需要,这套教材计划包括计算机专业类教材和非计算

机专业类教材,从层次上则可划分为研究生层次、本科生层次、高职高专层次、中职中专层次、中小学层次等若干个子系列,将陆续分批出版。

当今世界,信息革命方兴未艾,知识经济已见端倪,教育观念正面临从注重以知识为主体向以能力为主体的转变。我们在对教材进行规划和评审时,尤其注重把提高学生素质、培养学生的应用能力和创新能力作为首要的评价标准,同时注意教材的特色和教学的实用性,反映最新的教学和科研成果,体现时代特征。

限于水平和经验,这批教材的编写、出版还存在不足,希望使用教材的学校、教师和学生以及其他读者积极提出批评意见,以便我们及时更新、修订,以满足读者要求。

面向 21 世纪计算机教材出版指导委员会主任

陈火旺 (中国科学院院士,国防科技大学教授)

沈绪榜 (中国科学院院士,华中理工大学教授)

邹寿彬 (华中理工大学教授,副校长)

2000 年 2 月 10 日

人们已普遍认识到:21 世纪是信息时代,以计算机为核心的信息技术是 21 世纪科技发展的大趋势。那么,作为计算机专业人才培养基地的大学计算机专业,如何适应这种发展,培养出符合时代要求、具有创新能力的人才呢?这是近年来计算机教育界讨论的热门话题,也是我们长期思考并努力探索的课题。

教材是人才培养的基础。在华中理工大学出版社的倡议和委托下,我们自 1998 年下半年起就开始讨论、筹划编写一套适应 21 世纪人才培养需要的计算机本科专业系列教材。在此基础上,我们组织了武汉大学、华中理工大学、华中师范大学、武汉测绘科技大学、武汉水利电力大学、武汉交通科技大学等院校的部分教师共同编写了这套“面向 21 世纪计算机系列教材”,以期总结我们在教学内容和课程体系改革方面的体会和做法,在适应 21 世纪的教材建设方面做出自己的努力。

值得欣慰的是,在教材的编写过程中,全国计算机专业教学指导委员会、中国计算机学会教育委员会联合推出了“计算机学科教学计划 2000”(简称“2000 教程”),这就更增强了我们编好这套教材的信心。在编写过程中,我们吸收了其中与我们内容相异的新内容。因此,也完全可以说,这套教材是与“2000 教程”完全配套的教材。

我们这套系列教材的编写计划分为两个阶段:第一阶段,在 2000 年内出版“2000 教程”中所涉及的所有专业课和部分专业基础课教材;第二阶段,在 2000 年以后出版与这套教材相配套的实践课和实验课教材,以及教学辅导书。

我们希冀这套教材具有以下特点:

1. 基础性和先进性相结合。与其他学科相比,计算机学科的一个显著特点就是知识内容更新更快,这对教学内容的选取、课程知识结构的构建提出了挑战。基于大学教育应努力实现知识、能力、素质三者辩证统一的目标,我们把编写的重点放在基础知识、基本技能和基本方法上,希望提高学生的理论素养和分析问题、解决问题的能力;与此同时,注重介绍最新的技术和方法,以拓展学生的知识面,激发他们学习的积极性和创新意识。

2. 理论性与应用性相结合。理论是规律的表现形式,良好的理论素养是应用

的前提,而掌握理论的目的就是应用。在教材的编写过程中,我们注意了理论的系统性,在讲深讲透主要知识的基础上,各门课程知识点的选取做到尽量广一些;融理论性和应用性于一体,在阐述理论的同时,尤其注意理论方法的讲授,以培养学生应用理论和技术的能力。此外,精心设计了比较多的习题,以加强应用能力和创造能力的培养。

3. 时代性和实用性相结合。力求精简旧的知识点,增加新的知识点,使整个知识建立在“高”、“新”平台上,体现教材的时代特征。但是,并不片面追求“高”、“新”,而是实事求是地充分考虑一般高校目前所拥有的教学设备、师资条件,注重教材的实用性。我们以为,教材建设不可能毕其功于一役,而必须根据学科的发展和客观环境以及条件的变化不断努力和改进。需要说明的是,与“2000 教程”相比,我们根据人的认识规律和教学安排的需要,将有些课程进行了划分或合并,以便于教师根据需要灵活安排。

