



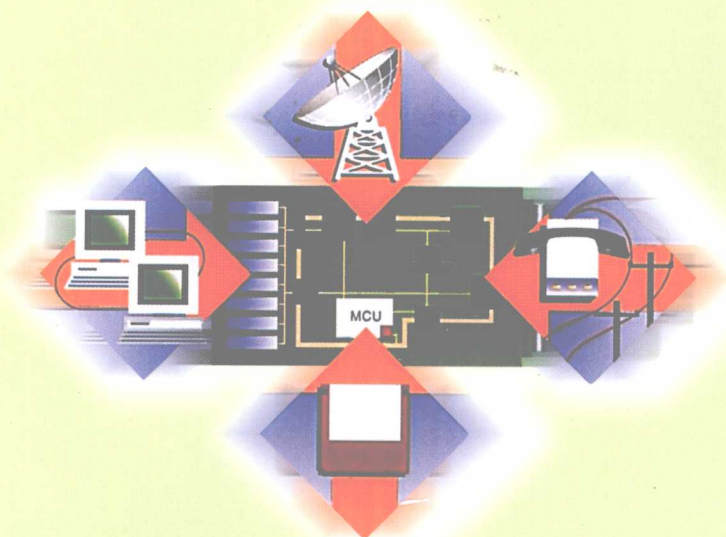
普通高等教育“十一五”国家级规划教材

计算机编译原理

——编译程序构造实践

第二版

张幸儿 编著



 科学出版社
www.sciencep.com

普通高等教育“十一五”国家级规划教材

计算机编译原理

——编译程序构造实践

(第二版)

张幸儿 编著

科学出版社

北京

内 容 简 介

本书共分两篇,第一篇概论包括编译程序概述与程序(软件)研制的实践指南,第二篇实践篇包括文法及相关概念、词法分析、语法分析、语义分析和目标代码生成、目标代码优化。第一篇系统而概要地介绍了以C型语言编译系统作为开发平台进行界面设计、研制应用程序的操作步骤;第二篇中的实习题涵盖了编译程序构造的各个主要环节,以C语言为规范,给出了相关的数据结构、众多的C型伪代码程序,供读者参考。书中所配光盘内容是本书的附录,其中附录一给出了语义分析中语义子程序的设计;附录二给出了覆盖编译全过程的可执行样本程序,让读者深切感受从源程序字符序列到等价的目标代码的翻译全过程的感性体验;附录三与附录四给出了以VC++为开发平台开发的应用系统实例;附录五是以赋初值方式给出的各章实习题调试实例。

期望读者通过阅读本书,能了解C型语言研制应用程序的要点,汲取丰富的编程方法和技巧,对照操作,便可以轻松开发具有良好界面的应用程序(包括编译程序)。

本书可供计算机及相关专业作为编译原理课程的富有启发性的配套实践教材,同时也可作为计算机软件工作者、研究生以及广大计算机爱好者的参考用书。

图书在版编目(CIP)数据

计算机编译原理:编译程序构造实践/张幸儿编著. —2版. —北京:科学出版社,2009

(普通高等教育“十一五”国家级规划教材)

ISBN 978-7-03-022203-9

I. 计… II. 张… III. 编译程序-程序设计-高等学校-教材 IV. TP314

中国版本图书馆CIP数据核字(2008)第079031号

责任编辑:巴建芬 潘继敏/责任校对:朱光光

责任印制:张克忠/封面设计:陈 敬

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

深 海 印 刷 有 限 责 任 公 司 印 刷

科学出版社发行 各地新华书店经销

*

2005年7月第 一 版 开本:B5(720×1000)

2009年1月第 二 版 印张:21 3/4

2009年1月第三次印刷 字数:414 000

印数:5 001—8 500

定价:36.00元(含光盘)

(如有印装质量问题,我社负责调换〈路通〉)

