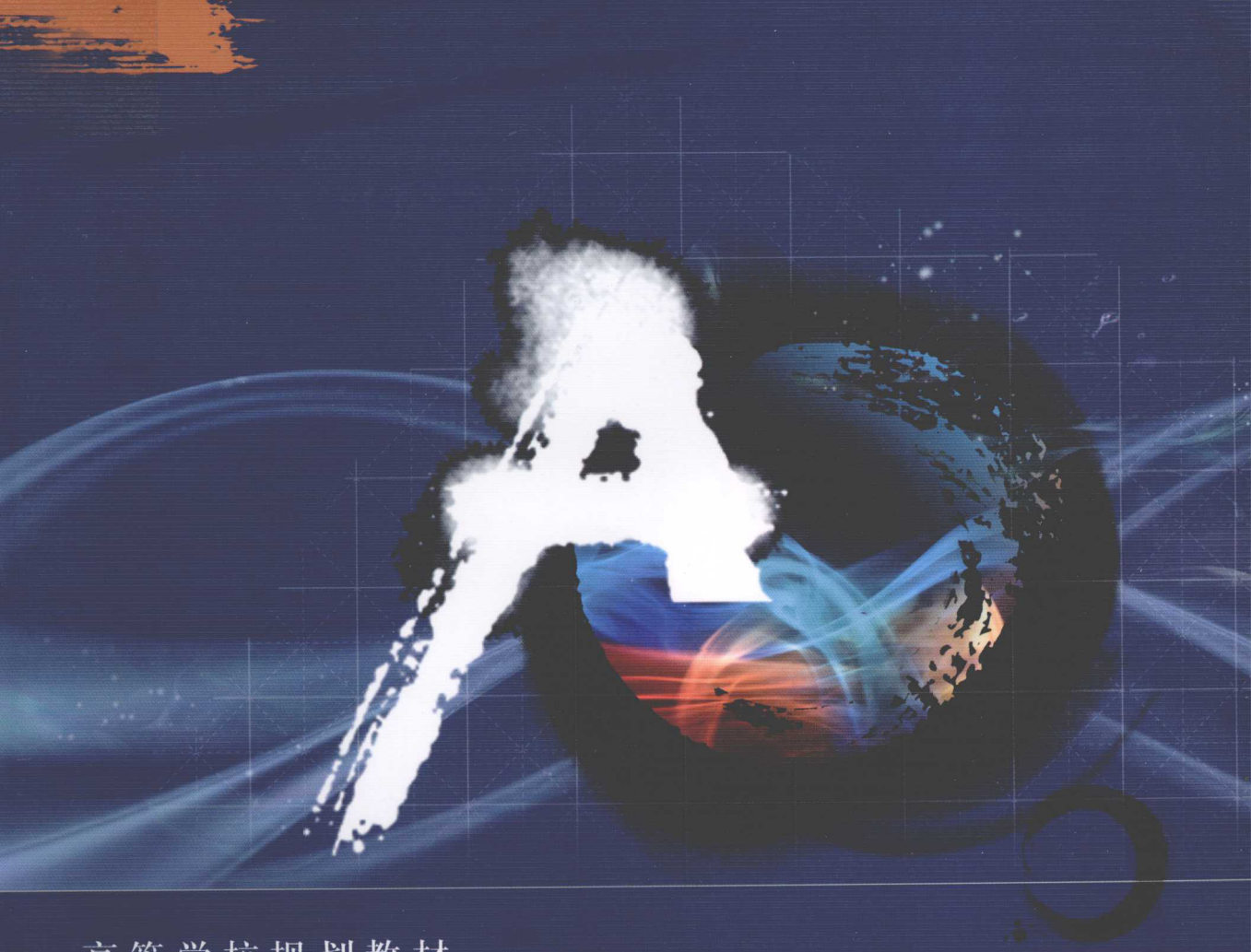

高等学校规划教材

编译原理简明教程

◎杨 明 姜乃松 蔡维玲 编著

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

高等学校规划教材

编译原理简明教程

杨 明　姜乃松　蔡维玲　编著

電子工業出版社
Publishing House of Electronics Industry
北京 • BEIJING

内 容 简 介

本书是精品课程建设成果。全书分 12 章，内容包括：编译概述、文法和语言、词法分析与有限自动机、自顶向下的语法分析、自底向上的语法分析、属性文法、语义分析与语法制导的翻译、运行时环境、目标代码生成、代码优化、并行编译技术、面向对象语言的语法制导翻译。本教材先进、实用、简洁、通俗，可读性好，还免费提供电子教案和习题解答。

