普通高等教育“十一五”国家级规划教材

重点大学计算机专业系列教材

编译原理及编译程序构造

张莉 杨海燕 史晓华 金茂忠 高仲仪 编著

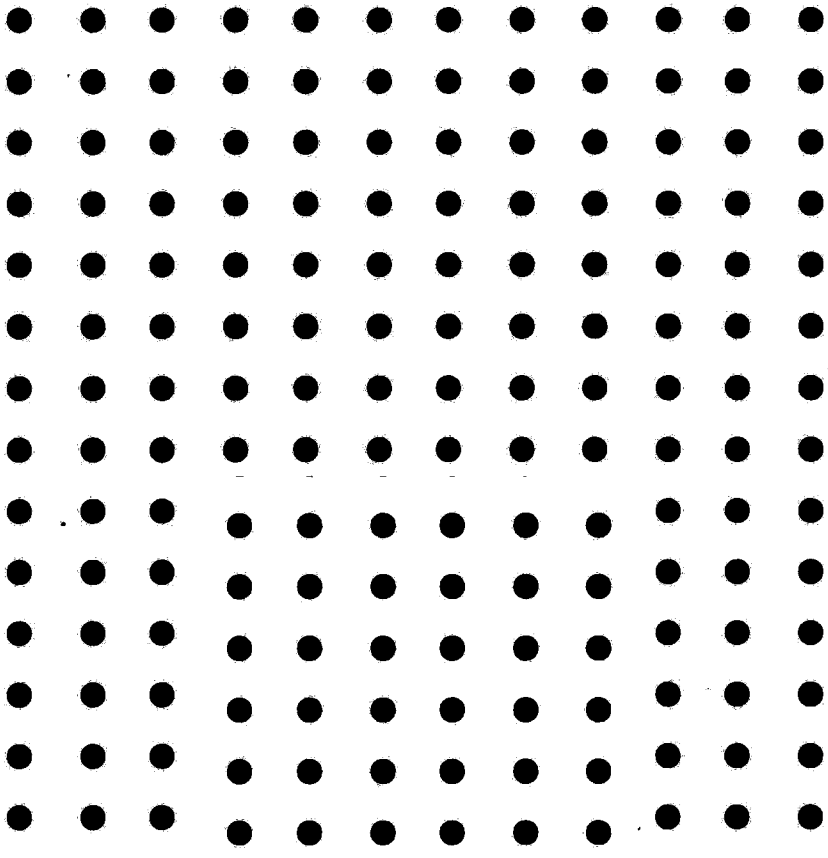
清华大学出版社

普通高等教育“十一五”国家级规划教材

重点大学计算机专业系列教材

编译原理及编译程序构造

张莉 杨海燕 史晓华 金茂忠 高仲仪 编著



清华大学出版社
北京

内 容 简 介

本书全面地介绍编译系统的构造和相关原理与技术。全书共15章，力求展示一个完整的编译过程，在此基础上介绍与编译系统相关的理论和方法。本书围绕这个完整的过程，还介绍并讨论了计算机领域三个非常重要的原理、概念和技术：高级程序设计语言的工作原理、程序模型间的转换方法，以及软件系统的概念。本书强调编译系统的构造及其相关技术，突出对工程师人才的培养要求。书中的算法和示例程序全部采用C语言风格。

本书适合作为高校计算机科学与技术专业本科“编译原理”、“编译技术”等专业课程的教材，也可供相关研究开发人员自学参考。

图书在版编目(CIP)数据

编译原理及编译程序构造/张莉等编著. —北京：清华大学出版社，2011.6
(重点大学计算机专业系列教材)
ISBN 978-7-302-26314-2

Ⅰ. ①编… Ⅱ. ①张… Ⅲ. ①编译程序－程序设计－高等学校－教材 Ⅳ. ①TP314

中国版本图书馆CIP数据核字(2011)第149905号

责任编辑：梁　颖　战晓雷
责任校对：焦丽丽
责任印制：王秀菊
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：北京市世界知识印刷厂
装 订 者：三河市溧源装订厂
经　　销：全国新华书店
开　　本：185×260　**印　张**：28.5　**字　数**：713千字
版　　次：2011年6月第1版　　**印　　次**：2011年6月第1次印刷
印　　数：1～3000
定　　价：45.00元