第二版前言

本书作为编译程序构造原理和实践相结合的范例呈现于广大读者面前,让读者理解从高级程序设计语言源程序到等价的低级语言目标程序的翻译全过程,认识编译原理课程中讨论的实现技术与实际实现之间的差异,尤其是语义分析时翻译方案(语法制导定义)中语义动作(语义规则)的书写形式与实际实现之间的真实联系。

为了有助于读者进一步掌握编译程序的构造技术和方法,轻松实现编译程序的构造,第二版在下列几方面开展了工作:

(1) 以 C 语言为规范,写出伪代码程序,包括数据结构与控制部分。这些 C 型伪代码程序与 C 语言程序的区别仅在于标识符名字是汉字,且包含一些用文字叙述的功能,这或相当于函数调用,或相当于宏功能。附录一(见光盘)给出了语义分析中语义子程序的设计(C 型程序)。

(2) 对伪代码程序进一步细化,不仅数据结构以 C 语言形式给出,而且对一些功能的实现给出了具体的 C 型语言语句,仅标识符名是汉字而已。对少数不确切的表示法进行了改进。例如,建立某种结点的指针: $P = \text{pointer}(\{N, Q\})$,确切地指明是哪一类结点的指针,分为 Apointer(属性信息结点)、Dpointer(依赖属性结点)和 Ipointer(目标代码结点)。关键部分的伪代码程序进行了上机验证。

(3) 对实现有一定难度的功能,增加了必要的伪代码程序,使实现进一步具体化,特别是在附录二(见光盘)中提供了可执行的 C 语言程序,覆盖文法的输入输出、词法分析、语法分析与目标代码的生成,使读者对翻译全过程有一个深切的感性体验。为方便读者上机实践,附录五(见光盘)中包含了以赋初值方式给出的各章习题调试实例。

(4) 附录四(见光盘)给出了在 Visual C++ 平台上开发应用系统的样本程序。

一个应用程序的开发,一般涉及界面设计、程序设计以及数据库三个方面。作为基础,要先解决界面设计和程序设计,为此需解决界面、数据与代码三者之间的联系,附录三(见光盘)展示了何处定义待处理的数据,如何把从界面输入的数据保存到计算机内,如何由代码处理,又如何通过界面显示输出处理结果。读者把这些与第一篇第 2 章关于界面的实现讨论中所列操作步骤结合起来,进行演练,将取得好的效果。

本书期望有助于:

- (1) 选修编译原理课程的学生,作为学习的辅助教材,加深对编译原理课程概念的理解;
- (2) 自学编译原理课程的读者,加深对编译原理课程的理解和要点的领会;
- (3) 打算使用 C/C++ 语言开发平台研制应用程序的读者,作为这些平台的应用入门;
- (4) 编译程序构造实践者,作为实现的指南。

注:书中标星号“*”的章节是选读内容。

期望读者阅读本书第二版之后,掌握程序设计中的的一些常用基本技术和方法,提高 C 语言程序设计的能力,特别是能应用 Visual C++ 平台开发应用程序。

最后,感谢张福炎教授对本书的大力支持。感谢李意勉女士,在她的支持和大力帮助下,本书才能顺利完成。最后,要感谢巴建芬编辑,是她的关心和大量的辛勤工作,才使第二版得以顺利面世。

本书是编译原理课程领域的一个初步尝试,尽管对第一版进行了补充和完善,但由于笔者水平限制,错误与不妥之处在所难免,也不乏改进与完善的余地。欢迎读者批评指正。

联系地址:南京大学计算机科学与技术系 210093

E-mail: zhangxr0@sina.com 或 zhang_xinger@hotmail.com

作者

南京大学计算机科学与技术系

2008年5月

第一版前言

编译原理讨论编译程序构造的基本原理和技术,它以形式语言理论的基本概念为基础,讨论编译全过程,剖析编译程序的构造,既有深沉的理论性,又有浓厚的实践性。编译原理课程与计算机科学中的众多学科相关联,其基本原理与技术对计算机应用领域的许多方面有着相当大的影响,对程序研制也有一定的指导意义。

