

C 学习笔记

泰山学人夏翁

2020 年 5 月 18 日

目录

第一章 初识 C 语言	11
1.1 C 语言的起源	11
1.2 选择 C 语言的理由	12
1.2.1 设计特性	12
1.2.2 高效性	13
1.2.3 可移植性	13
1.2.4 强大而灵活	14
1.2.5 面向程序员	15
1.2.6 缺点	15
1.3 C 语言的应用范围	16
1.4 计算机能做什么	18
1.5 高级计算机语言和编译器	22
1.6 语言标准	23
1.6.1 第 1 个 ANSI/ISO C 标准	24
1.6.2 C99 标准	25
1.6.3 C11 标准维护标准任重道远。	25
1.7 使用 C 语言的 7 个步骤	26
1.7.1 第 1 步: 定义程序的目标	26
1.7.2 第 2 步: 设计程序	26
1.7.3 第 3 步: 编写代码	27
1.7.4 第 4 步: 编译	28

1.7.5	第 5 步: 运行程序	28
1.7.6	第 6 步: 测试和调试程序	29
1.7.7	第 7 步: 维护和修改代码 45	29
1.7.8	说明	29
1.8	编程机制	30
1.8.1	目标代码文件、可执行文件和库	31
1.8.2	UNIX 系统	33
1.8.3	GNU 编译器集合和 LLVM 项目	35
1.8.4	Linux 系统	36
1.8.5	PC 的命令行编译器	36
1.8.6	集成开发环境 (Windows)	37
1.8.7	Windows/Linux	39
1.8.8	Macintosh 中的 C	40
1.9	本书的组织结构	40
1.10	本书的约定	41
1.10.1	字体本书用类似在屏幕上或打印输	41
1.10.2	程序输出	41
1.10.3	特殊元素	43
1.11	本章小结	43
1.12	复习题	43
1.12.1	1. 对编程而言, 可移植性意味着什么?	44
1.12.2	2. 解释源代码文件、目标代码文件和可执行文件有什么区别?	44
1.12.3	3. 编程的 7 个主要步骤是什么?	44
1.12.4	4. 编译器的任务是什么?	44
1.12.5	5. 链接器的任务是什么?	44
1.13	编程练习我们尚未要求你编写 C 代码, 该练习侧重于编程过程的早期步骤。	44

目录	5
第二章 C 语言概述	45
2.1 简单的 C 程序示例	46
2.2 示例解释	47
2.2.1 第 1 遍：快速概要	49
2.2.2 第 2 遍：程序细节	50
2.3 简单程序的结构	61
2.4 提高程序可读性的技巧	62
2.5 进一步使用 C	63
2.5.1 程序说明	64
2.5.2 多条声明	64
2.5.3 乘法	65
2.5.4 打印多个值	65
2.6 多个函数	66
2.7 调试程序	68
2.7.1 语法错误	68
2.7.2 语义错误	69
2.8 关键字和保留标识符	71
2.9 关键概念	72
2.10 本章小结	73
2.11 复习题	74
2.11.1 C 语言的基本模块是什么？（函数）	74
2.11.2 什么是语法错误？写出一个英语例子和 C 语言例子。	74
2.11.3 写出一个英语例子和 C 语言例子。	74
2.11.4 Indiana Sloth 编写了下面的程序，并征求你的意见。 请帮助他评定。	75
2.11.5 假设下面 4 例子都是完整程序中的一部分，它们都输出什么结果？	75
2.11.6 在 main、int、function、char、= 中，哪些是 C 语言的关键字？	76

2.11.7	如何以下面的格式输出变量 words 和 lines 的值（这里，3020 和 350 代表两个变量的值）?	76
2.11.8	考虑下面的程序:	76
2.11.9	考虑下面的程序:	77
2.12	编程练习	77
2.12.1	1.	78
2.12.2	2	78
2.12.3	3	78
2.12.4	4	79
2.12.5	5	80
2.12.6	6	81
2.12.7	7	82
2.12.8	8	82
第三章	数据和 C	85
3.1	示例程序	85
3.1.1	程序中的新元素	87
3.2	变量与常量数据	88
3.2.1	常量 (constan)	88
3.2.2	变量 (variable)	89
3.3	数据: 数据类型关键字	89
3.3.1	位、字节和字	90
3.3.2	整数和浮点数	91
3.3.3	整数	91
3.3.4	浮点数	91
3.4	C 语言基本数据类型	92
3.4.1	int 类型	93
3.4.2	其他整数类型	98
3.4.3	使用字符: char 类型	104

3.4.4	<code>_Bool</code> 类型	110
3.4.5	可移植类型: <code>stdint.h</code> 和 <code>inttypes.h</code>	111
3.4.6	<code>float</code> 、 <code>double</code> 和 <code>long double</code>	113
3.4.7	得数和虚数类型	119
3.4.8	其他类型	119
3.4.9	类型大小	122
3.4.10	使用数据类型	123
3.4.11	参数和陷阱	124
3.4.12	转义序列示例	126
3.4.13	关键概念	129
3.4.14	本章小结	129
3.4.15	复习题	130
3.4.16	编程练习	133
第四章	总结与思考	135
4.1	数的类型	135
4.1.1	整形	135
4.1.2	浮点型	135
4.1.3	字符型	136
4.2	目标:	137

%#+OPTIONS: toc:nil

前言

1984 年 C Primer Plus 第一版问世。各版本累计销售超 55 万册。

C 语言从早期的非正式的 KR 标准，发展到 1990 年 ISO/ANSI 标准，进而发展到 2011 年 ISO/IEC 标准。本书第六版。

C primer Plus Sixty Edition

- 经久不衰的 C 语言畅销经典教程
- 针对 C11 标准进行全面更新

美 |Stephen Prate| 著姜佑 | 译

第一章 初识 C 语言

- C 的历史和特性
- 编写程序的步骤
- 编译器和链接器的知识
- C 标准

C 是一门功能强大、专业化的编程语言。

1.1 C 语言的起源

1972 年，贝尔实验室的丹尼斯·里奇 (Dennis Ritchie) 和肯·汤普逊 (Ken Thompson) 在开发 UNIX 操作系统时设计了 C 语言。

- 设计者：Dennis Ritchie Ken Thompson
- 时间 1972 年
- UNIX¹
- 贝尔实验室

¹Unix 是 20 世纪 70 年代初出现的一个操作系统，除了作为网络操作系统之外，还可以作为单机操作系统使用。Unix 作为一种开发平台和台式操作系统获得了广泛使用，目前主要用于工程应用和科学计算等领域。[1]

Pascal² 的主发目标是为更好地学习编程原理提供扎实的基础；而 BASIC³ 的主要目标是开发出类似英文的语言，让不熟悉计算机的人轻松学习。

1.2 选择 C 语言的理由

在过去 40 年里，C 成为最重要、最流行的语言之一。突出的优点

1.2.1 设计特性

C 是一门流行的语言，融合了计算机科学理论和实践的控制特性。C 语言的设计理念让用户能轻松地完成 * 自顶向下的规划、结构化编程和模块化设计。* 因此，用 C 语言编写的程序更 * 易懂、更可靠 *。

融合了

- 计算机科学理论
- 实践的控制特性

²Pascal 的名称是为了纪念十七世纪法国著名哲学家和数学家 Blaise Pascal 而来的，它由瑞士 Niklaus Wirth 教授于六十年代末设计并创立的。Pascal 语言语法严谨，层次分明，程序易写，可读性强，是第一个结构化编程语言。Pascal 语言广泛用于各种软件，程序分为名称（program 后自拟）、设置（var 后规定）、开始（begin）、程序（正文）、读取（read/read ln）、结束（end），结构层次强，严谨而又紧密。

³BASIC（来自英语：Beginner's All-purpose Symbolic Instruction Code 的缩写），又译培基，一种直译式程序设计语言。名称字面为意思“初学者的全方位符式指令代码”，设计给初学者使用的编程语言，在完成编写后不须经由编译及链接等手续，经过解释器即可运行，但如果需要单独运行时仍然需要将其创建成可执行文件。

BASIC 语言是由 Dartmouth 学院 John G. Kemeny 与 Thomas E. Kurtz 两位教授于 20 世纪 60 年代中期所创。由于立意甚佳，BASIC 语言简单、易学的基本特性，很快地就普遍流行起来，几乎所有小型、微型以家用电脑，甚至部分大型电脑，都有提供使用者以此种语言撰写程式。在微电脑方面，则因为 BASIC 语言可配合微电脑操作功能的充分发挥，使得 BASIC 早已成为微电脑的主要语言之一。

设计特性

- 自顶向下的规划
- 结构化编程
- 模块化设计

程序特点

- 更易懂
- 更可靠

例如 linux 内核用 C 语言实现，而不是 C++

1.2.2 高效性

C 是高效的语言。在设计上，它充分利用了当前计算机的优势，因此 C 程序相对/更紧凑/，而且/运行速度很快/。实际上，C 语言具有通常是汇编语言才具有的微调控制能力（汇编语言是为特殊的中央处理单元设计的一系列内部指令，使用助记符来表示；不同的 CPU 系列使用不同的汇编语言），可以根据具体情况微调程序以获得最大运行速度或最有效地使用内存。

- 更紧凑。矮小精悍（具有汇编语言的微调控制能）
- 运行速度快、高效（最优运行速度或最有效地使用内存）

1.2.3 可移植性

C 是可移植的语言。这意味着，在一种系统中编写的 C 程序稍作修改或不修改就能在其他系统运行。如需修改，也只需简单更改主程序头文件中的少许项即可。大部分语言都希望成为可移植语言，但是，如果经历过把

IBM PC BASIC 程序转换成苹果 BASIC(两者是近亲), 或者在 UNIX 系统中运行 IBM 大型机的 FORTRAN 程序的人都知道, 移植是最麻烦的事。C 语言是可移植方面的佼佼者。从 8 位微处理器到克雷超级计算机, 许多计算机体系结构都可以使用 C 编译器 (C 编译器是把 C 代码转换成计算机内部指令的程序)。但是要注意, 程序中针对特殊硬件设备 (如, 显示监视器) 或操作系统特殊功能 (如, Windows 8 或 OS X) 编写的部分, 通常是不可移植的。

由于 C 语言与 UNIX 关系密切, UNIX 系统通常会将 C 编译器作为软件包的一部分。安装 Linux 时, 通常也会安装 C 编译器。供个人计算机使用的 C 编译器很多, 运行各种版本的 Windows 和 Macintosh(即, Mac) 的 PC 都能找到合适的 C 编译器。因此, 无论是使用家庭计算机、专业工作站, 还是大型机, 都能找到针对特定系统的 C 编译器。

- 不可移植: 针对特殊硬件设备或操作系统的特殊功能编写的部分

1.2.4 强大而灵活

C 语言功能强大且灵活 (计算机领域经常使用这两个词)。例如, 功能强大且灵活的 UNIX 操作系统, 大部分是用 C 语言写的; 其他语言 (如, FORTRAN、Perl、Python、Pascal、LISP、Logo、BASIC) 的许多编译器和解释器都是用 C 语言编写的。因此, 在 UNIX 机上使用 FORTRAN 时, 最终是由 C 程序生成最后的可执行程序。C 程序可以用于解决物理学和工程学的问题, 甚至可用于制作电影的动画特效。

- 编写操作系统 (UNIX,linux)
- 编写语言的编译器和解释器 (python,lisp FORTRAN BASIC)
- 物理、工程、电影特效等。

1.2.5 面向程序员

C 语言是为了满足程序员的需求而设计的，程序员利用 C 可以访问硬件、操控内存中的位。C 语言有丰富的运算符，能让程序员简洁地表达自己的意图。C 没有 Pascal 严谨，但是却比 C++ 的限制多。这样的灵活性既是优点也是缺点。优点是，许多任务用 C 来处理都非常简洁 (如，转换数据的格式); 缺点是，你可能会犯一些莫名其妙的错误，这些错误不可能在其他语言中出现。C 语言在提供更多自由的同时，也让使用者承担了更大的责任。

另外，大多数 C 实现都有一个大型的库，包含众多有用的 C 函数。这些函数用于处理程序员经常需要解决的问题。

- 可访问硬件、操控内存中的位
- 有丰富的运算符，可简洁地表达意图
- 没有 pascal 严谨，但比 C++ 限制多，灵活。
- 比其它语言易犯错误
- C 语言用大型的库，以 C 函数为基础。方便使用。

1.2.6 缺点

人无完人，金无足赤。C 语言也有一些缺点。例如，前面提到的，要享受用 C 语言自由编程的乐趣，就必须承担更多的责任。特别是，C 语言使用指针，而涉及指针的编程错误往往难以察觉。有句话说的好: 想拥有自由就必须时刻保持警惕。

C 语言紧凑简洁，结合了大量的运算符。正因如此，我们也可以编写出让人极其费解的代码。虽然没必要强迫自己编写晦涩的代码，但是有兴趣写写也无妨。试问，除 C 语言外还为哪种语言举办过年度混乱代码大赛 [1]?