4. 科学性与通俗性相结合。概念原理、新技术的阐述力求准确、精练;写作风格上尽量通俗易懂、深入浅出、图文并茂,增加可读性,便于学生自学。

如果说科学技术快速发展是 21 世纪的一个重要特征的话,那么,教学改革将是 21 世纪教育工作永恒的主题,是需要不断探索的课题。我们要达到以上目标,还需要不断地努力实践和完善。欢迎使用这套教材的教师、学生和其他读者提出宝贵意见。

最后,衷心感谢参加这套教材编写的所有作者所贡献的成果和辛勤的汗水,对为这套教材的编写提供支持的有关学校、院系的领导和老师表示诚挚的谢忱!感谢华中理工大学出版社为本系列教材的出版所付出的艰辛和努力!

面向 21 世纪计算机专业本科系列教材编委会主任

何炎祥 (武汉大学教授)

卢正鼎 (华中理工大学教授)

1999 年 11 月 20 日

前 言

编译程序(Compiler)是计算机的重要系统软件,是高级程序设计语言的支撑基础。本书主要介绍设计和构造编译程序的基本原理和方法。

本书共分 14 章。

第 1 章讲述编译程序的功能、结构、工作过程、组织方式、编译程序与高级语言的关系以及编译自动化方面的基本知识。第 2 章介绍形式语言理论,我们仅仅给出了便于理解、有助于研究各种分析方法和设计构造编译程序的形式语言理论,并着重介绍了上下文无关文法。

有穷自动机是描述词法的有效工具,也是进行词法分析的主要理论基础。因此,第 3 章专门讨论有穷自动机,它与正规文法、正规表达式之间的对应关系以及它的确定化和最小化方面的知识,略去了像 Turing 机及可计算性理论方面的内容。第 4 章讨论词法分析的功能和词法分析程序的设计方法。

上下文无关文法可用于描述现今大多数高级程序设计语言的语法,也是语法分析的主要理论支柱。为此,在接下来的几章里,主要讨论了与上下文无关文法相关的各类语法分析方法。

第 5 章介绍自上而下分析方法,包括 LL(k)文法、LL(1)分析方法和应用十分广泛的递归下降分析方法。第 6 章讨论自下而上分析方法的一般原理和优先分析方法,包括简单优先分析技术和算符优先分析方法。第 7 章专门讨论自下而上的 LR(k)分析方法,包括 LR(0)、SLR(1)、规范 LR(1)以及 LALR 分析表的构造算法。

第 8 章介绍语法制导翻译方法,主要讨论了 SDTS 的基本原理、属性翻译文法以及它们在中间代码生成中的应用。

第 9 章讨论运行时的存储组织与管理,其中考虑了一些重要的语言特征,如过程调用、参数传递、数组和记录的存取方式以及多种存储分配技术。

第 10 章讨论符号表的组织和存取符号表的各种方法。第 11 章介绍常用的优化方法。第 12 章简单讨论代码生成的原理。

第 13 章、第 14 章中花了较大篇幅分别介绍了词法分析器生成工具 LEX 和语法分析器生成工具 YACC,以便于本课程的教学实习和课程设计。

我们认为某些形式语言理论和自动机理论对设计构造编译程序是极其有用

的,但现有的不少形式语言理论及自动机理论与设计和构造编译程序的关系不大。本书试图在沟通设计和构造编译程序的理论与实践、原理与方法等方面作一点尝试。

编译原理这门课程是计算机专业的主干课和必修课,也是计算机专业高年级课程中较难学习的一门课程,其先导课程是汇编语言程序设计、计算机组成原理、数据结构、高级语言程序设计和离散数学等。本课程的参考学时数为 72,使用者可根据具体情况对教材内容进行取舍,例如,工科院校的学生可略过第 7 章、第 8 章并可精简第 2 章的内容,从而使授课学时数减至 54。