产品编号：013803-01

INTRODUCTION

出版说明

随着国家信息化步伐的加快和高等教育规模的扩大，社会对计算机专业人才的需求不仅体现在数量的增加上，而且体现在质量要求的提高上，培养具有研究和实践能力的高层次的计算机专业人才已成为许多重点大学计算机专业教育的主要目标。目前，我国共有16个国家重点学科、20个博士点一级学科、28个博士点二级学科集中在教育部部属重点大学，这些高校在计算机教学和科研方面具有一定优势，并且大多以国际著名大学计算机教育为参照系，具有系统完善的教学课程体系、教学实验体系、教学质量保证体系和人才培养评估体系等综合体系，形成了培养一流人才的教学和科研环境。

重点大学计算机学科的教学与科研氛围是培养一流计算机人才的基础，其中专业教材的使用和建设则是这种氛围的重要组成部分，一批具有学科方向特色优势的计算机专业教材作为各重点大学的重点建设项目成果得到肯定。为了展示和发扬各重点大学在计算机专业教育上的优势，特别是专业教材建设上的优势，同时配合各重点大学的计算机学科建设和专业课程教学需要，在教育部相关教学指导委员会专家的建议和各重点大学的大力支持下，清华大学出版社规划并出版本系列教材。本系列教材的建设旨在“汇聚学科精英、引领学科建设、培育专业英才”，同时以教材示范各重点大学的优秀教学理念、教学方法、教学手段和教学内容等。

本系列教材在规划过程中体现了如下一些基本组织原则和特点。

(1) 面向学科发展的前沿，适应当前社会对计算机专业高级人才的培养需求。教材内容以基本理论为基础，反映基本理论和原理的综合应用，重视实践和应用环节。

(2) 反映教学需要，促进教学发展。教材要能适应多样化的教学需要，正确把握教学内容和课程体系的改革方向。在选择教材内容和编写体系时注意体现素质教育、创新能力与实践能力的培养，为学生知识、能力、素质协调发展创造条件。

(3) 实施精品战略，突出重点，保证质量。规划教材建设的重点依然是专业基础课和专业主干课；特别注意选择并安排了一部分原来基础比较好的优秀教材或讲义修订再版，逐步形成精品教材；提倡并鼓励编写体现重点大学

计算机专业教学内容和课程体系改革成果的教材。

(4) 主张一纲多本,合理配套。专业基础课和专业主干课教材要配套,同一门课程可以有多本具有不同内容特点的教材。处理好教材统一性与多样化的关系;基本教材与辅助教材以及教学参考书的关系;文字教材与软件教材的关系,实现教材系列资源配套。

(5) 依靠专家,择优落实。在制订教材规划时要依靠各课程专家在调查研究本课程教材建设现状的基础上提出规划选题。在落实主编人选时,要引入竞争机制,通过申报、评审确定主编。书稿完成后要认真实行审稿程序,确保出书质量。

繁荣教材出版事业,提高教材质量的关键是教师。建立一支高水平的以老带新的教材编写队伍才能保证教材的编写质量,希望有志于教材建设的教师能够加入到我们的编写队伍中来。

教材编委会

FOREWORD

前言

本教材是在高仲仪、金茂忠老师 1991 年出版的《编译原理及编译程序构造》的基础上修改补充编纂而成的，融入了我们课程组十多年的教学经验。在教材的整体结构上保持了原书的结构，做了部分调整，对原书中的部分内容进行了修订，算法和示例程序改为 C 语言风格，删除了一些已不常用的语言，补充了一些新的内容，增加了编译优化部分的内容。由于本书仍然定位为本科生的第一门编译课程的教材，因此对于新出现的并行编译、较难的一些优化技术，本书没有涉及。

考虑到北京航空航天大学作为工科类院校的特点，我们的编译课程一直采用了“编译技术”这个名字，本书也将继续采用原书的书名《编译原理及编译程序构造》，意在强调编译系统的构造及其相关技术，突出对工程师人才的培养要求。

我们认为编译技术是计算机专业的一门非常重要的课程。虽然本书重在介绍编译系统的构造和相关原理与技术，但是在这个过程中，我们介绍并讨论了计算机领域三个非常重要的原理、概念和技术：