瑕不掩瑜，C 语言的优点比缺点多很多。我们不想在这里多费笔墨，还是来聊聊 C 语言的其他话题。

- 自由就易犯错
- 指针的编程错误难以觉察
- 紧凑、简洁，大量的运算符，代码极其费解。
- 代码任意排列，更难理解（混乱代码大赛）⁴（插入脚注命令 C-C

C-X f, 可修改脚注名称）

1.3 C 语言的应用范围

早在 20 世纪 80 年代，C 语言就已经成为小型计算机 (UNIX 系统) 使用的主流语言。从那以后，C 语言的应用范围扩展到微型机 (个人计算机) 和大型机 (庞然大物)。如图 1.2 所示，许多软件公司都用 C 语言来开发文字处理程序、电子表格、编译器和其他产品，因为用 C 语言编写的程序紧凑而高效。更重要的是，C 程序很方便修改，而且移植到新型号的计算机中也没什么问题。

⁴国际 C 语言混乱代码大赛 (IOCCC, The International Obfuscated C Code Contest) 是一项国际编程赛事，从 1984 年开始，每年举办一次 (1997 年、1999 年、2002 年、2003 年和 2006 年例外)。目的是写出最有创意的最让人难以理解的 C 语言代码。作品从联机提交，并经过好几回合的裁判的审核。作品审核是基于滥用混乱代码的程度 (以及创造性的滥用)。作品通过最后一轮审核的会被归成其特别的一类以示嘉奖，例如“最滥用 C 预处理器”或者“最古怪的行为”，并且发表在官方 IOCCC 网站。作品没有通过者会被删除或遗忘；该赛事主办声明：作品被声明在 IOCCC 网站是竞赛的锦标。IOCCC 是由蓝登·克特·诺尔 (Landon Curt Noll) 与拉里·贝索 (Larry Bassel) 在 1984 年受雇于国家半导体 (National Semiconductor) 的 Genix 程序移植事业群时开始的。比赛的点子是来自他们俩比较彼此有关于它们得修正的某些写得很烂代码的笔记。

获奖作品获奖作品规则每年，IOCCC 的比赛规则会张贴在其网站上。规则每年不同，并且会附上一组的指导方针以试图表达规则的精神。

这些规则通常是蓄意书写成文，伴随着精巧的漏洞让参赛者有所鼓励去发现并滥用。比赛结果就是“软件开发过程的讽刺体”。作品占某些规则里头的漏洞之便者 (不管它是否通过最后一轮审核) 会造成下年度比赛规则的调整 (虽然常常其他微妙的漏洞会又被裁判存心放水)。

无论是软件公司、经验丰富的 C 程序员，还是其他用户，都能从 C 语言中受益。越来越多的计算机用户已转而求助 C 语言解决一些安全问题。不一定非得是计算机专家也能使用 C 语言。

20 世纪 90 年代，许多软件公司开始改用 C++ 来开发大型的编程项目。C++ 在 C 语言的基础上嫁接了面向对象编程工具（面向对象编程是一门哲学，它通过对语言建模来适应问题，而不是对问题建模以适应语言）。C++ 几乎是 C 的超集，这意味着任何 C 程序差不多就是一个 C++ 程序。学习 C 语言，也相当于学习了许多 C++ 的知识。

虽然这些年来 C++ 和 JAVA 非常流行，但是 C 语言仍是软件业中的核心技能。在最想具备的技能中，C 语言通常位居前十。特别是，C 语言已成为嵌入式系统编程的流行语言。也就是说，越来越多的汽车、照相机、DVD 播放机和其他现代化设备的微处理器都用 C 语言进行编程。除此之外，C 语言还从长期被 FORTRAN 独占的科学编程领域分得一杯羹。最终，作为开发操作系统的卓越语言，C 在 Linux 开发中扮演着极其重要的角色。因此，在进入 21 世纪的第 2 个 10 年中，C 语言仍然保持着强劲的势头。

简而言之，C 语言是最重要的编程语言之一，将来也是如此。如果你想拿下一份编程的工作，被问到是否会 C 语言时，最好回答“是”。

- 紧凑、高效、方便移植
- 非专业也可以用
- C++ 是 C 的超集
- []C++ 在 C 的基础上嫁接了面向对象编程工具

☐ 面向对象编程是一门哲学，通过对语言建模来适应问题而非对问题建模以适应语言

- C 语言仍是软件业中的核心技能（嵌入式系统编程的流行语言）

1.4 计算机能做什么

在学习如何用 C 语言编程之前，最好先了解一下计算机的工作原理。这些知识有助于你理解用 C 语言编写程序和运行 C 程序时所发生的事情之间有什么联系。

现代的计算机由多种部件构成。中央处理单元 (CPU⁵) 承担绝大部分

⁵CPU 中央处理单元 (控制与运算)

Central⁶processing unit

计算机系统的运算和控制中心, 信息处理和程序运行的最终执行单元其作用⁷构成由⁸, 发展历史有六个阶段⁹, 工作原理以冯诺依曼体系结构是现代计算机的基础。在该体系结构下, 程序和数据统一存储, 指令和数据需要从同一存储空间存取, 经由同一总线传输, 无法重叠执行。根据冯诺依曼体系, CPU 的工作分为以下 5 个阶段: 取指令阶段¹⁰、指令译码阶段¹¹、执行指令阶段¹²、访存取数¹³。[1] 性能指标有¹⁴, CPU 结构可以大致分为运算逻辑部件¹⁵、寄存器部件。而计算机的内存分为: 随机存取存储器 (RAM) (ROM)¹⁶。

⁶fetch 英 [fet] 美 [fet] v. (去) 拿来; (去) 请来; 售得, 卖得 (某价); [例句] Sylvia fetched a towel from the bathroom 西尔维娅去卫生间拿了一条毛巾。[其他] 第三人称单数: fetches 现在分词: fetching 过去式: fetched 过去分词: fetched

⁷- 解释与执行指令

- 处理计算机软件中的数据
- 读取指令、指令译码、执行指令

⁸- 控制器

- 运算器
- 高速缓存

⁹- 专用计算》通用计算

- 4 位-8 位-16 位-32 位-64 位 1. 第一阶段 (71-73 年)

4 位及 8 位 intel 4004

1971 年, Intel 生产的 4004 微处理器将运算器和控制器集成在一个芯片上, 标志着 CPU 的诞生;

2. 第二阶段 (74-77 年)

高档 8 位处理器, 代表产品是 Intel 8080。此时指令系统已经比较完善了 3. 第三阶段 (78-84 年)

16 位微处理器的时代, 代表产品是 Intel 8086。相对而言已经比较成熟了。4. 第四阶段 (85-92 年)

32 位微处理器时代, 代表产品是 Intel 80386。已经可以胜任多任务、多用户的作业。1989 年发布的 80486 处理器实现了 5 级标量流水线, 标志着 CPU 的初步成熟, 也标志着传统处理器发展阶段的结束。

5. 第五阶段 (1993-2005 年)

奔腾系列微处理器的时代

1995 年 11 月, Intel 发布了 Pentium 处理器, 该处理器首次采用超标量指令流水结构, 引入了指令的乱序执行和分支预测技术, 大大提高了处理器的性能, 因此, 超标量指令流水线结构一直被

后续出现的现代处理器，如 AMD (Advanced Micro devices) 的 K9、K10、Intel 的 Core 系列等所采用。

第 6 阶段 0 (2005 年至今)。

是酷睿系列微处理器的时代，这是一款领先节能的新型微架构，设计的出发点是提供卓然出众的性能和能效。[3]

为了满足操作系统的上层工作需求，现代处理器进一步引入了诸如

- 并行化、
- 多核化、
- 虚拟化
- 以及远程管理系统等功能，不断推动着上层信息系统向前发展。[1]

¹⁰1. 取指令

(IF, instruction¹⁷fetch⁶), 即将一条指令从主存储器中取到指令寄存器的过程。程序计数器中的数值, 用来指示当前指令在主存中的位置。当一条指令被取出后, PC 中的数值将根据指令字长度自动递增。[1]

¹¹instruction 英 [nstrkn] 美 [nstrkn] n. 用法说明; 操作指南; 指示; 命令; 吩咐; (计算机的) 指令; adj. 说明用法的; 操作指南的; [例句]Each candidate is given instruction in safety 向每位申请人都讲授了安全知识。[其他] 复数: instructions

¹²指令译码阶段

(ID, instruction decode¹⁸), 取出指令后, 指令译码器按照预定的指令格式, 对取回的指令进行拆分和解释, 识别区分出不同的指令类别以及各种获取操作数的方法。[1]

¹³decode 英 [dikd] 美 [dikod] v. 解 (码); 破译 (尤指密码); 译解 (电子信号); 译解, 理解 (外文); [例句]All he had to do was decode it and pass it over 他需要做的就是将它破译然后转给他人。[其他] 第三人称单数: decodes 现在分词: decoding 过去式: decoded 过去分词: decoded

¹⁴3. 执行指令阶段 (EX, execute¹⁹),

具体实现指令的功能。CPU 的不同部分被连接起来, 以执行所需的操作。

¹⁵execute 英 [ekskjut] 美 [ekskjut] v. (尤指依法) 处决, 处死; 实行; 执行; 实施; 成功地完成 (技巧或动作); [例句]He was executed by lethal injection earlier today 他于今天早些时候被注射处死。[其他] 第三人称单数: executes 现在分词: executing 过去式: executed 过去分词: executed

¹⁶4. 访存取数阶段 (MEM, memory),

根据指令需要访问主存、读取操作数, CPU 得到操作数在主存中的地址, 并从主存中读取该操作数用于运算。部分指令不需要访问主存, 则可以跳过该阶段。[1]

¹⁷5. 结果写回阶段 (WB, write back),

作为最后一个阶段, 结果写回阶段把执行指令阶段的运行结果数据“写回”到某种存储形式。结果数据一般会被写到 CPU 的内部寄存器中, 以便被后续的指令快速地存取; 许多指令还会改变程序状态字寄存器中标志位的状态, 这些标志位标识着不同的操作结果, 可被用来影响程序的动作。[1]

的运算工作。随机存取内存 (RAM) 是存储程序和文件的工作区; 而永久内存存储设备 (过去一般指机械硬盘, 现在还包括固态硬盘) 即使在关闭计算机后, 也不会丢失之前储存的程序和文件。另外, 还有各种外围设备 (如, 键盘、鼠标、触摸屏、监视器) 提供人与计算机之间的交互。CPU 负责处理程序, 接下来我们重点讨论它的工作原理。CPU 的工作非常简单, 至少从以下简短的描述中看是这样。它从内存中获取并执行一条指令, 然后再从内存中获取并执行下一条指令, 诸如此类 (一个吉赫兹的 CPU 一秒钟能重复这样的操作大约十亿次, 因此, CPU 能以惊人的速度从事枯燥的工作)。CPU 有自己的小工作区——由若干个寄存器组成, 每个寄存器都可以储存一个数字。一个寄存器储存下一条指令的内存地址, CPU 使用该地址来获取和更新下一条指令。在获取指令后, CPU 在另一个寄存器中储存该指令, 并更新第 1 个寄存器储存下一条指令的地址。CPU 能理解的指令有限 (这些指令的集合叫作指令集)。而且, 这些指令相当具体, 其中的许多指令都

在指令执行完毕、结果数据写回之后, 若无意外事件 (如结果溢出等) 发生, 计算机就从程序计数器中取得下一条指令地址, 开始新一轮的循环, 下一个指令周期将顺序取出下一条指令。[1]

¹⁸对于 CPU 而言, 影响其性能的指标主要有主频、CPU 的位数以及 CPU 的缓存指令集。所谓 CPU 的主频, 指的就是时钟频率, 它直接决定了 CPU 的性能, 因此要想 CPU 的性能得到很好地提高, 提高 CPU 的主频是一个很好地途径。而 CPU 的位数指的就是处理器能够一次性计算的浮点数的位数, 通常情况下, CPU 的位数越高, CPU 进行运算时候的速度就会变得越快。现在 CPU 的位数一般为 32 位或者 64 位。以前人们使用的计算机都是 32 位系统, 近年来人们使用的计算机的处理器中 64 位所占用的比例则显得更多, 这是因为 64 位的计算机的运行速度变得更快, 提高了人们的工作效率。而 CPU 的缓存指令集是存储在 CPU 内部的, 主要指的是能够对 CPU 的运算进行指导以及优化的硬程序。一般来讲, CPU 的缓存可以分为一级缓存、二级缓存和三级缓存, 而那些处理能力比较强的处理器则一般具有较大的三级缓存。[4]

¹⁹所谓运算逻辑部件, 主要能够进行相关的逻辑运算, 如: 可以执行移位操作以及逻辑操作, 除此之外还可以执行定点或浮点算术运算操作以及地址运算和转换等命令, 是一种多

²⁰而寄存器部件则是用来暂存指令、数据和地址的。

²¹控制部件则是主要用来对指令进行分析并且能够发出相应的控制信号。

²²随机存取存储器能够与 CPU 直接的进行数据的交换, 也可以将其称为主存。对于 RAM 可以随时的进行读写, 而且这个过程的速度很快, 因此由于主存所具有的这个优点也往往将其作为操作系统或其他正在运行中的程序的临时数据存储媒介