为什么编译原理课程是理论性很强、实践性也很强的课程?这是因为应把编译原理的基本概念、技术与方法应用于编译程序构造的实践。然而目前不足的是,还没有一本编译原理教材明确地阐明如何把编译程序构造的各阶段有机地联系起来,实现从输入源程序到输出等价的目标代码的全过程。现有的编译原理教材基本上可以说还是仅仅局限于原理性、概念性的理论介绍的。

事实上,编译原理课程中讨论的实现技术与实际实现有着一定的距离,仅考察下列几点便已清楚:

- 文法的书写形式与实际的机内存放形式;
- 消去左递归文法等价变换的算法与实际实现;
- 非确定有穷状态自动机的确定化与实际实现;
- 预测识别算法与语法分析树的生成;
- LR(1)识别算法与实际实现。

尤其是语义分析时的翻译方案(语法制导定义)中语义动作(语义规则)的书写形式与实际实现。

鉴于语义动作等在翻译方案中仅是一个字符序列,要执行它们,实现语义分析,就必须有一个符号处理程序来分析处理这些语义动作(字符序列),产生相应的目标代码从而实现这些语义动作。这要与语法规则相结合,其复杂程度相当于构造一个高级程序设计语言编译程序。

本书尝试以实习题的形式探讨编译程序构造全过程的实现,使读者对于从源程序字符串到等价的目标代码的翻译全过程有深刻的理解。

概括起来,本书旨在:

- 简略回顾与编译程序构造相关的概念,使读者加深对编译原理课程要点的理解与掌握;
- 简略介绍程序(软件)开发方法,尤其是应用 C、C++ 与 VC++ 系统平台来设计与实现应用程序界面,使读者了解与掌握这些系统平台的初步使用;
- 结合编译程序的构造,把各个阶段有机地联系起来,使读者深切地体会从源

程序到目标代码的实际实现全过程。为编译原理课程提供上机实验题的富有启发性的参考资料。

笔者殷切期望,通过对本书的阅读,读者将能:

- 了解如何分析一个问题,明确自己的思路,从而有助于解决问题;
- 掌握各种实际实现方法和技巧,从而有助于读者对其他程序(软件)的研制。

若能如此,对笔者将是莫大的鼓舞。

本书的内容主要源自笔者历年来的教学和科研,其中包括了笔者部分本科生和研究生的实践成果。感谢曾奇、严萍与李进为本书积累了部分资料。感谢张福炎教授对本书成稿的大力支持。感谢李意勉女士,在她的支持和大力帮助下,本书才能顺利完成。最后,感谢资丽芳编辑,是她的关心、大力支持和辛勤工作,才使笔者得以实现奉献此书给广大读者的愿望。

本书是在编译原理课程领域的一个初步尝试,限于笔者水平限制,错误与不妥之处在所难免,综观全书,也不乏需要改进与完善之处。欢迎读者批评指正。

联系地址:南京大学计算机科学与技术系(210093)

E-mail 地址:zhangxr0@sina.com 或 zhang_xinger@hotmail.com

作 者

2005年2月于南京大学计算机科学与技术系

目 录

第二版前言

第一版前言

第一篇 概 论

第 1 章 编译程序概述	1
1.1 编译程序及其构造	1
1.2 编译程序构造实践的必要性	5
1.3 编译程序实现要点	6
1.4 本书阅读指南	8
第 2 章 实践指南	13
2.1 程序(软件)的一般研制过程.....	13
2.1.1 中大型软件的开发	13
2.1.2 一般程序的研制	14
2.2 界面的设计与实现.....	19
2.2.1 界面设计的必要性	19
2.2.2 界面的风格	20
2.2.3 界面的设计	23
2.2.4 界面的实现	26
2.3 上机实习报告及其设计.....	54
2.3.1 书写上机实习报告的必要性	54
2.3.2 上机实习报告的设计	54
2.3.3 上机实习报告举例	62