本书适合作为高等学校计算机及相关专业教材，也可作为 IT 技术人员的参考书。

图书在版编目（CIP）数据

编译原理简明教程 / 杨明，姜乃松，蔡维玲编著. —北京：电子工业出版社，2012.1
高等学校规划教材
ISBN 978-7-121-14596-4

Ⅰ. ①编… Ⅱ. ①杨…②姜…③蔡… Ⅲ. ①编译程序—高等学校—教材 Ⅳ. ①TP314

中国版本图书馆 CIP 数据核字（2011）第 186722 号

策划编辑：童占梅
责任编辑：童占梅
印　　刷：
装　　订：北京市李史山胶印厂
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：787×1 092　1/16　印张：15.25　字数：384 千字
印　　次：2012 年 1 月第 1 次印刷
印　　数：3 000 册　　定价：29.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

致阅读本书的师生朋友

春华秋实，岁月有辛勤付出才美丽；

桃李芬芳，人生看学生成才而快乐！

学习是终生成长的过程，她为我们打开认识世界、自然、社会乃至我们自己的窗口，让我们有勇气、责任、知识和能力来改变现状，创造未来。

人生因学习而改变，世界因你而不同。

——编辑心语

前　言

编译原理是计算机专业的重要专业课之一，主要涉及程序设计语言编译程序构造的基本原理和基本实现方法。本书是精品课程建设的成果，编者根据多年讲授“编译原理”课程的经验，编写了这本适用于高校计算机专业本科生使用的编译原理简明教材。

本教材分为 12 章，内容包括：编译概述、文法和语言、词法分析与有限自动机、自顶向下的语法分析、自底向上的语法分析、属性文法、语义分析与语法制导的翻译、运行时环境、目标代码生成、代码优化、并行编译技术、面向对象语言的语法制导翻译。本教材由七大模块组成，如图 1 所示。

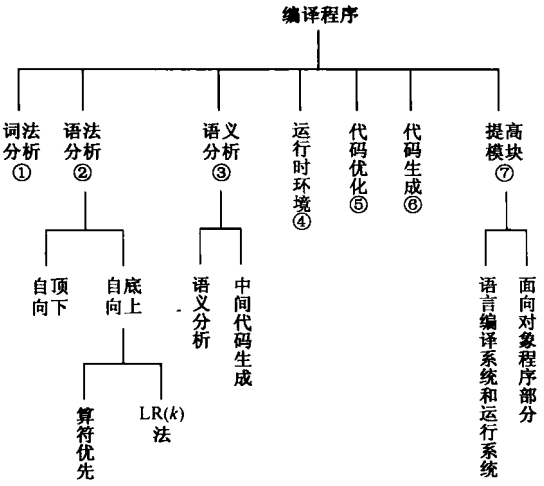

图 1　编译原理简明教程课程内容结构图

各模块具体内容如下：

① 词法分析模块（包括 1 个子模块）——涉及编译程序的词法分析部分；

② 语法分析模块（包括 3 个子模块）——涉及编译程序的语法分析部分；

③ 语义分析模块（包括 2 个子模块）——涉及编译程序的语义分析部分；

④ 运行时环境模块（包括 1 个子模块）——涉及编译程序的数据存储；

⑤ 代码优化模块（包括 1 个子模块）——涉及编译程序的代码优化部分；

⑥ 代码生成模块（包括 2 个子模块）——涉及编译程序的后端代码生成部分；

⑦ 提高模块（包括 1 个子模块）——涉及编译程序的语言编译系统和运行系统，以及面向对象程序部分，学生可以自学。

本教材参考学时为 60 学时，书中的**例题和习题以 C 和 C++语言为背景**。为了方便学生了解教材的重点，教材在每一章的开头都列出了本章的主要学习内容，在每一章的末尾给出本章小结，同时每章还附有相关练习题。除此之外，本教材主要有以下几个特点：