(1) 高级程序设计语言的工作原理；

(2) 程序/模型间转换的方法：实现模型从一种语言表达形式到另一种语言表达形式的(语义)等价转换方法及其相关技术；

(3) 软件系统的概念。对于大多数本科生而言，编译系统是他们构造的第一个相对完整的软件系统，因此我们非常强调系统的完整性。

我从 1987 年学编译，到 1998 年开始讲编译，这么多年来我不断地听到大家说编译是公认的难学亦难教的一门课，理论的东西太多，同时实践性很强。经过十多年的教学，我感觉编译应该不那么难学，因为编译的主线非常清晰。可为什么大家觉得难学呢？因为我们开始讲了过多的理论，从文法基础，到词法分析部分的自动机，再到各种语法分析方法。同学们被灌输了很多的算法，反而忽略了编译的根本任务：翻译，即程序从一种语言形式到另一种语言形式的等价翻译。这就是为什么学生往往在完成课程设计(一个小型编译系统)之后才恍然大悟：我有点明白编译是干什么的了。

显然，这是由于在教学中我们有些舍本求末了。编译技术作为一门重要

的专业基础课程，它讲述的是计算机领域中的一个重要的主题：即程序从一种语言形式到另一种语言形式的等价转换。编译课程通过从高级语言程序到低级语言程序的转换向我们介绍了相关的理论、方法和技术，并进一步阐明了高级语言的工作原理。因此，我们有责任给学生一个完整的编译过程。在这个过程的基础上或者这个过程中介绍相关的理论和方法。

本书的编写仍然沿用了传统的组织形式，以保证知识体系的完整。但是，在课程安排上，可以采用多种方式组织教学，比如：

（1）首先给学生一个完整的最简单的编译过程（翻译过程），见图 1 中 A 和 B 列对应的知识点，在此基础上再根据培养要求选讲代码优化、词法分析器的自动生成技术和其他语法分析技术（都可以自动生成），对应于图 1 中 C 列。我们经过三年的尝试，发现教学效果较好。在讲完 A 和 B 列后，学生可以看懂附录中的编译程序，并可以自己设计一个小型编译器。

（2）按照本书的章节顺序讲解，但是建议教师能随时回到编译过程，让学生理解当前讲解的内容处于编译过程的何处。

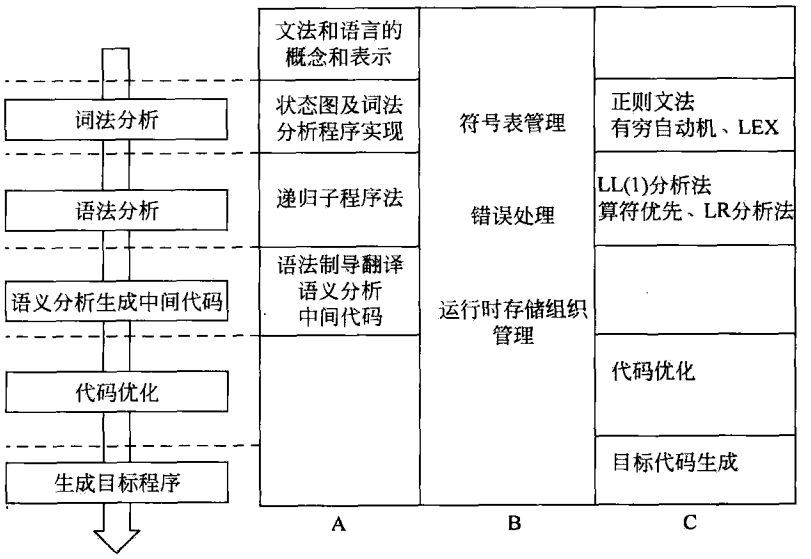


图 1　编译过程和知识点组织

因此，我们建议采用本书的老师可根据各学校的培养定位选择讲课的顺序和内容。

本书的完成凝聚着北航编译课程组全体老师的心血，几代师生共同完成了本书的撰写。金茂忠负责第 1、4、13 章，高仲仪、张莉负责第 2、3 章，张莉负责第 9、10 章，杨海燕负责第 5、6、7、8、14 章，史晓华负责第 11、12、15 章。张莉负责对全书进行了审校。还要感谢邱翔、田亚伟、申菊芳等在本书的习题、排版、校对等方面给予的帮助。

在本书编写过程中，还参考和引用了国内外优秀编译教材和著作中的相关内容，在此谨向原书著（译）者深表敬意和感谢。本书难免还有不少疏漏和差错之处，欢迎读者批评指正，以便在本书再版时进行修正。