第二篇 实 践 篇

第 1 章 文法及相关概念	71
1.1 基本概念.....	71
1.1.1 文法与句子	71
1.1.2 文法等价变换	72
1.1.3 句型分析.....	73
1.2 主要数据结构.....	73

1.3 实习题	76
实习题 1.1 文法的输入输出	76
实习题 1.2 推导的建立	85
实习题 1.3 Chomsky 文法类识别	93
实习题 1.4 压缩文法的等价变换	97
实习题 1.5 消去左递归的文法等价变换	106
实习题 1.6 从推导建立语法分析树	112
实习题 1.7 文法管理系统	120
第 2 章 词法分析	121
2.1 基本概念	121
2.2 主要数据结构	124
2.3 实习题	128
实习题 2.1 状态转换图的生成	128
实习题 2.2 运行状态转换图	133
实习题 2.3 有穷状态自动机(FA)的构造	137
* 实习题 2.4 NFA 的确定化	140
实习题 2.5 运行 FA	144
实习题 2.6 词法分析程序的实现	148
实习题 2.7 正则文法管理系统	157
第 3 章 语法分析——自顶向下分析技术	158
3.1 基本概念	158
3.1.1 自顶向下分析技术	158
3.1.2 递归下降分析技术	159
3.1.3 预测分析技术	159
3.2 主要数据结构	160
3.3 实习题	161
实习题 3.1 文法之 first 集合与 follow 集合的计算	161
实习题 3.2 递归下降识别程序的构造	167
实习题 3.3 LL(1)文法的判定	172
实习题 3.4 预测分析表的构造	176
实习题 3.5 应用预测分析技术句型分析	180
* 实习题 3.6 自顶向下分析系统	188
第 4 章 语法分析——自底向上分析技术	190
4.1 基本概念	190
4.1.1 自底向上分析技术	190

4.1.2	算符优先分析技术	190
4.1.3	LR(k)分析技术	192
4.2	主要数据结构	194
4.3	实习题	197
实习题 4.1	有穷字母表上关系传递闭包的计算(Warshall 算法)	197
实习题 4.2	算符优先矩阵构造	200
实习题 4.3	应用算符优先分析技术句型分析	204
实习题 4.4	优先函数构造	209
*实习题 4.5	SLR(1)分析表自动构造	214
实习题 4.6	应用 LR(1)分析技术句型分析	220
*实习题 4.7	自底向上分析系统的设计与实现	230
第 5 章	语义分析和目标代码生成	231
5.1	基本概念	231
5.2	主要数据结构	234
5.3	实习题	238
实习题 5.1	注释分析树生成	238
实习题 5.2	基于自顶向下方式翻译方案的属性值计算	249
实习题 5.3	基于自底向上方式翻译方案的属性值计算	254
实习题 5.4	赋值语句目标代码生成	262
实习题 5.5	条件语句目标代码生成	273
实习题 5.6	while 语句目标代码生成	282
实习题 5.7	赋值语句四元式序列生成	290
实习题 5.8	条件语句四元式序列生成	295
实习题 5.9	赋值语句三元式序列生成	305
实习题 5.10	从四元式序列生成目标代码	309
第 6 章	代码优化	317
6.1	基本概念	317
6.2	主要数据结构	318
6.3	实习题	319
实习题 6.1	流图构造	319
实习题 6.2	构造 DAG	325
实习题 6.3	从 DAG 还原成四元式序列	332
参考文献		338

附录(见光盘)

附录一 第二篇第 5 章语义分析中语义子程序的设计(C 型程序)