1．先进性

内容先进。在精选经典内容并保持完整性的基础上，介绍了目前较新的内容如：垃圾回收、寄存器着色分配算法等。

2．实用性

结合当前软件技术，选择C/C++语言作为背景，讨论了C语言和C++面向对象语言编译技术，在目前教材中不多见。读者通过本课程学习，还可达到深入理解与掌握C/C++语言的目的。**并免费提供电子课件和习题解答**。

3．简洁

避开高深的理论，简洁并富有启发性地引导、展开所需讲述的基本原理与技术。

4．通俗

一图抵千言。用丰富图例，以图文并茂形式展现内容，读者容易掌握与理解。

本教材第1～4章由杨明老师编写，第5～9、12章由姜乃松老师编写，第10～11章由蔡维玲老师编写。在本书的编写过程中，南京师范大学的许多老师和同学提出了宝贵的意见，在此表示诚挚的感谢！

限于作者水平，书中难免有疏漏和错误之处，恳求读者批评指正！

作者E-mail：myang@njnu.edu.cn。

编著者

于南京师范大学

目　录

第1章 编译概述

本章是编译原理的基本知识，主要介绍编译程序、编译过程、编译程序的结构与分类、编译程序的生成。

1.1 翻译程序

目前，世界上存在着多种语言，人们为了交流和通信的便利，需要实现各种语言之间的翻译。人与计算机之间的信息交流，同样存在一个翻译问题。人类语言存在多样性，计算机语言也存在多样性，因为不同类型计算机均有自己特有的指令系统，即所谓的该种机器的机器语言。由于人们直接运用机器语言编写程序极为不便，且编写的机器语言程序难读、难维护、结构性差，因此目前程序员主要采用接近自然语言的高级程序设计语言来编写程序，然后再借助特定的软件（翻译程序）将它翻译成机器语言程序。目前，常用的高级程序设计语言包括 FORTRAN，Pascal，C，C++，Java 等。而机器语言和面向硬件的语言被称为低级语言，如各种汇编语言是面向硬件的语言，属于低级语言。

所谓翻译程序是指这样的程序，它把用某种程序设计语言（源语言）编写的程序（源程序）翻译成与之等价的另一种语言（目标语言）的程序（目标程序），其作用如图 1.1 所示。

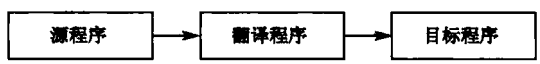

图 1.1 翻译程序的功能

编译是指如果一个翻译程序的源语言是某种高级语言，其目标语言是相对于某一种计算机的汇编语言或机器语言，则称这种翻译程序为编译程序（或称编译器）。可见，编译程序是一种翻译程序，它将高级语言编写的源程序翻译成等价的机器语言或汇编语言的目标程序。特别需要指出的是，对应于机器语言的目标程序可在对应的机器上运行，而对应于汇编语言的目标程序不能在目标机器上运行，还需要进一步通过汇编程序进行翻译，将其翻译成与之等价的机器语言的目标程序，如图 1.2 所示。

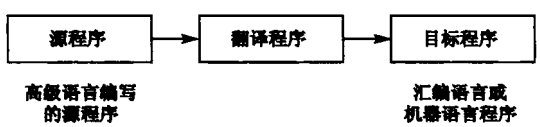

图 1.2 编译程序功能

编译原理课程重点介绍如何将用高级语言编写的源程序翻译成与之等价的某种中间语言或汇编语言的目标程序。这是因为不同类型机器具有不同的指令系统，且经编译程序翻译得到的中间语言或汇编语言的目标程序可方便翻译成与其等价的机器语言的目标程序。