为了满足教学和自学的需要,我们还编写了配套教材《编译原理学习与解题指南》,供读者选用。

本书的第 1~6 章由李晓燕编写,第 7~12 章及 1.6~1.7 节由何炎祥编写,第 13 章、第 14 章由王汉飞编写,全书由何炎祥统稿。

本书可作为高等院校计算机专业的教材,也可供教师、研究生及有关科技工作者学习和参考。

本书成书过程中,得到了华中科技大学出版社的鼎力协助,此外,书中还引用了一些专家学者的研究成果,在此一并表示感谢。

本次再版,对全书内容进行了细致的审定和修改,以使本书趋于完善。

作 者

2005 年 5 月于武昌珞珈山

目 录

第 1 章 引论	(1)
1.1 翻译程序	(1)
1.1.1 程序设计语言	(1)
1.1.2 翻译程序	(2)
1.2 编译程序的工作过程	(3)
1.3 编译程序的结构	(5)
1.4 编译程序的组织方式	(6)
1.5 编译程序的自展、移植与自动化	(7)
1.5.1 高级语言的自编译性	(7)
1.5.2 编译程序的自展技术	(8)
1.5.3 编译程序的移植	(9)
1.5.4 编译程序的自动化	(9)
1.6 翻译程序编写系统	(10)
1.7 并行编译程序	(12)
1.8 小结	(13)
习题一	(14)
第 2 章 形式语言概论	(15)
2.1 字母表和符号串	(15)
2.2 文法及其分类	(17)
2.2.1 文法	(17)
2.2.2 文法分类	(17)
2.2.3 文法举例	(19)
2.3 语言和语法树	(20)
2.3.1 推导和规范推导	(20)
2.3.2 句型、句子和语言	(22)
2.3.3 语法树	(22)

2.3.4 产生式树	(23)
2.4 关于文法和语言的几点说明	(25)
2.5 分析方法简介	(27)
2.5.1 自上而下分析方法	(27)
2.5.2 确定的自上而下分析方法	(29)
2.5.3 自下而上分析方法	(30)
2.6 小结	(32)
习题二	(32)
第3章 有穷自动机	(35)
3.1 有穷自动机的形式定义	(35)
3.1.1 状态转换表	(35)
3.1.2 状态转换图	(36)
3.1.3 自动机的等价性	(37)
3.1.4 非确定有穷自动机	(38)
3.2 NDFA 到 DFA 的转换	(39)
3.2.1 空移环路的寻找和消除	(39)
3.2.2 消除空移	(40)
3.2.3 确定化——子集法	(41)
3.2.4 确定化——造表法	(43)
3.2.5 ϵ NDFA 的确定化	(44)
3.2.6 消除不可达状态	(45)
3.2.7 DFA 的化简	(46)
3.2.8 从化简后的 DFA 到程序表示	(47)
3.3 正规文法与有穷自动机	(48)
3.3.1 从正规文法到 FA	(48)
3.3.2 从 FA 到正规文法	(49)
3.4 正规表达式与 FA	(50)
3.4.1 正规表达式的定义	(50)
3.4.2 正规表达式与 FA 的对应性	(52)
3.4.3 正规表达式到 NDFA 的转换	(52)
3.4.4 NDFA 到正规表达式的转换	(53)
3.4.5 从正规文法到正规表达式	(55)

3.5 DFA 在计算机中的表示	(56)
3.5.1 矩阵表示法	(56)
3.5.2 表结构	(57)
3.6 小结	(57)
习题三	(58)
第4章 词法分析	(60)
4.1 词法分析器与单词符号	(60)
4.1.1 词法分析器	(60)
4.1.2 单词符号	(60)
4.2 扫描程序的设计	(62)
4.2.1 预处理	(62)
4.2.2 状态转换图	(62)
4.2.3 根据状态图设计词法分析程序	(64)
4.3 标识符的处理	(66)
4.3.1 类型的机内表示	(66)
4.3.2 标识符的语义表示	(66)
4.3.3 符号表(标识符表)	(67)
4.3.4 标识符处理的基本思想	(67)
4.4 设计词法分析程序的直接方法	(68)
4.4.1 由正规文法设计词法分析程序	(68)
4.4.2 由正规表达式设计词法分析程序	(69)
4.5 小结	(70)
习题四	(70)
第5章 自上而下语法分析	(72)
5.1 消除左递归方法	(72)
5.1.1 文法的左递归性	(72)
5.1.2 用扩展的 BNF 表示法消除左递归	(72)
5.1.3 直接改写法	(73)
5.1.4 消除左递归算法	(74)
5.2 LL(k)文法	(75)
5.2.1 LL(1)文法的判断条件	(75)