联系方式 compile@ buaa.edu.cn

课程网站：http://www.sei.buaa.edu.cn/compile/（国家精品课程网站）

张　莉

2011 年 5 月

CONTENTS

目录

CHAPTER 1

绪　论　第1章

把高级程序设计语言翻译成汇编语言或机器语言的工作称为编译，完成这项翻译工作的软件系统称为编译程序或编译器。本书介绍编译原理及编译器的设计和实现技术。同时，编译原理和技术也是程序设计语言的工作原理和设计高质量程序的基础。

编译程序设计原理和构造技术具有十分普遍的意义，是计算机科学工作者、软件工程技术人员所必须掌握的专业基础知识，可以说每一个计算机科学工作者在他一生的研究生涯中，都会反复用到与编译有关的原理和技术。编译程序的设计涉及程序设计语言、计算机体系结构、形式语言理论、算法、数据结构和软件工程等学科。本章通过描述程序设计语言、编译过程、编译程序的组成以及编译程序的工作环境来简要介绍本书的主要内容。

1.1　什么是程序设计语言

语言是思维的表现形式，人与人之间通过自然语言来传达彼此的思想，因此，语言是人与人之间用来传递信息的媒介和手段。在计算机的应用领域，程序设计语言（也称为“程序语言”或“编程语言”）充当了人与问题以及协助解决该问题的计算机之间的通信工具，简而言之，人与计算机之间的中介就是程序设计语言。一种有效的程序设计语言能提高计算机程序的开发效率和对问题的表达能力，它必须能填补人们通常的思维方式与计算机执行所要求的精确性之间的沟壑。程序设计语言正是沿着为更好地满足这方面要求的方向而向前发展的。

一个好的程序设计语言应该容易表达人想要计算机所做的事情和便于进行程序的开发，即图 1.1 中的间隔（1）要小。如果沿着使该间隔减小的方向来看一下程序设计语言的发展道路，就是从机器语言、汇编语言等与机器相关的低级语言向通用编程语言（面向过程语言）、面向问题语言、逻辑编程语言、面向对象语言等与机器无关的高级语言发展的过程。

但是，越是沿着这条道路前进，图 1.1 中的间隔（2）就会越来越大，而用

来填补该间隔的就是程序设计语言的处理系统。

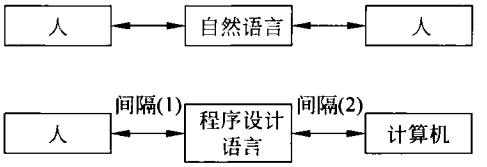


图 1.1　程序设计语言的位置

1.1.1　程序设计语言的定义方法

程序设计语言是一种人工语言。各种程序设计语言是有严格定义的(即该语言的规格说明)。语言的定义由语法和语义来规定。

例如,考虑以下简单的赋值语句:

```
s := (up + low) * h/2                                    (1-1)
```

赋值语句的形式为:

```
标识符 := 表达式
```

而表达式的形式为:

```
表达式 * 表达式
表达式 / 表达式
表达式 + 表达式
(表达式)
⋮
```

我们把对上述语言结构形式的描述称为文法(grammar)或语法(syntax)。

下面再来看一下式(1-1)的含义。它表示变量 up 和变量 low 相加,再和变量 h 相乘,然后除以 2,最后把结果赋给变量 s。如果这里 up、low、h、s 是实型变量,则 +、*、/ 就是实型数的加法、乘法和除法,式(1-1)中的 2 应该先转换成实数 2.0 以后再进行实型数的除法运算。我们把上述语言所表达内容的含义称为语义。

如何表示语言的语法和语义,将在第 2 章中详述。程序设计语言的定义由语言的规格说明书来描述。通常用巴科斯范式(BNF,Backus-Naur Form)来表示上下文无关文法语言的语法。

而表示程序设计语言的语义有很大困难。通常我们在程序设计语言规格说明书中用自然语言来描述语义。语义的形式化定义在形式语义学中研究,但这些定义难以理解。根据形式语义构造的语言分析系统效率较低且实现困难,至今尚未有成功的实用处理系统。本书将采用近年来广泛应用的属性文法来描述程序设计语言的语义。属性文法是一种半形式化的语义描述方法,它比较容易理解,而且可根据它构造出相当实用的分析系统。

1.1.2　程序设计语言的处理系统

程序设计语言的规格说明书给出了该语言的定义。当输入该语言的程序时,程序语言