对高级语言编写的源程序，其执行除了可用编译程序将其翻译成可在目标机器上运行的目

标程序外，也可通过解释程序进行源程序的执行。所谓解释程序是指翻译一句执行一句，不生成目标文件，直接执行源代码文件，如 BASIC 程序，可用 QBASIC 直接解释一句执行一句。而汇编程序是指把用汇编语言编写的源程序翻译成与之等价的机器语言的目标程序。

总体上看，翻译程序有三种方式：编译程序、解释程序、汇编程序。翻译后的程序称为目标代码或目标程序。

1.2　编译过程和编译程序的基本结构

编译程序的功能是，将高级语言编写的源程序翻译成与之等价的机器语言或汇编语言表示的目标程序。也就是说，编译过程是一种语言的翻译过程，其工作原理类似于外文的翻译过程。

例如，若要将英语句子“I wish you success”翻译成汉语句子，首先要对句子中出现的单词进行词法分析。构成该句子的单词有 I，wish，you 和 success。通过查字典可知，I 是代词，wish 是动词，you 是代词，success 是名词，即这些单词都是合法的单词。那么由这些有效的单词能否构成符合英语语法的句子呢？这就需要进行语法分析。按照英语的主、谓、宾语法结构可以看出，“I wish you success”是一个符合英语语法的合法句子。至此，我们还没有完成该句子的翻译，因为我们的目标是将其翻译成汉语句子，这还需要我们进行更深入的语义分析。由于 I 的意思是“我”，wish 的意思是“希望”、“祝愿”等，you 的意思是“你”、“你们”，success 的意思是“成功”、“成就”等。于是，英文句子“I wish you success”可直接翻译成汉语句子“我祝你（们）成功”。然而，英文句子直接翻译成汉语句子往往比较生硬，因此结合上下文关系及汉语语法规则，可对汉语句子进一步修饰，使其更加优美。通过上述三个过程，即词法分析、语法分析、语义分析及优化，句子“I wish you success”可翻译成：“祝你（们）成功”。

类似外文句子的翻译，编译程序将输入的源程序翻译成与之等价的输出用的目标程序的工作过程大致可分为 5 个阶段：词法分析、语法分析、语义分析和中间代码生成、代码优化、目标代码生成，其结构框图如图 1.3 所示。

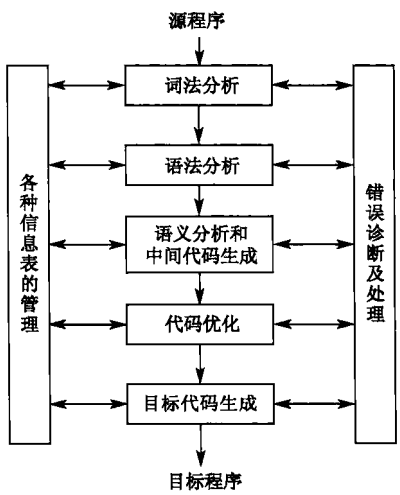

图 1.3　编译程序结构框图

由图 1.3 可知，编译程序的 5 个阶段是依次执行的，每个阶段均有自己特定的任务。各阶段的任务如下：

① 词法分析。对构成输入源程序的字符串进行扫描和分解，识别出一个个单词符号，如标识符、常数、界限符等。

② 语法分析。在词法分析的基础上，根据语言的语法规则，把单词符号分解成各类语法单位（语法范畴），如短语、句子、程序段等。

③ 语义分析与中间代码生成。首先对各种语法范畴进行静态语义检查，如果正确则进行中间代码的翻译。该阶段遵循的是语言的语义规则，通常使用属性文法描述语义规则。中间代码的形式有四元式、三元式、间接三元式、逆波兰式和树形表示。

④ 代码优化。优化所遵循的是程序的等价变换规则。其方法有公共子表达式的提取、循环优化、删除无用代码等。

⑤ 目标代码生成。把中间代码（或经优化处理后）变换成特定机器上的低级语言代码。它有赖于硬件系统结构和机器指令的含义。如何充分利用寄存器、合理选择指令以生成尽可能短且有效的目标代码等都与目标机器的硬件结构有关。