²³而只读存储器 ROM 是一种只能读出事先所存数据的存储器, 使用者对于其内部存储的资料没有改变的权限也无法对其进行删除, 并且在电源关闭以后资料并不会消失。这种内存也得到了广泛的应用, 在那些资料不需要经常变更的电子或电脑系统中得到了很好地应用。

是用于请求计算机把一个数字从一个位置移动到另一个位置。例如，从内存移动到寄存器。

下面介绍两个有趣的知识。其一，储存在计算机中的所有内容都是数字。计算机以数字形式储存数字和字符 (如，在文本文档中使用的字母)。每个字符都有一个数字码。计算机载入寄存器的指令也以数字形式储存，指令集中的每条指令都有一个数字码。其二，计算机程序最终必须以数字指令码 (即，机器语言) 来表示。

简而言之，计算机的工作原理是：如果希望计算机做某些事，就必须为其提供特殊的指令列表 (程序)，确切地告诉计算机要做的事以及如何做。

你必须用计算机能直接明白的语言 (机器语言) 创建程序。这是一项繁琐、乏味、费力的任务。计算机要完成诸如两数相加这样简单的事，就得分成类似以下几个步骤。

1. 从内存位置 2000 上把一个数字拷贝到寄存器 1。
2. 从内存位置 2004 上把另一个数字拷贝到寄存器 2。
3. 把寄存器 2 中的内容与寄存器 1 中的内容相加，把结果储存在寄存器 1 中。
4. 把寄存器 1 中的内容拷贝到内存位置 2008。而你要做的是，必须用数字码来表示以上的每个步骤！

如果以这种方式编写程序很合你的意，那不得不说抱歉，因为用机器语言编程的黄金时代已一去不复返。但是，如果你对有趣的事情比较感兴趣，不妨试试高级编程语言。

1.5 高级计算机语言和编译器

高级编程语言 (如，C) 以多种方式简化了编程工作。首先，不必用数字码表示指令；其次，使用的指令更贴近你如何想这个问题，而不是类似计算机那样繁琐的步骤。使用高级编程语言，可以在更抽象的层面表达你的想法，不用考虑 CPU 在完成任务时具体需要哪些步骤。例如，对于两数相加，可以这样写：

```
total = mine + yours;
```

对我们而言，光看这行代码就知道要计算机做什么；而看用机器语言写

成的等价指令 (多条以数字码形式表现的指令) 则费劲得多。但是, 对计算机而言却恰恰相反。在计算机看来, 高级指令就是一堆无法理解的无用据。编译器在这里派上了用场。编译器是把高级语言程序翻译成计算机能理解的机器语言指令集的程序。程序员进行高级思维活动, 而编译器则负责处理冗长乏味的细节工作。

编译器还有一个优势。一般而言, 不同 CPU 制造商使用的指令系统和编码格式不同。例如, 用 Intel Core i7 (英特尔酷睿 i7) CPU 编写的机器语言程序对于 ARM Cortex-A57 CPU 而言什么都不是。但是, 可以找到与特定类型 CPU 匹配的编译器。因此, 使用合适的编译器或编译器集, 便可把一种高级语言程序转换成供各种不同类型 CPU 使用的机器语言程序。一旦解决了一个编程问题, 便可让编译器集翻译成不同 CPU 使用的机器语言。

简而言之, 高级语言 (如 C、Java、Pascal) 以更抽象的方式描述行为, 不受限于特定 CPU 或指令集。而且, 高级语言简单易学, 用高级语言编程比用机器语言编程容易得多。

1964 年, 控制数据公司 (Control Data Corporation) 研制出了 CDC 6600 计算机。这台庞然大物是世界上首台超级计算机, 当时的售价是 600 万美元。它是高能核物理研究的首选。然而, 现在的普通智能手机在计算能力和内存方面都超过它数百倍, 而且能看视频, 放音乐。

1964 年, 在工程和科学领域的主流编程语言是 FORTRAN。虽然编程语言不如硬件发展那么突飞猛进, 但是也发生了很大变化。为了应对越来越大型的编程项目, 语言先后为结构化编程和面向对象编程提供了更多的支持。随着时间的推移, 不仅新语言层出不穷, 而且现有语言也会发生变化。

1.6 语言标准

目前, 有许多 C 实现可用。在理想情况下, 编写 C 程序时, 假设该程序中未使用机器特定的编程技术, 那么它的运行情况在任何实现中都应该

相同。要在实践中做到这一点，不同的实现要遵循同一个标准。

C 语言发展之初，并没有所谓的 C 标准。1987 年，布莱恩·柯林汉 (Brian Kernighan) 和丹尼斯·里奇 (Dennis Ritchie) 合著的 *The C Programming Language* (《C 语言程序设计》) 第 1 版是公认的 C 标准，通常称之为 K&R C 或经典 C。特别是，该书中的附录中的“C 语言参考手册”已成为实现 C 的指导标准。例如，编译器都声称提供完整的 K&R 实现。虽然这本书中的附录定义了 C 语言，但却没有定义 C 库。与大多数语言不同的是，C 语言比其他语言更依赖库，因此需要一个标准库。实际上，由于缺乏官方标准，UNIX 实现提供的库已成为了标准库。

1.6.1 第 1 个 ANSI/ISO C 标准

随着 C 的不断发展，越来越广泛地应用于更多系统中，C 社区意识到需要一个更全面、更新颖、更严格的标准。鉴于此，美国国家标准协会 (ANSI) 于 1983 年组建了一个委员会 (X3J11)，开发了一套新标准，并于 1989 年正式公布。该标准 (ANSI C) 定义了 C 语言和 C 标准库。国际标准化组织于 1990 年采用了这套 C 标准 (ISO C)。ISO C 和 ANSI C 是完全相同的标准。ANSI/ISO 标准的最终版本通常叫作 C89(因为 ANSI 于 1989 年批准该标准) 或 C90(因为 ISO 于 1990 年批准该标准)。另外，由于 ANSI 先公布 C 标准，因此业界人士通常使用 ANSI C。

在该委员会制定的指导原则中，最有趣的可能是：保持 C 的精神。委员会在表述这一精神时列出了以下几点：

- 信任程序员；

- 不要妨碍程序员做需要做的事；

- 保持语言精练简单；只提供一种方法执行一项操作；让程序运行更快，即使不能保证其可移植性。

在最后一点上，标准委员会的用意是：作为实现，应该针对目标计算机来定义最合适的某特定操作，而不是强加一个抽象、统一的定义。在学习 C 语言过程中，许多方面都反映了这一哲学思想。

1.6.2 C99 标准

1994 年，ANSI/ISO 联合委员会 (C9X 委员会) 开始修订 C 标准，最终发布了 C99 标准。该委员会遵循了最初 C90 标准的原则，包括保持语言的精练简单。委员会的用意不是在 C 语言中添加新特性，而是为了达到新的目标。第 1 个目标是，支持国际化编程。例如，提供多种方法处理国际字符集。第 2 个目标是，“调整现有实践致力于解决明显的缺陷”。因此，在遇到需要将 C 移至 64 位处理器时，委员会根据现实生活中处理问题的经验来添加标准。第 3 个目标是，为适应科学和工程项目中的关键数值计算，提高 C 的适应性，让 C 比 FORTRAN 更有竞争力。

这 3 点 (国际化、弥补缺陷和提高计算的实用性) 是主要的修订目标。在其他方面的改变则更为保守，例如，尽量与 C90、C++ 兼容，让语言在概念上保持简单。用委员会的话说：“委员会很满意让 C++ 成为大型、功能强大的语言”。

C99 的修订保留了 C 语言的精髓，C 仍是一门简洁高效的语言。本书指出了许多 C99 修改的地方。虽然该标准已发布了很长时间，但并非所有的编译器都完全实现 C99 的所有改动。因此，你可能发现 C99 的一些改动在自己的系统中不可用，或者只有改变编译器的设置才可用。

1.6.3 C11 标准维护标准任重道远。

标准委员会在 2007 年承诺 C 标准的下一个版本是 C1X，2011 年终于发布了 C11 标准。此次，委员会提出了一些新的指导原则。出于对当前编程安全的担忧，不那么强调“信任程序员”目标了。而且，供应商并未像对 C90 那样很好地接受和支持 C99。这使得 C99 的一些特性成为 C11 的可选项。因为委员会认为，不应要求服务小型机市场的供应商支持其目标环境中用不到的特性。另外需要强调的是，修订标准的原因不是因为原标准不能用，而是需要跟进新的技术。例如，新标准添加了可选项支持当前使用多处理器的计算机。对于 C11 标准，我们浅尝辄止，深入分析这部分内容已超出本书讨论的范围。

注意

本书使用术语 ANSI C、ISO C 或 ANSI/ISO C 讲解 C89/90 和较新标准共有的特性，用 C99 或 C11 介绍新的特性。有时也使用 C90(例如，讨论一个特性被首次加入 C 语言时)。

1.7 使用 C 语言的 7 个步骤

C 是编译型语言。如果之前使用过编译型语言 (如, Pascal 或 FORTRAN), 就会很熟悉组建 C 程序的几个基本步骤。但是, 如果以前使用的是解释型语言 (如, BASIC) 或面向图形界面语言 (如, Visual Basic), 或者甚至没接触过任何编程语言, 就有必要学习如何编译。别担心, 这并不复杂。首先, 为了让读者对编程有大概的了解, 我们把编写 C 程序的过程分解成 7 个步骤 (见图 1.3)。注意, 这是理想状态。在实际的使用过程中, 尤其是在较大型的项目中, 可能要做一些重复的工作, 根据下一个步骤的情况来调整或改进上一个步骤。

图 1.3 编程的 7 个步骤

1.7.1 第 1 步: 定义程序的目标

在动手写程序之前, 要在脑中有清晰的思路。想要程序去做什么首先要明确自己想做什么, 思考你的程序需要哪些信息, 要进行哪些计算和控制, 以及程序应该要报告什么信息。在这一步骤中, 不涉及具体的计算机语言, 应该用一般术语来描述问题。

1.7.2 第 2 步: 设计程序

对程序应该完成什么任务有概念性的认识后, 就应该考虑如何用程序来完成它。例如, 用户界面应该是怎样的? 如何组织程序? 目标用户是谁? 准备花多长时间来完成这个程序?

除此之外，还要决定在程序 (还可能是辅助文件) 中如何表示数据，以及用什么方法处理数据。学习 C 语言之初，遇到的问题都很简单，没什么可选的。但是，随着要处理的情况越来越复杂，需要决策和考虑的方面也越来越多。通常，选择一个合适的方式表示信息可以更容易地设计程序和处理数据。

再次强调，应该用一般术语来描述问题，而不是用具体的代码。但是，你的某些决策可能取决于语言的特性。例如，在数据表示方面，C 的程序员就比 Pascal 的程序员有更多选择。

1.7.3 第 3 步: 编写代码

设计好程序后，就可以编写代码来实现它。也就是说，把你设计的程序翻译成 C 语言。这里是真正需要使用 C 语言的地方。可以把思路写在纸上，但是最终还是要将代码输入计算机。这个过程机制取决于编程环境，我们稍后会详细介绍一些常见的环境。一般而言，使用文本编辑器创建源代码文件。该文件中内容就是你翻译的 C 语言代码。程序清单 1.1 是一个 C 源代码的示例。

程序清单 1.1 C 源代码示例

```
#include <stdio.h>

int main(void)
{
    int dogs;
    printf("How many dogs do you have?\n");
    scanf("%d", &dogs);
    printf("So you have %d dog(s)!\n", dogs);
    return 0;
}
```

在这一步骤中，应该给自己编写的程序添加文字注释。最简单的方式是使用 C 的注释工具在源代码中加入对代码的解释。第 2 章将详细介绍如何在代码中添加注释。

1.7.4 第 4 步：编译

接下来的这一步是编译源代码。再次提醒读者注意，编译的细节取决于编程的环境，我们稍后马上介绍一些常见的编程环境。现在，先从概念的角度讲解编译发生了什么事情。

前面介绍过，编译器是把源代码转换成可执行代码的程序。可执行代码是用计算机的机器语言表示的代码。这种语言由数字码表示的指令组成。如前所述，不同的计算机使用不同的机器语言方案。C 编译器负责把 C 代码翻译成特定的机器语言。此外，C 编译器还将源代码与 C 库（库中包含大量的标准函数供用户使用，如 `printf()` 和 `scanf()`）的代码合并成最终的程序（更精确地说，应该是由一个被称为链接器的程序来链接库函数，但是在大多数系统中，编译器运行链接器）。其结果是，生成一个用户可以运行的可执行文件，其中包含着计算机能理解的代码。

编译器还会检查 C 语言程序是否有效。如果 C 编译器发现错误，就不生成可执行文件并报错。理解特定编译器报告的错误或警告信息是程序员要掌握的另一项技能。

1.7.5 第 5 步：运行程序

传统上，可执行文件是可运行的程序。在常见环境（包括 Windows 命令提示符模式、UNIX 终端模式和 Linux 终端模式）中运行程序要输入可执行文件的文件名，而其他环境可能要运行命令（如，在 VAX 中的 `VMS[2]`）或一些其他机制。例如，在 Windows 和 Macintosh 提供的集成开发环境（IDE）中，用户可以在 IDE 中通过选择菜单中的选项或按下特殊键来编辑和执行 C 程序。最终生成的程序可通过单击或双击文件名或图标直接在操作系统中运行。