5.2.2 集合 FIRST、FOLLOW 与 SELECT 的构造	(76)
5.3 确定的 LL(1)分析器的构造	(78)
5.3.1 构造分析表 M 的算法	(78)
5.3.2 LL(1)分析器的总控算法	(80)
5.4 递归下降分析程序及其设计	(81)
5.4.1 框图设计	(82)
5.4.2 程序设计	(84)
5.5 带回溯的自上而下分析法	(85)
5.5.1 文法在内存中的存放形式	(85)
5.5.2 其它信息的存放	(86)
5.5.3 带回溯的自上而下分析算法	(86)
5.6 小结	(90)
习题五	(91)

第 6 章 自下而上分析和优先分析方法

6.1 自下而上分析	(93)
6.2 短语和句柄	(94)
6.3 移进-归约方法	(96)
6.4 有关文法的一些关系	(97)
6.4.1 关系	(97)
6.4.2 布尔矩阵和关系	(98)
6.4.3 Warshall 算法	(99)
6.4.4 关系 FIRST 与 LAST	(101)
6.5 简单优先分析方法	(103)
6.5.1 优先关系	(103)
6.5.2 简单优先关系的形式化构造方法	(104)
6.5.3 简单优先文法及其分析算法	(109)
6.5.4 简单优先分析方法的局限性	(111)
6.6 算符优先分析方法	(112)
6.6.1 算符优先文法	(112)
6.6.2 OPG 优先关系的构造	(113)
6.6.3 素短语及句型的分析	(115)
6.6.4 算符优先分析算法	(116)

6.7 优先函数及其构造	(118)
6.7.1 优先函数	(118)
6.7.2 Bell 方法	(119)
6.7.3 Floyd 方法	(120)
6.7.4 两种方法的比较	(122)
6.7.5 运用优先函数进行分析	(122)
6.8 两种优先分析方法的比较	(123)
6.9 小结	(124)
习题六	(124)
第 7 章 自下而上的 LR(k) 分析方法	(127)
7.1 LR(k) 文法和 LR(k) 分析器	(128)
7.2 LR(0) 分析表的构造	(131)
7.2.1 规范句型的活前缀	(132)
7.2.2 LR(0) 项目	(132)
7.2.3 文法 G 的拓广文法	(133)
7.2.4 CLOSURE(I) 函数	(133)
7.2.5 goto(I, X) 函数	(134)
7.2.6 LR(0) 项目集规范族	(135)
7.2.7 有效项目	(137)
7.2.8 举例	(138)
7.2.9 LR(0) 文法	(141)
7.2.10 构造 LR(0) 分析表的算法	(141)
7.2.11 构造 LR(0) 分析表的步骤小结	(142)
7.3 SLR 分析表的构造	(143)
7.4 规范 LR(1) 分析表的构造	(147)
7.5 LALR 分析表的构造	(153)
7.6 无二义性规则的使用	(157)
7.7 小结	(159)
7.7.1 LR 分析程序	(159)
7.7.2 LR 分析表的自动构造	(159)
7.7.3 文法间的关系	(160)
7.7.4 LR 文法举例	(161)

7.7.5 有关 LR 文法的几个结论	(164)
习题七	(164)
第 8 章 语法制导翻译法	(166)
8.1 一般原理和树变换	(166)
8.1.1 一般原理	(166)
8.1.2 树变换	(169)
8.2 简单 SDTS 和自上而下翻译器	(171)
8.3 简单后缀 SDTS 和自下而上翻译器	(174)
8.3.1 后缀翻译	(175)
8.3.2 IF-THEN-ELSE 控制语句	(175)
8.3.3 函数调用	(177)
8.4 抽象语法树的构造	(177)
8.4.1 自下而上构造 AST	(180)
8.4.2 AST 的拓广	(180)
8.5 属性文法	(181)
8.5.1 L 属性文法	(182)
8.5.2 S 属性文法	(182)
8.6 中间代码形式	(183)
8.6.1 逆波兰表示法	(183)
8.6.2 逆波兰表示法的推广	(184)
8.6.3 四元式	(185)
8.6.4 三元式	(187)
8.7 属性翻译文法的应用	(188)
8.7.1 综合属性与自下而上定值	(188)
8.7.2 继承属性和自上而下定值	(189)
8.7.3 布尔表达式到四元式的翻译	(190)
8.7.4 条件语句的翻译	(191)
8.7.5 迭代语句的翻译	(193)
8.8 小结	(195)
习题八	(196)