实习题 5.2 基于自顶向下方式翻译方案的属性值计算

实习题 5.4 赋值语句目标代码生成

实习题 5.5 条件语句目标代码生成

实习题 5.6 while 语句目标代码生成

实习题 5.7 赋值语句四元式序列生成

实习题 5.8 条件语句四元式序列生成

实习题 5.9 赋值语句三元式序列生成

附录二 样本 C 语言程序(可在 Turbo C 2.0 环境下运行)

1. 文法输入输出 Iogammar. c

2. 词法分析 Lexical. c

3. 应用 LR 分析技术句型分析 LRParsing. c

4. 赋值语句目标代码生成 CodeAssign. c

附录三 在 Visual C++ 开发平台上开发文法输入输出系统的程序代码

附录四 在 Visual C++ 开发平台上开发应用系统的样本程序(可在 VC++ 6.0 环境下运行)

附录五 以赋初值方式给出的各章实习题调试实例

第一篇 概 论

第 1 章 编译程序概述

1.1 编译程序及其构造

1. 编译程序的引进

一个高级程序设计语言程序归根结底是一个基本符号序列,或进一步说是一个字符序列。例如,下列 C 语言程序:

```
main( )
{ float x,y;
  printf("Input values of x and y. \n");
  scanf("% f", % f", &x, &y);
  if(x>y)
    printf("max(x,y) = % f",x);
  else
    printf("max(x,y) = % f",y);
}
```

当上机运行时,计算机并不能直接接受、理解与运行该程序。计算机所能直接接受、理解与运行的是机器语言的二进位串,而现在却是字母 m、a、i、n 与括号 (、) 等字符。为此需要一个软件处理这个字符序列,识别出这是 main,那是 float 等,进一步把它们翻译成可由计算机执行的等价的目标代码,达到运行且获得所期望效果的目的。这种把高级程序设计语言程序翻译成等价的低级语言程序的软件就是编译程序。高级程序设计语言程序称源程序,所翻译成的等价的低级程序设计语言程序称为目标程序或目标代码。

编译程序是一种符号处理工具,它把高级程序设计语言源程序翻译成等价的低级语言目标代码。此处低级语言可以是汇编语言,也可以是机器语言。等价指的是源程序和目标程序两者效果必须完全一样,也就是说,当源程序正确时运行得到的效果,运行目标代码时同样能得到;若源程序有错误,目标代码也将有相应的反映。

以翻译方式运行的任何一种高级程序设计语言程序都必须有编译程序。这种

编译程序在当前都被开发成了集编辑、编译、调试与运行于一体的程序设计语言支持环境,例如, Turbo PASCAL、Turbo C、Borland C++ Builder 与 Visual C++ 等。可视化版本编译程序的引进更为用户设计应用程序界面带来了极大的方便。由于编译程序一般较为庞大,结构复杂,往往称为编译系统。这些编译系统软件往往由著名的公司开发,具有人机接口良好、使用方便、目标代码质量高等优点。

这些编译系统软件垄断了绝大部分的软件市场,国内自行开发编译系统已近乎没有,一些学生可能会说,我们毕业后不会去开发编译程序,没有必要学习编译原理课程。殊不知,编译程序作为符号处理的工具,其基本原理和技术有着典型性和广泛性,在软件工程、逆向工程、软件再工程、语言转换及其他领域都有着广泛的应用。例如,语法制导的程序编辑软件、程序结构分析软件、程序调试与测试软件、反汇编软件、符号执行软件与超文本语言的实现等都需要应用编译程序构造的基本原理与技术;对软件的开发也有一定的启发与指导作用。编译程序作为符号处理的工具,只要涉及符号处理,就需要采用编译程序实现的基本原理与技术。