1.7.6 第 6 步: 测试和调试程序

程序能运行是个好迹象，但有时也可能会出现运行错误。接下来，应该检查程序是否按照你所设计的思路运行。你会发现你的程序中有一些错误，计算机行话叫作 bug。查找并修复程序错误的过程叫调试。学习的过程中不可避免会犯错，学习编程也是如此。因此，当你把所学的知识应用于编程时，最好为自己会犯错做好心理准备。随着你越来越老练，你所写的程序中的错误也会越来越不易察觉。

将来犯错的机会很多。你可能会犯基本的设计错误，可能错误地实现了一个好想法，可能忽视了输入检查导致程序瘫痪，可能会把圆括号放错地方，可能误用 C 语言或打错字，等等。把你将来犯错的地方列出来，这份错误列表应该会很长。

看到这里你可能会有些绝望，但是情况没那么糟。现在的编译器会捕获许多错误，而且自己也可以找到编译器未发现的错误。在学习本书的过程中，我们会给读者提供一些调试的建议。

1.7.7 第 7 步: 维护和修改代码 45

创建完程序后，你发现程序有错，或者想扩展程序的用途，这时就要修改程序。例如，用户输入以 Zz 开头的姓名时程序出现错误、你想到了一个更好的解决方案、想添加一个更好的新特性，或者要修改程序使其能在不同的计算机系统中运行，等等。如果在编写程序时清楚地做了注释并采用了合理的设计方案，这些事情都很简单。

1.7.8 说明

编程并非像描述那样是一个线性的过程。有时，要在不同的步骤之间往复。例如，在写代码时发现之前的设计不切实际，或者想到了一个更好的解决方案，或者等程序运行后，想改变原来的设计思路。对程序做文字注释为今后的修改提供了方便。

许多初学者经常忽略第 1 步和第 2 步 (定义程序目标 and 设计程序), 直接跳到第 3 步 (编写代码)。刚开始学习时, 编写的程序非常简单, 完全可以在脑中构思好整个过程。即使写错了, 也很容易发现。但是, 随着编写的程序越来越庞大、越来越复杂, 动脑不动手可不行, 而且程序中隐藏的错误也越来越难找。最终, 那些跳过前两个步骤的人往往浪费了更多的时间, 因为他们写出的程序难看、缺乏条理、让人难以理解。要编写的程序越大越复杂, 事先定义和设计程序环节的工作量就越大。

磨刀不误砍柴工, 应该养成先规划再动手编写代码的好习惯, 用纸和笔记录下程序的目标和设计框架。这样在编写代码的过程中会更加得心应手、条理清晰。

1.8 编程机制

生成程序的具体过程因计算机环境而异。C 是可移植性语言, 因此可以在许多环境中使用, 包括 UNIX、Linux、MS-DOS (一些人仍在使用)、Windows 和 Macintosh OS。有些产品会随着时间的推移发生演变或被取代, 本书无法涵盖所有环境。

首先, 来看看许多 C 环境 (包括上面提到的 5 种环境) 共有的一些方面。虽然不必详细了解计算机内部如何运行 C 程序, 但是, 了解一下编程机制不仅能丰富编程相关的背景知识, 还有助于理解为何要经过一些特殊的步骤才能得到 C 程序。

用 C 语言编写程序时, 编写的内容被储存在文本文件中, 该文件被称为源代码文件 (source code file)。大部分 C 系统, 包括之前提到的, 都要求文件名以 .c 结尾 (如, wordcount.c 和 budget.c)。在文件名中, 点号 (.) 前面的部分称为基本名 (basename), 点号后面的部分称为扩展名 (extension)。因此, budget 是基本名, c 是扩展名。基本名与扩展名的组合 (budget.c) 就是文件名。文件名应该满足特定计算机操作系统的特殊要求。例如, MS-DOS 是 IBM PC 及其兼容机的操作系统, 比较老旧, 它要求基本名不能超过 8 个字符。因此, 刚才提到的文件名 wordcount.c 就是无效的 DOS 文件名。

有些 UNIX 系统限制整个文件名 (包括扩展名) 不超过 14 个字符, 而有些 UNIX 系统则允许使用更长的文件名, 最多 255 个字符。Linux、Windows 和 Macintosh OS 都允许使用长文件名。

接下来, 我们来看一下具体的应用, 假设有一个名为 `concrete.c` 的源文件, 其中的 C 源代码如程序清单 1.2 所示。

程序清单 1.2 c 程序

```
#include <stdio.h>
int main(void)
{
    printf("Concrete contains gravel and cement.\n");
    return 0;
}
```

如果看不懂程序清单 1.2 中的代码, 不用担心, 我们将在第 2 章学习相关知识。

1.8.1 目标代码文件、可执行文件和库

C 编程的基本策略是, 用程序把源代码文件转换为可执行文件 (其中包含可直接运行的机器语言代码)。典型的 C 实现通过编译和链接两个步骤来完成这一过程。编译器把源代码转换成中间代码, 链接器把中间代码和其他代码合并, 生成可执行文件。C 使用这种分而治之的方法方便对程序进行模块化, 可以独立编译单独的模块, 稍后再用链接器合并已编译的模块。通过这种方式, 如果只更改某个模块, 不必因此重新编译其他模块。另外, 链接器还将你编写的程序和预编译的库代码合并。

用程序把源代码文件转换为可执行文件（编译程序及链接程序）

两步：编译和链接

编译

编译器把源代码转换成中间代码（obj）

链接

把中间代码和其他代码合并，生成可执行文件。

分而治之方便对程序模块化。

可以独立编译单独的模块，稍后再用链接器合并已编译的模块。通过这种方式，如果只改某个模块，不必因此重新编译其他模块。另外，链接器还将你编写程序和预编译的库代码合并。

中间文件有多种形式。

我们在这里描述的是最普遍的一种形式，即把源代码转换为机器语言代码，并把结果放在目标代码文件（或简称目标文件）中（这里假设源代码只有一个文件）。虽然目标文件中包含机器语言代码，但是并不能直接运行该文件。因为目标文件中储存的是编译器翻译的源代码，这还不是一个完整的程序。

目标代码文件缺失启动代码（startup code）。

1. 启动代码充当着程序和操作系统之间的接口。

例如，可以在 MS Windows 或 Linux 系统下运行 IBM PC 兼容机。这两种情况所使用的硬件相同，所以目标代码相同，/但是 Windows 和 Linux 所需的启动代码不同，/因为这些系统处理程序的方式不同。

2. 目标代码还缺少库函数。

几乎所有的 C 程序都要使用 C 标准库中的函数。例如，concrete.c 中就使用了 printf() 函数。目标代码文件并不包含该函数的代码，它只包含了使用 printf() 函数的指令。printf() 函数真正的代码储存在另一个被称为库的文件中。

3. 库文件中有许多函数的目标代码。

链接器

的作用是，把你编写的目标代码、系统的标准启动代码和库代码这 3 部分合并成一个文件，即可执行文件。对于库代码，链接器只会把程序中要用到的库函数代码提取出来 (见图 1.4)。

图 1.4 编译器和链接器

简而言之，目标文件和可执行文件都由机器语言指令组成的。然而，目标文件中只包含编译器为你编写的代码翻译的机器语言代码，可执行文件中还包含你编写的程序中使用的库函数和启动代码的机器代码。

在有些系统中，必须分别运行编译程序和链接程序，而在另一些系统中，编译器会自动启动链接器，用户只需给出编译命令即可。

接下来，了解一些具体的系统。

目标文件和可执行文件都由机器语言指令组成

目标文件只包含编译器为你编写的代码翻译的机器语言代码

可执行文件中还包含你的程序中使用的库函数和启动代码的机器代码。

即：可执行文件 = 目标代码 + 库函数 + 启动代码

1.8.2 UNIX 系统

由于 C 语言因 UNIX 系统而生，也因此而流行，所以我们从 UNIX 系统开始 (注意: 我们提到的 UNIX 还包含其他系统，如 FreeBSD，它是 UNIX 的一个分支，但是由于法律原因不使用该名称)。

1. 在 UNIX 系统上编辑

UNIX C 没有自己的编辑器，但是可以使用通用的 UNIX 编辑器，如 emacs、jove、vi 或 X Window System 文本编辑器。作为程序员，要负责输入正确的程序和为储存该程序的文件起一个合适的文件名。如前所述，文件名应该以.c 结尾。注意，UNIX 区分大小写。因此，budget.c、BUDGET.c 和 Budget.c 是 3 个不同但都有效的 C 源文件名。但是 BUDGET.C 是无效文件名，因为该名称的扩展名使用了大写 C 而不是小写 c。

假设我们在 vi 编译器中编写了下面的程序，并将其储存在 inform.c 文件中：

```
#include <stdio.h>
int main(void)
{
    printf("A .c is used to end a C program filename.\n");
    return 0;
}
```

以上文本就是源代码，inform.c 是源文件。注意，源文件是整个编译过程的开始，不是结束。

2. 在 UNIX 系统上编译

虽然在我们看来，程序完美无缺，但是对计算机而言，这是一堆乱码。计算机不明白 #include 和 printf 是什么（也许你现在也不明白，但是学到后面就会明白，而计算机却不会）。如前所述，我们需要编译器将我们编写的代码（源代码）翻译成计算机能看懂的代码（机器代码）。最后生成的可执行文件中包含计算机要完成任务所需的所有机器代码。

以前，UNIX C 编译器要调用语言定义的 cc 命令。但是，它没有跟上标准发展的脚步，已经退出了历史舞台。但是，UNIX 系统提供的 C 编译器通常来自一些其他源，然后以 cc 命令作为编译器的别名。因此，虽然在不同的系统中会调用不同的编译器，但用户仍可以继续使用相同的命令。

编译 `inform.c`，要输入以下命令：

```
cc inform.c
```

几秒钟后，会返回 UNIX 的提示，告诉用户任务已完成。如果程序编写错误，你可能会看到警告或错误消息，但我们先假设编写的程序完全正确（如果编译器报告 `void` 的错误，说明你的系统未更新成 ANSI C 编译器，只需删除 `void` 即可）。如果使用 `ls` 命令列出文件，会发现有一个 `a.out` 文件（见图 1.5）。该文件是包含已翻译（或已编译）程序的可执行文件。要运行该文件，只需输入：

```
a.out
```

输出内容如下：

```
A .c is used to end a C program filename.
```

图 1.5 用 UNIX 准备 C 程序如果要储存可执行文件 (`a.out`)，应该把它重命名。否则，该文件会被下一次编译程序时生成的新 `a.out` 文件替换。如何处理目标代码？C 编译器会创建一个与源代码基本名相同的目标代码文件，但是其扩展名是 `.o`。在该例中，目标代码文件是 `inform.o`。然而，却找不到这个文件，因为一旦链接器生成了完整的可执行程序，就会将其删除。如果原始程序有多个源代码文件，则保留目标代码文件。学到后面多文件程序时，你会明白到这样做的好处。

1.8.3 GNU 编译器集合和 LLVM 项目

GNU 项目始于 1987 年，是一个开发大量免费 UNIX 软件的集合（GNU 的意思是“GNU’s Not UNIX”，即 GNU 不是 UNIX）。GNU 编译器集合（也被称为 GCC，其中包含 GCC C 编译器）是该项目的产品之一。GCC 在一个指导委员会的带领下，持续不断地开发，它的 C 编译器紧跟 C 标准的改动。GCC 有各种版本以适应不同的硬件平台和操作系统，包括 UNIX、Linux 和 Windows。用 `gcc` 命令便可调用 GCC C 编译器。许多使用 `gcc` 的系统都用 `cc` 作为 `gcc` 的别名。

LLVM 项目成为 `cc` 的另一个替代品。该项目是与编译器相关的开源软件集合，始于伊利诺伊大学的 2000 份研究项目。它的 Clang 编译器处理 C

代码，可以通过 clang 调用。有多种版本供不同的平台使用，包括 Linux。2012 年，Clang 成为 FreeBSD 的默认 C 编译器。Clang 也对最新的 C 标准支持得很好。

GNU 和 LLVM 都可以使用 -v 选项来显示版本信息，因此各系统都使用 cc 别名来代替 gcc 或 clang 命令。以下组合：

```
cc -v
```

显示你所使用的编译器及其版本。

gcc 和 clang 命令都可以根据不同的版本选择运行时选项来调用不同 C 标准。gcc -std=c99 inform.c[3] gcc -std=c1x inform.c gcc -std=c11 inform.c

第 1 行调用 C99 标准，第 2 行调用 GCC 接受 C11 之前的草案标准，第 3 行调用 GCC 接受的 C11 标准版本。Clang 编译器在这一点上用法与 GCC 相同。

1.8.4 Linux 系统

Linux 是一个开源、流行、类似于 UNIX 的操作系统，可在不同平台（包括 PC 和 Mac）上运行。在 Linux 中准备 C 程序与在 UNIX 系统中几乎一样，不同的是要使用 GNU 提供的 GCC 公共域 C 编译器。编译命令类似于：

```
gcc inform.c
```