同时，为了有效关联各阶段的任务，在图 1.3 中各种信息表（存储信息的各种数据结构，如符号表）及其管理必不可少。也就是说，编译程序在工作过程中依赖各种表格，这些表格登记了源程序的各类信息和编译各阶段的相关信息。合理设计和使用表格是编译程序构造的重要任务之一。在编译程序使用的表格中，最重要的表格是符号表，它包括源程序中出现的每个单词及单词的各种属性等。例如，一个单词是常量名、变量名，还是过程名；若是变量名，它的类型是什么、所占内存有多大、地址是什么。通常，在词法分析时，填入单词的各种属性到符号表中，而当处理到单词时，查找符号表，对单词的属性进行查证。当然，对某些单词，词法分析过程并不能完成所有属性的登记，如对变量名而言，其类型等属性要等到语义分析时才能确定，而变量地址可能要到目标代码生成时才能确定。

此外，为方便对源程序的有效修改，图 1.3 中还包含对各种词法、语法错误的诊断和处理部分，即一个编译程序不仅能对书写正确的程序进行翻译，而且能对出现在源程序中的错误进行处理。若源程序有错误，编译程序应尽可能多地发现错误，并把错误的性质和位置准确地告之用户，以便用户快速有效地进行纠正。我们总是希望一个好的编译程序能发现源程序中几乎所有的错误，并对部分错误自动纠正。然而，自动纠错的代价非常大。 总体上看，编译程序能发现的错误通常分为语法错误和语义错误两大类。语法错误是指源程序中不符合语法（或词法）规则的错误，如“非法字符”、“不合法的标识符”等词法错误，“括号不匹配”、“缺少”之类的语法错误等。语义错误是指源程序中不符合语义规则的错误，这些错误通常在语义分析或运行时才能检测出来。语义错误通常包括说明错误、作用域错误、某运算符与其运算对象的类型不相容等。

特别需要注意的是，图 1.3 中的 5 个阶段仅是逻辑功能上的一种划分，不一定是执行时间上的先后关系。具体实现时，依据源语言的特点、设计要求、使用对象和计算机条件（如主存容量），往往将编译程序的组织分为若干遍（Pass）。所谓“遍”是指对源程序或其等价的中间语言程序从头到尾扫描一遍，并完成规定的加工处理过程。例如，可将图 1.3 中 5 个阶段的工作结合在一起，对源程序从头到尾扫描一遍来完成编译的各项工作，这种编译程序称为一遍扫描的编译程序。而在多遍扫描的编译程序结构中，每遍可完成上述某个阶段的一部分、全部或多个阶段的工作，且每一遍的工作是从前一遍获得的工作结果开始的，最后一遍的工作结果是目标程序，第一遍的输入则是用户书写的源程序。

相对一遍扫描而言，多遍扫描编译程序的优点是，占存储空间少，各遍所完成的功能相对独立，编译程序逻辑结构清晰。其缺点是，输入/输出开销增大，编译效率降低。因此，究竟采用多少遍扫描来设计编译程序要根据具体情况来确定。

为便于对编译程序各阶段的理解，我们通过一个简单的C程序段来分析5个阶段的任务。

【例1.1】 计算圆柱体侧面积的C程序段如下：

```
float r,h,s;
s=2*3.14159*r*h;
```

第1阶段　词法分析

词法分析阶段的任务是，依据语言的词法规则对构成源程序的符号串进行扫描和分解，识别出一个个的单词（亦称单词符号，或简称符号）。其中，词法规则是单词符号的形成规则，而描述词法规则的有效工具是正规式和有限自动机。通过正规式或有限自动机，可有效确定哪些字符串构成了一个单词符号。上述程序段通过词法分析识别出如下单词符号：

基本字　float
标识符　r，h，s
常数　　2，3.14159
运算符　*
界限符　，　；　=

第2阶段　语法分析

语法分析的任务是在词法分析的基础上，根据语言的语法规则，把单词符号串分解成各类语法单位，如短语、子句、句子（语句）、程序段和程序等。通过语法分析，判断整个输入串是否构成语法上正确的“程序”。其中，语法规则是语法单位的形成规则，它规定了如何从单词符号形成语法单位。