2. 编译程序的功能

基于对编译程序作用的认识,总的来说,编译程序的功能是把高级程序设计语言源程序翻译成等价的低级语言目标代码。

鉴于目标代码一般总是与机器相关的,也就是说为某种型号计算机产生机器语言目标代码,必须对此种型号计算机的指令系统(包括指令格式、不同操作码和寻址方式等)有详细的了解。这样需要花费大量的时间,而且仅限于一种型号的计算机,有相当的局限性。因此在学校教学中往往采用臆设的虚拟机,即按照教学的要求设计指令格式、操作码和寻址方式。这样就无须把大量时间花在对具体计算机细节的了解,而且可以根据需要,随时扩充。

编译程序的输入是某种高级程序设计语言源程序字符序列(字符串),以 C 语言为背景时输入就是 C 程序字符串,输出是等价的虚拟机目标代码。

一个编译程序输入高级程序设计语言源程序字符串,要识别出一个个具有独立含义的最小语法单位——符号(或称单词),例如,识别出 main、float 与 if 等,并把它们加工处理成内部中间表示——属性字,这就是词法分析。词法分析时还进行一些力所能及的工作,例如,删除注解、空格与不可见的控制字符等。对识别出的连续若干个符号,编译程序要识别出各个语法成分,例如,if 语句、输入输出语句与 main 函数定义,直到识别出程序,这时属性字序列被加工处理成了一种内部中间表示,即推导或语法分析树等,这就是语法分析。

对于语法分析树或推导这些内部中间表示,识别所相应的各个语法成分的含义,为此进行确定类型、类型检查,以及识别语义并进行相应的处理,例如,生成一些表格和生成目标代码或者内部中间表示代码,这就是语义分析。语义分析时还

进行若干静态语义检查。

引进内部中间表示代码的目的是为了改进目标代码的质量。例如,在编译时刻计算好 $2 * 3.1416$ 的值,运行时刻直接使用该乘积值 6.2832,或者把循环时每次重复都不改变值的表达式外提,以大幅度减少重复计算量等,可以提高目标代码运行效率。这种在编译时刻为改进目标代码的质量而进行的工作就是目标代码优化。

概括起来,一个编译程序一般要经历下列五个阶段,即词法分析、语法分析、语义分析、目标代码优化和目标代码生成。

词法分析:扫描(读入)源程序字符序列,识别开单词(符号),把它们加工处理成内部中间表示形式(属性字),并进行力所能及的一些工作(如删除注解与空格等无效字符,以及预处理工作)。

语法分析:对符号序列识别出各个语法成分,把它们加工处理成等价的内部中间表示(语法分析树或推导)。

语义分析:确定类型、类型检查、识别含义并做相应处理(生成目标代码或内部中间表示代码),还做一些静态语义检查。当生成中间表示代码时,这些中间表示代码可以是抽象语法树、逆波兰表示、四元式序列和三元式序列。

目标代码优化:在中间表示代码(主要是四元式序列或三元式序列)的基础上进行基本块的优化和与循环有关的优化,直至全局优化。代码优化后生成优化了的中间表示代码。

目标代码生成:当语义分析所生成的是中间表示代码时,就需要对这中间表示代码来生成目标代码。为了教学的目的,此目标代码是虚拟机目标代码。不言而喻,此虚拟机目标代码是不能直接运行的。为了能运行,需要设计并实现一个解释程序来模拟执行虚拟机目标代码。编译原理通常不讨论此问题。

上述这五个阶段有机地结合起来就可实现相应的编译程序原型。

在实际的编译系统中,往往有不同的结构。例如,对源程序字符序列从头到尾扫描一次就完成词法分析、语法分析、语义分析与目标代码生成;也可以词法分析时从头到尾扫描一次源程序字符序列,生成相应的属性字序列,再以属性字序列作为语法分析与语义分析的输入,同时完成语法分析与语义分析,生成相应的中间表示代码,最终从头到尾地扫描中间表示代码而生成目标代码。一个编译系统,从头到尾扫描源程序或中间表示代码一次,就称为一遍(或一趟),总共几遍就称为几遍的编译系统。在本书的讨论中,每个阶段看作一遍。