注意，在安装 Linux 时，可选择是否安装 GCC。如果之前没有安装 GCC，则必须安装。通常，安装过程会将 cc 作为 gcc 的别名，因此可以在命令行中使用 cc 来代替 gcc。

欲详细了解 GCC 和最新发布的版本，请访问 <http://www.gnu.org/software/gcc/index.html>。

1.8.5 PC 的命令行编译器

C 编译器不是标准 Windows 软件包的一部分，因此需要从别处获取并安装 C 编译器。可以从互联网免费下载 Cygwin 和 MinGW，这样便可

在 PC 上通过命令行使用 GCC 编译器。Cygwin 在自己的视窗运行，模仿 Linux 命令行环境，有一行命令提示。MinGW 在 Windows 的命令提示模式中运行。这和 GCC 的最新版本一样，支持 C99 和 C11 最新的一些功能。Borland 的 C++ 编译器 5.5 也可以免费下载，支持 C90。

源代码文件应该是文本文件，不是字处理器文件（字处理器文件包含许多额外的信息，如字体和格式等）。因此，要使用文本编辑器（如，Windows Notepad）来编辑源代码。如果使用字处理器，要以文本模式另存文件。源代码文件的扩展名应该是.c。一些字处理器会为文本文件自动添加.txt 扩展名。如果出现这种情况，要更改文件名，把 txt 替换成 c。

通常，C 编译器生成的中间目标代码文件的扩展名是.obj（也可能是其他扩展名）。与 UNIX 编译器不同，这些编译器在完成编译后通常不会删除这些中间文件。有些编译器生成带.asm 扩展名的汇编语言文件，而有些编译器则使用自己特有的格式。

一些编译器在编译后会自动运行链接器，另一些要求用户手动运行链接器。在可执行文件中链接的结果是，在原始的源代码基本名后面加上.exe 扩展名。例如，编译和链接 concrete.c 源代码文件，生成的是 concrete.exe 文件。可以在命令行输入基本名来运行该程序：

```
C>concrete
```

1.8.6 集成开发环境 (Windows)

许多供应商（包括微软、Embarcadero、Digital Mars）都提供 Windows 下的集成开发环境，或称为 IDE（目前，大多数 IDE 都是 C 和 C++ 结合的编译器）。可以免费下载的 IDE 有 Microsoft Visual Studio Express 和 Pelles C。利用集成开发环境可以快速开发 C 程序。关键是，这些 IDE 都内置了用于编写 C 程序的编辑器。这类集成开发环境都提供了各种菜单（如，命名、保存源代码文件、编译程序、运行程序等），用户不用离开 IDE 就能顺利编写、编译和运行程序。如果编译器发现错误，会返回编辑器中，标出有错误的行号，并简单描述情况。

初次接触 Windows IDE 可能会望而生畏，因为它提供了多种目标（tar-

get), 即运行程序的多种环境。例如, IDE 提供了 32 位 Windows 程序、64 位 Windows 程序、动态链接库文件 (DLL) 等。许多目标都涉及 Windows 图形界面。要管理这些 (及其他) 选择, 通常要先创建一个项目 (project), 以便稍后在其中添加待使用的源代码文件名。不同的产品具体步骤不同。一般而言, 首先使用【文件】菜单或【项目】菜单创建一个项目。选择正确的项目形式非常重要。本书中的例子都是一般示例, 针对在简单的命令行环境中运行而设计。Windows IDE 提供多种选择以满足用户的不同需求。例如, Microsoft Visual Studio 提供【Win32 控制台应用程序】选项。对于其他系统, 查找一个诸如【DOS EXE】、【Console】或【Character Mode】的可执行选项。选择这些模式后, 将在一个类控制台窗口中运行可执行程序。选择好正确的项目类型后, 使用 IDE 的菜单打开一个新的源代码文件。对于大多数产品而言, 使用【文件】菜单就能完成。你可能需要其他步骤将源文件添加到项目中。

通常, Windows IDE 既可处理 C 也可处理 C++, 因此要指定待处理的程序是 C 还是 C++。有些产品用项目类型来区分两者, 有些产品 (如, Microsoft Visual C++) 用 .c 文件扩展名来指明使用 C 而不是 C++。当然, 大多数 C 程序也可以作为 C++ 程序运行。欲了解 C 和 C++ 的区别, 请参阅参考资料 IX。

你可能会遇到一个问题: 在程序执行完毕后, 执行程序的窗口立即消失。如果不希望出现这种情况, 可以让程序暂停, 直到按下 Enter 键, 窗口才消失。要实现这种效果, 可以在程序的最后 (return 这行代码之前) 添加下面一行代码:

```
getchar();
```

该行读取一次键的按下, 所以程序在用户按下 Enter 键之前会暂停。有时根据程序的需要, 可能还需要一个击键等待。这种情况下, 必须用两次 getchar():

```
getchar();
```

```
getchar();
```

例如, 程序在最后提示用户输入体重。用户键入体重后, 按下 Enter 键

以输入数据。程序将读取体重，第 1 个 `getchar()` 读取 Enter 键，第 2 个 `getchar()` 会导致程序暂停，直至用户再次按下 Enter 键。如果你现在不知所云，没关系，在学完 C 输出后就会明白。到时，我们会提醒读者使用这种方法。

虽然许多 IDE 在使用上大体一致，但是细节上有所不同。就一个产品的系列而言，不同版本也是如此。要经过一段时间的实践，才会熟悉编译器的工作方式。必要时，还需阅读使用手册或网上教程。

Microsoft Visual Studio 和 C 标准在 Windows 软件开发中，Microsoft Visual Studio 及其免费版本 Microsoft Visual Studio Express 都久负盛名，它们与 C 标准的关系也很重要。然而，微软鼓励程序员从 C 转向 C++ 和 C#。虽然 Visual Studio 支持 C89/90，但是到目前为止，它只选择性地支持那些在 C++ 新特性中能找到的 C 标准 (如，`long long` 类型)。而且，自 2012 版本起，Visual Studio 不再把 C 作为项目类型的选项。尽管如此，本书中的绝大多数程序仍可用 Visual Studio 来编译。在新建项目时，选择 C++ 选项，然后选择 **【Win32 控制台应用程序】**，在应用设置中选择 **【空项目】**。几乎所有的 C 程序都能与 C++ 程序兼容。所以，本书中的绝大多数 C 程序都可作为 C++ 程序运行。或者，在选择 C++ 选项后，将默认的源文件扩展名 `.cpp` 替换成 `.c`，编译器便会使用 C 语言的规则代替 C++。

1.8.7 Windows/Linux

许多 Linux 发行版都可以安装在 Windows 系统中，以创建双系统。一些存储器会为 Linux 系统预留空间，以便可以启动 Windows 或 Linux。可以在 Windows 系统中运行 Linux 程序，或在 Linux 系统中运行 Windows 程序。不能通过 Windows 系统访问 Linux 文件，但是可以通过 Linux 系统访问 Windows 文档。

1.8.8 Macintosh 中的 C

目前，苹果免费提供 Xcode 开发系统下载 (过去，它有时免费，有时付费)。它允许用户选择不同的编程语言，包括 C 语言。

Xcode 凭借可处理多种编程语言的能力，可用于多平台，开发超大型的项目。但是，首先要学会如何编写简单的 C 程序。在 Xcode 4.6 中，通过【File】菜单选择【New Project】，然后选择【OS X Application Command Line Tool】，接着输入产品名并选择 C 类型。Xcode 使用 Clang 或 GCC C 编译器来编译 C 代码，它以前默认使用 GCC，但是现在默认使用 Clang。可以设置选择使用哪一个编译器和哪一套 C 标准 (因为许可方面的事宜，Xcode 中 Clang 的版本比 GCC 的版本要新)。

UNIX 系统内置 Mac OS X，终端工具打开的窗口是让用户在 UNIX 命令行环境中运行程序。苹果在标准软件包中不提供命令行编译器，但是，如果下载了 Xcode，还可以下载可选的命令行工具，这样就可以使用 clang 和 gcc 命令在命令行模式中编译。

1.9 本书的组织结构

本书采用多种方式编排内容，其中最直接的方法是介绍 A 主题的所有内容、介绍 B 主题的所有内容，等等。这对参考类书籍来说尤为重要，读者可以在同一处找到与主题相关的所有内容。但是，这通常不是学习的最佳顺序。例如，如果在开始学习英语时，先学完所有的名词，那你的表达能力一定很有限。虽然可以指着物品说出名称，但是，如果稍微学习一些名词、动词、形容词等，再学习一些造句规则，那么你的表达能力一定会大幅提高。

为了让读者更好地吸收知识，本书采用螺旋式方法，先在前几个章节中介绍一些主题，在后面章节再详细讨论相关内容。例如，对学习 C 语言而言，理解函数至关重要。因此，我们在前几个章节中安排一些与函数相关的内容，等读者学到第 9 章时，已对函数有所了解，学习使用函数会更加容易。与此类似，前几章还概述了一些字符串和循环的内容。这样，读者

在完全看懂这些内容之前，就可以在自己的程序中使用这些有用的工具。

1.10 本书的约定

在学习 C 语言之前，先介绍一下本书的格式。

1.10.1 字体本书用类似在屏幕上或打印输

出时的字体（一种等宽字体），表示文本程序和计算机输入、输出。前面已经出现了多次，如果读者没有注意到，字体如下所示：

```
#include <stdio.h> int main(void) { printf("Concrete contains gravel  
and cement."); return 0; }
```

在涉及与代码相关的术语时，也使用相同的等宽字体，如 `stdio.h`。本书用等宽斜体表示占位符，可以用具体的项替换这些占位符。例如，下面是一个声明的模型：

`typename variablename`；这里，可用 `int` 替换 `typename`，用 `zebracount` 替换 `variablename`。

1.10.2 程序输出

本书用相同的字体表示计算机的输出，粗体表示用户输入。例如，下面是第 14 章中一个程序的输出：

```
Please enter the book title. Press [enter] at the start of a line to stop.  
My Life as a BudgieNow enter the author.Mack Zackle s 如上所示，以标准  
计算机字体显示的行表示程序的输出，粗体行表示用户的输入。可以通过  
多种方式与计算机交互。在这里，我们假设读者使用键盘键入内容，在屏  
幕上阅读计算机的响应。
```

1. 特殊的击键

通常，通过按下标有 Enter、c/r、Return 或一些其他文字的键来发送指令。本书将这些按键统一称为 Enter 键。一般情况下，我们默认你在每行输入的末尾都会按下 Enter 键。尽管如此，为了标示一些特定的位置，本书使用 [enter] 显式标出 Enter 键。方括号表示按下一次 Enter 键，而不是输入 enter。

除此之外，书中还会提到控制字符 (如，Ctrl+D)。这种写法的意思是，在按下 Ctrl 键 (也可能是 Control 键) 的同时按下 D 键。

2. 本书使用的系统

C 语言的某些方面 (如，储存数字的空间大小) 因系统而异。本书在示例中提到“我们的系统”时，通常是指在 iMac 上运行 OS X 10.8.4，使用 Xcode 4.6.2 开发系统的 Clang 3.2 编译器。本书的大部分程序都能使用 Windows7 系统的 Microsoft Visual Studio Express 2012 和 Pelles C 7.0，以及 Ubuntu13.04 Linux 系统的 GCC 4.7.3 进行编译。

3. 读者的系统

你需要一个 C 编译器或访问一个 C 编译器。C 程序可以在多种计算机系统中运行，因此你的选择面很广。确保你使用的 C 编译器与当前使用的计算机系统匹配。本书中，除了某些示例要求编译器支持 C99 或 C11 标准，其余大部分示例都可在 C90 编译器中运行。如果你使用的编译器是早于 ANSI/ISO 的老式编译器，在编译时肯定要经常调整，很不方便。与其如此，不如换个新的编译器。

大部分编译器供应商都为学生和教学人员提供特惠版本，详情请查看供应商的网站。

1.10.3 特殊元素

本书包含一些强调特定知识点的特殊元素，提示、注意、警告，将以如下形式出现在本书中：

- 边栏边栏提供更深入的讨论或额外的背景，有助于解释当前的主题。
- 提示提示一般都短小精悍，帮助读者理解一些特殊的编程情况。
- 警告

用于警告读者注意一些潜在的陷阱。注意提供一些评论，提醒读者不要误入歧途。

1.11 本章小结

C 是强大而简洁的编程语言。它之所以流行，在于自身提供大量的实用编程工具，能很好地控制硬件。而且，与大多数其他程序相比，C 程序更容易从一个系统移植到另一个系统。

C 是编译型语言。C 编译器和链接器是把 C 语言源代码转换成可执行代码的程序。用 C 语言编程可能费力、困难，让你感到沮丧，但是它也可以激发你的兴趣，让你兴奋、满意。我们希望你在愉快的学习过程中爱上 C。

1.12 复习题

复习题的参考答案在附录 A 中。

- 1.12.1 1. 对编程而言, 可移植性意味着什么?
 - 1.12.2 2. 解释源代码文件、目标代码文件和可执行文件有什么区别?
 - 1.12.3 3. 编程的 7 个主要步骤是什么?
 - 1.12.4 4. 编译器的任务是什么?
 - 1.12.5 5. 链接器的任务是什么?
- 1.13 编程练习我们尚未要求你编写 C 代码, 该练习侧重于编程过程的早期步骤。**