上述程序段，通过语法分解并依据语言的语法规则，识别出单词符号串 s=2*3.14159*r*h 具有<变量>=<表达式>构成<赋值语句>的语法单位。同时，由于 s 是变量，而通过识别可知 2*3.14159*r*h 是一个合法的表达式，因而上述程序段是一个语法上正确的程序段。

第3阶段　语义分析及中间代码生成

该阶段的任务是，对语法分析所识别出的各类语法单位，分析其含义，即对语言的各种语法单位赋予具体的意义，并进行初步翻译（产生中间代码）。该阶段通常包括两方面的工作：① 对每种语法单位进行静态语义检查，如变量是否定义、类型是否正确等；② 若语义正确，则用另一种语言形式（比源语言更接近于目标语言的一种中间代码或直接用目标语言）来描述这种语义。对上述赋值表达式而言，语法分析仅是判别该表达式是否合法，而并不计算其值，即不进行语义分析；同时，语法分析也不检查赋值号右边表达式和左边变量的类型是否一致。而通过语义分析，根据赋值语句的语义，对它进行翻译可得如下的四元式中间代码：

（1）（*，2，3.14159，T_1）
（2）（*，T_1，r，T_2）
（3）（*，T_2，h，T_3）
（4）（=，T_3，—，s）

其中，T_1，T_2，T_3 是编译程序引进的临时变量，存放每条指令的运算结果。对第（2）个四元式，其含义是，将 T_1*r 的值赋给临时变量 T_2，四元式（1）和（3）具有类似的含义。

于是，源语言形式的赋值语句就被翻译为四元式表示的另一种语言形式，这两种语言在表达形式上不同，但在语义上是等价的。

第4阶段 代码优化

优化的任务旨在对中间代码进行加工变换，以期在最后阶段能产生更为高效（省时间和空间）的目标代码。优化的主要策略有：公共表达式的提取、循环优化、删除无用代码等。对上述四元式经局部优化后可得：

（1）（*，6.28318，r，T_2）

（2）（*，T_2，h，T_3）

（3）（=，T_3，－，s）

其中，2和3.14159两个运算对象都是编译时的已知量，在编译时就可算出它的值6.28318，而不必等到程序执行时才计算，即不必生成（*，2，3.14159，T_1）的运算指令。

第5阶段 目标代码生成

目标代码生成的任务是将中间代码变换成特定机器上的绝对指令代码、可重定位的指令代码或汇编指令代码。

1.3 编译程序的分类和生成方法

1.3.1 编译程序的分类

按照用途或侧重点不同来划分，编译程序可分为如下5类。

1. 优化型编译程序

该类编译程序以产生高功效（即存储空间小，而运行时间短）的机器代码为目标，但代价是增加编译程序的复杂性和编译时间。要同时做到存储空间小且运行时间短是很难的，因此，没有最优目标程序可言，总是以满足用户某个重要方面上的优化为目标，很多优化型编译程序提供若干优化选项供用户选择，使其以合理的代价获得所期望的优化效果。

2. 可重定目标型编译程序

一般地，一个编译程序是为一个特定的程序设计语言和一种特定的目标计算机而设计的，从源程序编译生成的目标程序只能在某种特定计算机上运行。可重定目标型编译程序是不重写编译程序中与机器无关部分，而只需重写与机器相关部分，就可改变目标程序的编译程序。因此，可重定目标型编译程序是易移植的编译程序。

3. 诊断型编译程序

此类编译程序专门设计成以帮助开发与调试程序为主要目标，它能仔细地检查程序并发现程序中的错误。一些小错误可自动纠正，如遗漏逗号或括号等。对于不合法的下标、指针误用与不合法的文件管理等，诊断型编译程序能在目标程序中生成可查出运行时刻错误并因运行时刻错误而放弃程序执行的代码。该类编译程序只在程序开发的初始阶段使用，使得目标程序生成时不含诊断代码，以提高运行速度。