这里我们可以看到,在进行编译程序实践时,除了确定语言背景,即确定一个要进行编译程序实践的程序设计语言或其子集外,重要的是设计各个中间表示,即属性字的机内表示、语法分析树或推导的机内表示、中间表示代码(四元式等)的机内表示等。最后要设计虚拟机指令系统。

要说明的一点是:在编译原理课程中,编译期间的每一个阶段,除了词法分析以源程序字符序列作为输入外,其余各阶段都是以符号序列作为输入,并不是实际的内部中间表示。例如,语法分析以符号序列作为输入,并不以属性字序列作为输入,又如,语义分析阶段也不是以语法分析树作为输入,而是仍然以符号序列作为输入。这并不失一般性,事实上,我们可以在取符号函数 GetSymbol 中做相应的变动即可。下面以语法分析为例说明。

假定属性字的数据结构定义如下:

```
struct AttributeWordType /* 属性字之结构 */
{ int tag; /* 特定符号特征:tag = 1:特定符号,tag = 0:非特定符号 */
  int kind; /* 符号类 */
  char SymbolValue[MaxLength];
  /* 符号值,简单起见,就是符号本身,MaxLength 是符号最大长度 */
  ...
} AttributeWordS[MaxAttriWNumber];
```

为简单起见,用数组存放属性字序列,且假定符号(属性字)数最多为 MaxAttriWNumber 个。假定 i 是当前符号计数器,只需在 GetSymbol 函数中回送 AttributeWordS[i]. SymbolValue,使用函数调用语句:

```
strcpy(symbol, GetSymbol( ));
```

在变量 symbol 中设置好当前符号。又如,对于语法分析树的数据结构可如下定义:

```
struct NodeType
{ char SymbolName[MaxLength];
  /* 实现中可以用符号在相应符号集中的序号代表 */
  int NumberOfNode; /* 结点序号 */
  int NumberOfFatherNode; /* 父结点序号 */
  int NumberOfBrotherNode; /* 左兄结点序号 */
  int NumberOfSonNode; /* 右子结点序号 */
} 语法分析树[MaxNodeNumber];
```

同样地,为简单起见,用数组存放语法分析树结点,结点数最多为 MaxNodeNumber 个,只需同样地在 GetSymbol 函数中回送语法分析树[i]. SymbolName,使用函数调用语句 strcpy(symbol, GetSymbol())同样地可以在变量 symbol 中置好当前符号。

中间表示(数据结构)随实现而不同,省略了对具体表示法的描述,大大简化了讨论,但又不失一般性。本书讨论的编译程序各阶段的输入在语法分析与语义分析阶段都以符号序列作为输入,而不是具体的属性字序列或语法分析树。

3. 编译程序的构造

一般的编译程序由两大部分组成,即前端与后端。前端进行分析,包括词法分析、语法分析与语义分析;后端进行综合,包括代码优化与目标代码生成。

前端往往涉及一个程序设计语言的定义,而与具体目标机无关。后端在生成目标代码时一般与目标机有关,必须考虑各个具体细节,如指令系统(包括指令格式、不同的操作码和寻址方式)、寄存器个数与使用约定,以及特殊指令的使用等。当目标机是虚拟机时,情况要简单得多。

本书讨论的编译程序可以看作五遍的编译程序,即词法分析、语法分析、语义分析、代码优化和目标代码生成各看成一遍。这样的好处是结构清晰,重点突出,有利于理解与掌握编译程序构造原理及其实现。

1.2 编译程序构造实践的必要性

编译原理以形式语言理论相关概念为基础,尤其是实现语义分析的语法制导翻译,以基于形式定义的属性文法为基础,其本身决定了有较强的理论性,但所讨论的编译程序作为高级程序设计语言支持软件和符号处理的工具,它又有很强的实践性。要学好编译原理,必须兼顾理论性与实践性两方面。