1. 你刚被 MacroMuscle 有限公司聘用。该公司准备进入欧洲市场, 需要一个把英寸单位转换为厘米单位 (1 英寸 = 2.54 厘米) 的程序。该程序要提示用户输入英寸值。你的任务是定义程序目标 and 设计程序 (编程过程的第 1 步和第 2 步)。

[1]. 国际 C 语言混乱代码大赛 (IOCCC, The International Obfuscated C Code Contest)。这是一项国际编程赛事, 从 1984 年开始, 每年举办一次 (1997、1999、2002、2003 和 2006 年除外), 目的是写出最有创意且最让人难以理解的 C 语言代码。——译者注

[2]. VAX (Virtual Address eXtension) 是一种可支持机器语言和虚拟地址的 32 位小型计算机。VMS (Virtual Memory System) 是旧名, 现在叫 OpenVMS, 是一种用于服务器的操作系统, 可在 VAX、Alpha 或 Itanium 处理器系列平台上运行。——译者注

[3]. GCC 最基本的用法是: gcc [options] [filenames], 其中 options 是所需的参数, filenames 是文件名。——译者注

第二章 C 语言概述

本章介绍以下内容：

- 运算符：=
- 函数：main()、printf()
- 编写一个简单的 C 程序
- 创建整形变量，为其赋值并在屏幕上显示其值
- 换行字符
- 如何在程序中写注释，创建包含多个函数的程序，发现程序的错误
- 什么是关键字

C 程序是什么样子的？浏览本书，能看到许多示例。初见 C 程序会觉得有些古怪，程序中的许多 {、cp-tor 和 ptr++ 这样的符号。然而，在学习 C 的过程中，对这些符号和 C 语言特有的其他符号会越来越熟悉，甚至会喜欢上它们。如果熟悉与 C 相关的其他语言，会对 C 语言有似曾相识的感觉。本章，我们从演示一个简单的程序示例开始，解释该程序的功能。同时，强调一些 C 语言的基本特性。

2.1 简单的 C 程序示例

我们来看一个简单的 C 程序，如程序清单 2.1 所示。该程序演示了用 C 语言编程的一些基本特性。请先通读程序清单 2.1，看看自己是否能明白该程序的用途，再认真阅读后面的解释。

程序清单 2.1 first.c 程序

```
// This is 2.1 first.c
//first
//
//Created by Mac on<2020-04-16 Thu>
//Copyright @ 2020 Mac All rights reserved.
//
#include<stdio.h>
int main(void) /*一个简单的C程序*/
{
    int num; /*定义一个名为num的变量*/
    num=1; /*为num赋一个值*/
    printf("I am a simple"); /*使用printf()函数*/
    printf("computer.\n");
    printf("My favorite number is %d because it is first.\n",num);
    return ;
}
```

输出内容为：

I am a simple computer.

My favorite number is 1 because it is first.

程序调整

程序的输出是否在屏幕上一闪而过？某些窗口环境会在单独的窗口运行程序，然后在程序运行结束后自动关闭窗口。如果遇到这种情况，可以在程序中加代码，让窗口等待用户按下一个键后才关闭。一种方法是，在

程序的 `return` 语句前加一行代码：

```
getchar();
```

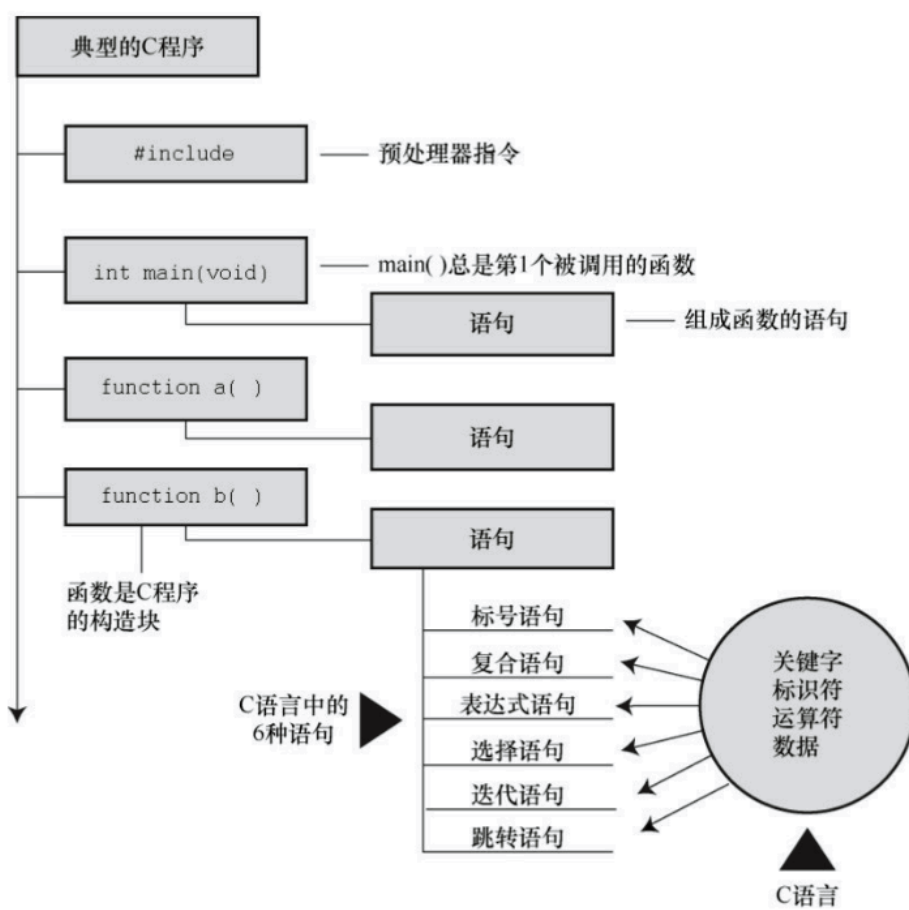
这行代码让程序等待击键，窗口会在用户按下一个键后才关闭。在第八章详细介绍 `getchar()` 的内容。

2.2 示例解释

把程序清单 2.1 分析两遍。第 1 遍（快速概要）概述程序中每行代码的作用，帮助读者初步了解程序。第 2 遍（程序细节）详细分析代码的具体含义，帮助读者深入理解程序。

图 2.1 总结了组成 C 程序的几个部分，图中包含的元素比第 1 个程序多

如何插入图：



函数中 C 的 6 种语句:

- 标号语句
- 复合语句
- 表达式语句
- 选择语句
- 迭代语句
- 跳转语句

C 语言语句的构成：

- 关键字
- 标识符
- 运算符
- 数据

2.2.1 第 1 遍：快速概要

本节简述程序中的每行代码的作用。下一节详细讨论代码的含义。

`#include<stdio.h>` 包含另一个文件

该行告诉编译器把 `stdio.h` 中的内容包含在当前程序中。`stdio.h` 是 C 编译器软件包的标准部分，它提供键盘输入和屏幕输出的支持。

`int main(void)` 函数名

C 程序包含一个或多个函数，它们是 Ctkcyr 基本模块。程序清单 2.1 的程序中有一个名为 `main()` 的函数。圆括号表明 `main()` 是一个函数名。`int` 表明 `main()` 函数返回一个整数，`void` 表明 `main()` 不带任何参数。`int` and `void` 是标准 ANSI C 定义 `main()` 的一部分。如果使用 ANSI C 之前的编译器，可省略 `void`；考虑兼容的问题，使用较新的 C 编译器。

一个简单的 C 程序 注释

注释在 `/*` 和 `*/` 两个符号之间，这些注释能提高程序的可读性。注释只是帮助读者理解程序，编译器会忽略它们。

{ 函数体开始

左花括号表示函数定义开始，右花括号 (`}`) 表示函数定义结束。

`int num;` 声明

该声明表明，将使用一个名为 `num` 的变量，而且 `num` 是 `int`(整形) 类型。

`num=1;` 赋值表达式语句

语句 `num=1`; 把值 1 赋给名为 `num` 的变量。

`printf("I am a simple");` 调用一个函数

该语句使用 `printf()` 函数，在屏幕上显示 `I am a simple`，光标停在同一行。`printf()` 是标准的 C 库函数。在程序中使用函数叫作 * 调用函数 *。

`printf("computer.\n");` 调用另一个函数

接下来调用的这个 `printf()` 函数在上条语句打印出来的内容后面加上 `"computer"`。代码告诉计算机另起一行，即把光标移至下一行。

`printf("My favorite number is %d because it is first.",num);`

最后调用的 `printf()` 把 `num` 的值 (1) 内嵌在用双引号括起来的内容中一并打印。`%d` 告诉计算机以何种形式输出 `num` 的值，打印在何处。

`return 0;` `return` 语句

C 函数可以给调用方提供（或返回）一个数。目前，可暂时把该行看作是结束 `main()` 函数的要求。

} 结束

必须以右花括号表示程序结束。

2.2.2 第 2 遍：程序细节

浏览完程序清单 2.2 后，仔细分析程序。再次强调，将逐行分析程序中的代码，以每行代码为出发点，深入分析代码背后的细节，为更全面的学习 C 语言编程的特性夯实基础。

1. `#include` 指令和头文件

```
#include<stdio.h>
```

这是程序的第一行。`#include<stdio.h>` 的作用相当于把 `stdio.h` 文件中的所有内容都输入该行所在的位置。实际上，这是一种“拷贝-粘贴”的操作。

`include` 文件提供了一种方便的途径共享许多程序共有的信息。

`#include` 这行代码是一条 C 预处理器指令 (preprocessor directive)。通常, C 编译器在编译 `#` 前会对源代码做一些准备工作, 即预处理 (preprocessing)。

所有的 C 编译器软件都提供 `stdio.h` 文件。该文件中包含了供编译器使用的输入和输出函数 (如, `printf()`) 信息。该文件名的含义是标准输入/输出头文件。通常, 在 C 程序顶部的信息集合被称为头文件 (header)。

在大多数情况下, 头文件包含了编译器创建最终可执行程序要用到的信息。例如, 头文件中可以定义一些常量, 或者指明函数名以及如何使用它们。但是, 函数的实际代码在一个预编译代码的库文件中。简而言之, 头文件帮助编译器把你的程序正确的组合在一起。

ANSI/ISO C 规定了 C 编译器必须提供哪些头文件。有些程序要包含 `stdio.h`, 而有些不用。特定 C 实现的文档中应该包含对 C 库函数的说明。这些说明确定了使用哪些函数需要包含哪些头文件。例如, 要使用 `printf()` 函数, 必须包含 `stdio.h` 头文件。省略必要的头文件可能不会影响某一特定程序, 但是最好不要这样做。本书每次用到库函数, 都会用 `#include` 指令包含 ANSI/ISO 标准指定的头文件。

注意为何不内置输入和输出

读者一定很好奇, 为何不把输入和输出这些基本功能内置在语言中。原因之一是, 并非所有的程序都会用到 I/O (输入/输出) 包。轻装上阵表现了 C 语言的哲学。正是这种经济使用资源的原则, 使得 C 语言成为流行的嵌入式编程语言 (例如, 编写控制汽车自动燃油系统或蓝光播放机芯片的代码)。`#include` 中的 `#` 符号表明, C 预处理器在编译器接手之前处理这条指令。本书后面章节中会介绍更多预处理器指令的示例, 第 16 章将更详细地讨论相关内容。

2.main() 函数

```
int main(void);
```

程序清单 2.2 中的第 2 行表明该函数名为 `main`。的确, `main` 是一个极其普通的名称, 但是这是唯一的选择。C 程序一定从 `main()` 函数开始执

行（目前不必考虑例外的情况）。坠于 `main()` 函数，你可以任意命名其他函数，而且 `main()` 函数必须是开始的函数。圆括号有什么功能？用于识别 `main()` 是一个函数。很快你将学到更多的函数。目前而言，只需记住函数是 C 语序的基本模块。

`int` 是 `main()` 函数的返回类型。表明 `main()` 函数返回的值是整数。返回到哪里？返回给操作系统。我们将在第 6 章再来探讨这个问题。

通常，函数名后面的圆括号中包含一些传入函数的信息。该例中没有传递任何信息。因此，圆括号内是单词 `void`（第 11 章次介绍把信息从 `main()` 函数传回操作系统的另一种形式）。

如果浏览旧式的 C 代码，会发现程序以如下形式开始：

```
main()
```

C90 标准勉强接受这种形式，但是 C99 和 C11 标准不允许这样写。因此，即使你使用的编译器允许，也不要这样写。

你还会看到下面这种形式：

```
void main()
```

一些编译器允许这样写，但是所有的标准都未认可这种写法。因此，编译器不必接受这种形式，而且许多编译器都不能这样写。需要强调的是，只要坚持使用标准形式，把程序从一个编译器移至另一个编译器时就不会出什么问题。

3. 注释

一个简单的程序

在程序中，被 `/**/` 两个符号括起来的部分是程序的注释。写注释能让他人（包括自己）更容易明白你所写的程序。C 语言注释的好处之一是，可将注释放在任意的地方，甚至是与要解释的内容在同一行。较长的注释可单独放一行或多行。在 `/*` 和 `*/` 之间的内容都会被编译器忽略。下面列出了一些有效和无效的注释形式：