(1) 通过实践,加深对概念的理解。

这里以一个简单的例子说明。

例如,设有文法 $G[Z]: Z ::= Zbc | AE \quad A ::= da$, 要求写出其非终结符号集合 V_N 。少数学生可能会写出 $V_N = \{Z, A, E\}$ 与 $V_T = \{a, b, c, d\}$, 因为在他(她)们心目中,大写字母代表非终结符号,小写字母代表终结符号。但若正确地实现了自动构造 V_N 与 V_T 的程序,就会发现对于上述文法应是 $V_N = \{Z, A\}$, $V_T = \{a, b, c, d, E\}$, 从而形成一个深刻的概念:只有出现在重写规则左部的才是非终结符号,与大小写无关。

另外一个例子是有穷状态自动机。随着有穷状态自动机构造程序的实现,将深刻体会到 DFA 与 NFA 之间的两个根本区别,即开始状态个数的差异与状态映像之间的差异。

其他例子是在采用 SLR(1)方法构造 SLR(1)分析表时后继项集的计算,容易对不同项集中相同的终结符号 T 之 T -后继取相同的后继项集名。通过 SLR(1)分析表自动生成程序的设计与实现,就会体会到必须不仅项的个数相同,而且所包含的项全部相同时,两个项集才能有相同的项集名。

事实上,对编译原理中某个方面编写实现程序的过程本身就是不断认真学习概念和消化理解概念的过程。

通过编译程序构造实践,对相关部分的概念理解和掌握将提升到一个新高度。

(2) 通过实践,深刻理解概念与具体实现之间存在的差异。

在编译原理课程上讲解编译程序的构造原理,应该说仅是概念上的,譬如说,应用 LR(1)分析技术进行语法分析时,分析表是关键,其中 ACTION 部分的元素确定分析动作。移入是 S_i ,归约是 r_j ,接受是 acc,而出错是空白元素。但具体实现时原封不动便难以实现。为方便处理,用一个整数值表示: >0 为移入, <0 为归约,出错为 0,而接受则可用某个特殊值,如 -111 或 999 等。诸如此类,通过实践,摸索出实际实现时可以采用的合适方法与技巧,为今后程序研制积累经验。

(3) 培养与提高上机动手与开发软件能力。

学习的一个重要方面是理论与实践相结合。学生做习题是实践的一个手段,通过做习题加深对概念的理解。仅仅做习题的一个缺陷是并不能提高学生上机动手与开发软件的能力。

在编译原理课程中安排学生一定的上机时间量,完成若干题目的上机实习,无疑对学生提高软件开发能力有显著的效果。学生将了解开发这类软件的全过程;经历这一全过程,从中积累知识和经验,了解在开发一个软件时所应经历的步骤,在每一步应考虑哪些问题,提高编程能力,尤其是提高上机动态调试的能力。这对于计算机专业学生的进一步发展有莫大的帮助。

本书期望通过实践,读者对从源程序字符序列到目标代码的生成全过程有一个深切的感性体验,并且在如何开发软件上对读者有启发提示与指导作用。通过阅读本书,结合上机实践,能大幅度提高软件开发与上机动手能力。

1.3 编译程序实现要点

对于一个完整而独立的编译程序的实现来说,其基本步骤大致如下:

步骤 1 确定背景程序设计语言。这可以是某个新设计的高级程序设计语言,也可以是现有的某个程序设计语言的扩充或子集。

步骤 2 确定编译程序设计指标与设计要求,包括确定目标机目标语言。

步骤 3 确定总体结构,包括分多少遍,各遍的功能与各遍之间的联系,并设计各个中间表示语言。

步骤 4 确定各遍的设计方案,并实现之。

1. 确定背景程序设计语言

编译程序的工作是把高级程序设计语言源程序翻译成等价的低级语言目标代码。因此必须确定是哪一种高级程序设计语言(源语言 SL)。依据该程序设计语言的定义(词法、语法与语义)来实现编译程序。