这是一条 C 注释。

`/* 这也是一条注释，`

被分成了两行。*/

/*

也可以这样写注释。

*/

/* 这条注释无效，因为缺少了结束标记。

C99 新增了另一种风格的注释，普遍用于 C++ 和 Java。这种新风格使用//符号创建注释，仅限于单行。

//这种注释只能写成一行。

int rigue;//这种注释也可置于此。

因为一行末尾就标志着注释的结束，所以这种风格的注释只需在注释开始下标明//符号即可。

这种新形式的注释是为了解决旧形式注释存在的潜在问题。假设有下面的代码：

```
/*
```

```
希望能运行
```

```
*/
```

```
x=100;
```

```
Y=200;
```

```
/* 其它内容已省略。 */
```

接下来，假设你决定删除第 4 行，但不小心删掉了第 3 行 (*/)。代码如下所示：

```
/*
```

希望能运行。

```
y=200;
```

```
/*其他内容已省略。*/
```

现在，编译器把第 1 行的 `/*` 和第 4 行的 `*/` 配对，导致 4 行代码全部成了注释（包括应作为代码的那一行）。而 `//` 形式的注释只对单行有效，不会导致这种“消失代码”的问题。

一些编译器可能不支持这一特性。还有一些编译器需要更改设置，才能支持 C99 或 C11 的特性。

考虑到只有一种注释风格过于死板乏味，本书在示例中采用两种风格的注释。

4. 花括号、函数体和块

```
{  
...  
}
```

程序清单 2.1 中，花括号把 `main()` 函数括起来。一般而言，所有的 C 函数都使用花括号标记函数体的开始和结束。这是规定，不能省略。只有花括号（`{}`）能起这种作用，圆括号（`()`）的方括号（`[]`）都不行。

花括号还可用于把函数中的多条语句合并为一个单元或块。如果读者熟悉 Pascal、ADA、Modula-2 或者 Algol，就会明白花括号在 C 语言中的作用类似于这些语言中的 `begin` 和 `end`。

5. 声明

```
int num;
```

程序清单 2.1 中，这行代码叫作声明 (declaration)。声明是 C 语言最重要的特性之一。在该例中，声明完成了两件事。

其一，在函数中有一个名为 num 的变量 (variable)。

其二，int 表明 num 是一个整数（即没有小数点或小数部分的数）。int 是一种数据类型。编译器使用这些信息为 num 变量在内存中分配存贮空间。分号在 C 语言中是大部分语句和声明的平分，不像在 Pascal 中只是语句间的分隔符。

int 是 C 语言的一个关键字 (keyword)，表示一种基本的 C 语言数据类型。关键字是语言定义的单词，不能做其他用途。例如，不能用 int 作为函数名或变量名。但是，这些关键字在该语言外不起作用，所以把一只猫或一个可爱的小孩叫 int 是可以的（尽管某些地方的当地习俗或法律可能不允许）。

示例中的 num 是一个标识符 (identifier)，也就一个变量、函数或其他实体的名称。因此，声明把特定标识符与计算机内存中的特定位置联系起来，同时也确定了储存在某位置的信息类型或数据类型。

在 C 语言中，所有变量都必须先声明才能使用。这章法着必须列出程序中用到的所有变量名及其类型。

以前的 C 语言，还要求把变量声明在块的顶部，其他语句不能在任何声明的前面。也就是说，main() 函数体如下所示：

```
int main() //旧规则
{
    int doors;
    int dogs;
    doors=5;
    dogs=3;
    //其他语句

}
```

C99 和 C11 遵循 C++ 的惯例，可以把声明放在块中任何位置。尽管如此，首次使用变量之前一定要先声明它。因此，如果编译器支持这一新特性，可以这样编写上面的代码：

```
int main()    //目前的C规则
{
    //一些语句
    int doors;
    doors=5; //第1次使用 doors
    //其他语句
    int dogs;
    dogs=3; //第1次使用 dogs
    //其他语句
}
```

为了与旧系统更好地兼容，本书沿用最初的规则（即把变量声明都写在块的顶部）。

现在，读者可能有 3 问题：

- 什么是数据类型？
- 如何命名？
- 为何要声明变量？

1. 数据类型

C 语言可以处理多种类型的数据，如整数、字符和浮点数。把变量声明为整形或字符类型，计算机才能正确地储存、读取和解释数据。下一章将详细介绍 C 语言中的各种数据类型。

2. 命名

给变量命名时要使用有意义的变量或标识符（如，程序中需要一个变量数狗，该变量名应该是 `dogcount` 而不是 `x3`）。如果变量名无法清楚

地表达自身的用途，可在注释中进一步说明。这是一种良好的编程习惯和编程技巧。

C99 和 C11 允许使用更长的标识符名，但是编译器中识别前 63 个字符。对于外部标识符（参阅第 12 章），只允许使用 31 个字符。（以前 C90 只允许 6 个字符），这是一个很大的进步。旧式编译器通常最多只允许使用 8 个字符）。实际上，你可以使用更长的字符，但是编译器会忽略超出的字符。也就是说，如果有两个标识符名都有 63 个字符，只有一个字符不同，那么编译器会识别这是两个不同的名称。如果两个标识符都是 64 个字符，只有最后一个字符不同，那么编译器可能交其视为同一个名称，也可能不会。标准并未定义在这种情况下会发生什么。

可以用小写字母、大写字母、数字和下划线（——）来命名。而且，名称的第 1 个字符必须是字母或下划线，不能是数字。表 2.1 给出了一些示例。表 2.1 有效和无效的名称

有效的名称	无效的名称	原因
wiggles	\$2]**	首字不为字母
cat2	2cat	
Hot _{Tub}	Hot-Tub	仅可用下划线
taxRate	tax rate	不能用空格
_kcab	don't	' 不能用

大小写字母

数字

下划线

首字不能为数字

C99 及 C11 长度 63 个字符

外部标识符长度为 31 个

C90 只允许 8 个字符

操作系统及 C 库标识符首字符默认下划线

操作系统和 C 库经常使用以一个或两个下划线字符开始的标识符(如, `kcab`), 因此最好避免在自己的程序中使用这种名称。标准标签都以一个或两个下划线字符开始, 如库标识符。这样的标识符都是保留的。这意味着, 虽然使用它们没有语法错误, 但是会导致名称冲突。

C 语言的名称区分大小写, 即把一个字母的大写和小写视为两个不同的字符。因此, `stars` 和 `Stars`、`STARS` 都不同。

为了让 C 语言更加国际化, C99 和 C11 根据通用字符名(即 UCN)机制添加了扩展字符集。其中包函了除英文字母以外的部分字符。欲了解详细内容, 请参阅附录 B 的“参考资料 VII: 扩展字符支持”。

3. 声明变量的四个理由

一些更老的语言(如, `fortran` 和 `basic` 的最初形式)都允许直接使用变量, 不必先声明。为何 c 语言不采用这种简单易行的方法? 原因如下。

把所有变量放在一处, 方便读者查找和理解程序的用途。如果变量名都是有意义的(如, `taxtate` 而不是 `r`), 这样做效果很好。如果变更名无法表述清楚, 在注释中解释变量的含义。这种方法让程序的可读性更高。

声明变量会促使在编写程序之前做一些计划。程序在开始时要获得哪些信息? 希望程序如何输出? 表示数据最好的方式是什么?

声明变量有助于发现隐藏在程序中的小错误, 如变量名拼写错误。例如, 假设在某些不需要声明就可以直接使用变量的语言中, 编写如下语句:

```
RADIUS1=20.R;
```

在后面的程序中, 误写成:

```
CIRCUM=6.28*RADIUSl;
```

你不小心把数字 1 打成小写字母 l。这些语言会创建一个新的变量 RADIUSl，并使用该变量中的值（也许是 0，也许是垃圾值），导致赋给 CIRCUM 的值是错误值。可能要花很长时间才能查出原因。这种错误在 c 语言中不会发生（除非你很不明智地声明了两个极其相似的变量），因为编译器在发现未声明的 RADIUS1 时会报错。

如果事先未声明变量，c 程序将无法通过编译。如果前几个理由还不足以说服你，这个理由总可以让你认真考虑一下了。

如果要声明变量，应该声明在何处？前面提到过，c99 之前的标准要求变量声明都置于块的顶部，这样规定的好处是：把声明放在一起更容易理解程序的用途。c99 允许在需要时才声明变量，这样做的好处是：在给变量赋值之前声明变量，就不会忘记给变量赋值。但是实际上，许多编译器都不支持 c99。

6. 赋值

```
num=1;
```

程序清单中的这行代码是赋值表达式语句。赋值是 c 语言的基本操作之一。该行代码的意思是“反值 1 赋给变量 num”。在执行 `int num;` 声明时，编译器在计算机内存中为变量 num 预留了空间，然后在执行这行赋值表达式语句时，把值储存在之前预留的位置。可以给 num 赋不同的值，这 num 之所以被称为变量（variable）的原因。注意，该赋值表达式语句从右侧把值赋到左侧。另外，该语句以分号结尾，如图 2.2 所示。

其中等号为赋值运算符

7.printf() 函数

```
printf("I am a simple");  
printf("computer.");  
printf("My favorite number is %d because it is first.",num);
```

这三行都使用了 C 语言的一个标准函数：printf()。圆括号表明 printf 是一个函数名。圆括号内的内容是从 main() 函数传递给 printf() 函数的信息。例如，上面的第 1 行把 I am a simple 传递给 printf() 函数。该信息被称为参数，或者更确切地说，是函数的实际参数 (actual argument)，【在 C 语言中，实际参数 (简称实参) 是传递给函数的特定值，形式参数 (简称形参) 是函数中用于储存值的变量。第 5 章中将详述相关内容。】printf() 函数用参数来做什么？该函数会查看双引号中的内容，并将其打印在屏幕上。

第 1 行 printf() 演示了在 C 语言中如何调用函数。只需输入函数名，把所需的参数填入圆括号即可。当程序运行到这一行时，控制权被转给命名的函数 (该例中是 printf())。函数执行结束后，控制权被返回至主调函数 (calling function)，该例中是 main()。

第 2 行 printf() 函数的双斜杠中的字符并未输出。这是为什么？的意思是换行。组合 (依次输入这两个字符) 代表一个换行符 (newline character)。对于 printf() 而言，这的意思是“在下一行的最左加开始新的一行”。也就是说，打印换行符的效果与在键盘按下 Enter 键相同。既然如此，为何不在键入 printf() 参数时直接使用 Enter 键？因为编辑器可能认为这是直接的命令，而不是储存在源代码中的指令。换句话说，如果直接按下 Enter 键，编辑器会退出当前行并开始新的一行。但是，换行符会影响程序输出的显示格式。

换行符是一个转义序列 (escape sequence)。转义序列用于代表难以表示或无法输入的字符。如，`\t` 表 Tab 键，`\b` 代表 Backspace 键 (退格键)。每个转义序列都以反斜杠字符 `\` 开始。我们在第 3 章再来探讨相关内容。

这样，就解释了为什么 3 行 printf() 语句只打印出两行：第 1 个 printf() 打印的内容中不含换行符，但是第 2 和第 3 个 printf() 中都有换行符。

第 3 个 printf() 还有一些不明之处：参数中的 %d 在打印时有什么作用？先来看该函数的输出：

```
My favorite number is 1 because it is first.
```

对比发现，参数中的 %d 被数字 1 代替了，而 1 就是变量 num 的值。%d 相当于是一个占位符，其作用是指明输出 num 值的位置。该行和下面

的 BASIC 语句很像：

```
PRINT"My favorite number is ";num;because it is first."
```

实际上，C 语言的 `printf()` 比 BASIC 的这条语句做的事情多一些。提醒程序，要在该处打印一个变量，`d` 表明把变量作为十进制整数打印。`printf()` 函数名中的 `f` 提醒用户，这是一种格式化打印函数。`printf()` 函数有多种打印变量的格式，包括小数和十六进制整数。后面章节在介绍数据类型时，会详细介绍相关内容。

8.return 语句

```
return 0;
```

`return` 语句是程序清单 2.1 的最后一条语句。`int main(void)` 中的 `int` 表明 `main()` 函数应返回一个整数。C 标准要求 `main()` 这样做。有返回值的 C 函数要有 `return` 语句。该语句以 `return` 关键字开始，后面是待返回的值，并以分号结尾。如果遗漏 `main()` 函数中的 `return` 语句，程序在运行至最外面的右花括号 `}` 时会返回 0。因此，可以省略 `main()` 函数末尾的 `return` 语句。但是，不要在其他有返回值的函数中漏掉它。因此，强烈建议读者养成在 `main()` 函数中保留 `return` 语句的好习惯。在这种情况下，可将其看作是统一代码风格。但对于某些操作系统（包括 Linux 和 UNIX），`return` 语句有实际的用途。第 11 章再详述这个主题。

2.3 简单程序的结构

分析过一个具体的程序示例后，了解一个程序的基本结构。程序由一个或多个函数组成，必须有 `main()` 函数。函数由函数头和函数体组成。函数头包括函数名、传入该函数的信息类型和函数的返回类型。通过函数名后的圆括号可识别出函数，圆括号里可能为空，可能有参数。函数体被花括号括起来，由一系列语句、声明组成，如图 2.4 所示。本章的程序示例中有一条声明，声明了程序使用的变量名和类型。然后是一条赋值表达式语

句，变量被赋给一个值。接下来是 3 条 printf() 语句，调用 printf() 函数 3 次。最后，main() 以 return 语句结束。

```
int main(void) //函数头
{
    int num; //声明
    num=1; //语句
    printf("%d is neat. \n",num); //语句
    return 0; //语句
}
```

图 2.4 函数包含函数头和函数体

总而言之，一个简单的 C 程序的格式如下：

```
#include<stdio.h>
int main(void)
{
    语句
    return 0;
}
```

(大部分语句都以分号结尾。)

2.4 提高程序可读性的技巧

编写可读性高和程序是良好的编程习惯。可读性高的程序更容易理解，以后也更容易修改和更正。抽调程序的可读性还有助于理清编程思路。

前面介绍过两种提高程序可读性的技巧：选择有意义的函数名和写注释。注意，使用这两种技巧时应相得益彰，避免重复啰嗦。如果变量名是 width，就不必写注释说明该变量表示宽度，但是如果变量名是 video_routine4，就要解释一下该变量名的含义。

提高程序可读性的第 3 个技巧是：在函数中用空行分隔概念上的多个部分。例如，程序清单 2.1 中用空行把声明部分和程序的其他部分区分开来。C 语言并未规定一定要使用空行，但是多使用空行能提高程序的可读性。

提高程序可读性的第 4 个技巧是：每条语句各占一行。同样，这也不是 C 语言的要求。C 语言的格式比较自由，可以把多条语句放在一行，也可以每条语句独占一行。下面的语句都没问题，但是不好看：

```
int main(void){int four; four
=
4;
printf(
"%d\n",
four);return 0;}
```

分号告诉编译器一条语句在哪里结束、下一条语句在哪里开始。如果按照本章示例的约定来编写代码（见图 2.5），程序的逻辑会更清晰。

```
int main(void)/*把2音寻（测水深的单位）转换成英尺*/——写注释
{
int feet ,fathoms; //使用有意义的变量名
                //使用空行
fathoms=2;
feet=6*fathoms;  //每行一条语句
printf("There are %d feet in %d fathoms!\n",feet ,fathoms);
return 0;
}
```

图 2.5 提高程序的可读性

2.5 进一步使用 C

本章的第 1 年程序相当简单，下面的程序清单 2.2 也不太难。

程序清单 2.2 fath_m_{ft}.c 程序

```
//fathmft.c ——把2音寻转换成英寸

#include<stdio.h>
int main(void)
{
    int feet ,fathoms;
    fathoms=2;
    feet=6*fathoms;
    printf("There are %d feet in %d fathoms!\n",feet ,fathoms);
    printf("Yes, I said%d feet!\n",6*fathoms);
    return 0;
}
```

与程序清单 2.1 相比，以上代码有什么新内容？这段代码提供了程序描述，声明了多个变量，进行了乘法运算，并打印了两个变量的值。下面更详细地分析这些内容。

2.5.1 程序说明

程序在开始处有一条注释（使用新的注释风格），给出了文件名和程序的目的。写这种程序说明很简单、不费时，而且在以后浏览或打印程序时很有帮助。

2.5.2 多条声明

接下来，程序在一条声明中声明了两个变量，而不是一个变量。为此，要在声明中用逗号隔开两个变量（feet 和 fathoms）。也就是说，

```
int feet,fathoms;
```

和


```
int feet;  
int fathoms;  
等价。
```

2.5.3 乘法

然后，程序进行了乘法运算。利用计算机强大的计算能力来计算 6 乘以 2。C 语言和许多其他语言一样，用 * 表示乘法。因此，语句

```
feet=6*fathoms;
```

的意思是“查找变量 fathoms 的值，用 6 乘以该值，并把计算结果赋给变量 feet”。

2.5.4 打印多个值

最后，程序以新的方式使用 printf() 函数。如果编译并运行该程序，输出应该是这样：

```
There are 12 feet in 2 fathoms!
```

```
Yes, I said 12 feet!
```

程序的第 1 个 printf() 中进行了两次替换。双引号后面的第 1 个变量 (feet) 替换了双引号中的第 1 个 %d；双引号后面的第 2 个变量 (fathoms) 替换了双引号中的第 2 个 %d。注意，待输出的变量列于双引号的后面。还要注意，变量之间要用逗号隔开。

第 2 个 printf() 函数说明待打印的值不一定是变量，只要可求值得出合适类型值的项即可，如 6*fathoms。

该程序涉及的范围有限，但它是把章寻转换成英寸程序的核心部分。还需要把其他值通过交互的方式赋给 feet，其方法将在后面章节中介绍。

2.6 多个函数

目前为止，介绍的几个程序只使用了 `printf()` 函数。程序清单 2.3 演示了除 `main()` 以外，如何把自己的函数加入程序中。

程序清单 2.3 `two_func.c` 程序

```
//two_func.c — 一个文件中包含两个函数
#include<stdio.h>
void butler(void); //ANSI/ISO C函数原型
int main(void)
{
    printf("I will summon the butler function.\n");
    butler();
    printf("Yes. Bring me some tea and writeable DVDs.\n");
    return 0;
}
void butler(void) //函数定义开始
{
    printf("You rang, sir?\n");
}
```

该程序输出如下：

I will summon the butler function.

You rang, sir?

Yes. Bring me some tea and writeable DVDs.

`butler()` 函数在程序中出现了 3 次。第 1 次是函数原型 (prototype)，告知编译器在程序中要使用该函数；第二次以函数调用 (function call) 的形式出现在 `main()` 中；最后一次出现在函数定义 (function definition) 中，函数定义既是函数本身的源代码。下面逐一分析。

C90 标准新增了函数原型，旧式的编译器可能无法识别（稍后将介绍），如果使用这种编译器应该怎么做）。函数原型是一种声明形式，告知编译器

正在使用某函数，因此函数原型也被称为函数声明（function declaration）。函数原型还指明了函数的属性。例如，`butler()` 函数原型中的第 1 个 `void` 表明，`butler()` 函数没有返回值（通常，被调函数会向主调函数返回一个值），但是 `butler()` 函数没有）。第 2 个 `void(butler(void)` 中的 `void)` 的意思是 `butler()` 函数不带参数。因此，当编译器运行至此，会检查 `butler()` 是否使用得当。注意，`void` 在这里的意思是“空的”，而不是“无效”。

早期的 C 语言支持一种更简单的函数声明，只需指定返回类型，不用描述参数：

```
void butler();
```

早期的 C 代码中的函数声明就类似上面这样，还是现在的函数原型。

C90、C99 和 C11 标准都承认旧版本的形式，但是也表明了会逐渐淘汰这种过时的写法。如果使用以前写的 C 代码，就需要把旧式声明转换成函数原型。后面章节会继续介绍函数原型的相关内容。

接下来继续分析程序。在 `main()` 中调用 `butler()` 很简单，写出函数名和圆括号即可。当 `butler()` 执行完毕后，程序会继续执行 `main()` 中的下一条语句。

程序的最后部分是 `butler()` 函数的定义，其形式和 `main()` 相同，都包含函数头和用花括号括起来的函数体。函数头重述了函数原型的信息：`butler()` 不带任何参数，且没有返回值。如果使用老式的编译器，请去掉圆括号中的 `void`。

注意，何时执行 `butler()` 函数取决于它在 `main()` 中被调用的位置，而不是 `butler()` 定义在文件中的位置。例如，把 `butler()` 函数的定义放在 `main()` 定义之前，不会改变程序的执行顺序，`butler()` 函数仍然在两次 `printf()` 调用之间被调用。记住，无论 `main()` 在程序文件处于什么位置，所有的 C 程序都从 `main()` 开始执行。但是，C 的惯例是把 `main()` 放在开头，因为它提供了程序的基本框架。

C 标准建议，要为程序中用到的所有函数提供函数原型。标准 `include` 文件（包含文件）为标准库函数提供可函数原型。例如，在 C 标准中，`stdio.h` 文件包含了 `printf()` 的函数原型。第 6 章最后一个示例演示了如何使用带

返回值的函数，第 9 章交详细全面地介绍函数。

2.7 调试程序

现在，可以编写一个简单的 C 程序，但是可能会犯一些简单的错误。程序的错误通常叫做 bug，找出并修正错误的过程叫做调试 (debug)。程序清单 2.4 是一个有错误的程序，看看你能找出几处。程序清单 2.4 nogood.c 程序

```
//nogood.c —有错误的程序
#include<stdio.h>
int main(void)
(
    int n,int n2, int n3;//该程序有多处错误
    n=5;
    n2=n*n;
    n3=n2*n2;
    printf("n=%d, n squared=%d, n cubed=%d\n",n,n2,n3)
    return 0;
)
```

2.7.1 语法错误

程序清单 2.4 中有多处语法错误。如果不遵循 C 语言的规则就会犯语法错误。这类似于英文中的语法错误。例如，看看这个句子：Bugs frustrate be can. 该句子中的英文单词都是有效的单词（即，拼写正确），但是并未按照正确的顺序组织句子，而且用词也不妥。C 语言的语法错误指的是，把有效的 C 符号放在错误的地方。

nogood.c 程序中有哪些错误？其一，main() 函数体使用圆括号来代替花括号。这就是把 C 符号用错了地方。其二，变量声明应该这样写：

```
int n,n2,n3;
```

或者，这样写：

```
int n;
```

```
int n2;
```

```
int n3;
```

其三，main() 中的注释末尾漏掉了 */ (另一种修改方案是，用//替换，吾即用此)。最后，printf() 语句末尾漏掉了分号。

如何发现程序的语法错误？首先，在编译之前，浏览源代码看是否能发现一些明显的错误。接下来，查看编译器是否发现错误，检查程序的语法错误是它的工作之一。在编译程序时，编译器发现错误会报告错误信息，指出每一处错误的性质和具体位置。

尽管如此，编译器也有出错的时候。也许某处隐藏的语法错误会导致编译器误判。例如，由于 nogood.c 程序未正确声明 n2 和 n3，会导致编译器在使用这些变量时发现更多问题。实际上，有时有用把编译器报告的所有错误逐一修正，仅修正第 1 条或前几处错误后，错误信息就会少很多。继续这样做，直到编译器不再报错。编译器另一个常见的毛病是，报错的位置比真正的错误位置滞后一行。例如，编译器在编译下一行时才会发现上一行缺少分号。因此，如果编译器报错某行缺少分号，请检查上一行。

2.7.2 语义错误

语义错误是指意思上的错误。例如，考虑这个句子：Scornful derivatives sing greenly(轻蔑的衍生物不熟练地唱歌)。句子中的形容词、名词、动词和副词都在正确的位置上，所以语法正确。但是却让人不知所云。在 C 语言中，如果遵循了 C 规则，但是结果不正确，那就是犯了语义错误。程序示例中有这样的错误：

```
n3=n2*n2;
```

此处，n3 原意表示 n 的 3 次方，但是代码中的 n3 被设置成 n 的 4 次方 (n2=n*n)。

编译器无法检测语义错误，因为这类错误并未违反 C 语言的规则。编译器无法了解你的真正意图，所以你能自己找出这些错误。例如，假设你修正了程序的语法错误，程序应该如程序清单 2.5 所示：

程序清单 2.5 stillbad.c 程序

```
//stillbad.c —修复了语法错误的程序
#include<stdio.h>
int main(void)
{
    int n,n2,n3;
    //该训育有一个语义错误
    n=5;
    n2=n*n;
    n3=n2*n2;
    printf("n=%d,n squared=%d, n cubed=%d\n",n,n2,n3);
    return 0;
}
```

该程序的输出如下：

n=5,n squared=25,n cubed=625

如果对简单的立方比较熟悉，就会注意到 625 不对。下一步是跟踪程序的执行步骤，找出程序如何得出这个答案。对于本例，通过查看代码就会发现其中的错误，但是，还应该学习更系统的方法。方法之一是，把自己想象成计算机，跟着程序的步骤一步一步地执行。下面，来试试这种方法。

main() 函数体一开始就声明了 3 个变量：n,n2,n3. 你可以画出 3 个盒子并把变量名写在盒子上来模拟这种情况（见图 2.6）。接下来，程序把 5 赋给变量 n。何以在标签为 n 的盒子里写上 5. 接着，程序把 n 和 n 相乘，并把乘积赋给 n2。因此，查看标签为 n 的盒子，其值是 5，5 乘以 5 得 25，于是把 25 放进标签为 n2 的盒子里。为了模拟下一条语句（n3=n2*n2）。查看 n2 盒子，发现其值是 25。25 乘以 25 得 625，把 625 放进标签为 n3 的盒子。原来如此！程序中计算的是 n2 的平方，不是 n2 乘以 n 得到 n 的

3 次方。

对于上面程序示例，检查程序的过程可能过于繁琐。但是，用这种方法一步一步查看程序的执行情况，通常是发现程序问题所在的良方。

***2.7.3 程序状态

通过逐步跟踪程序的执行步骤，并记录每个变量，便可监视程序的状态。程序状态 (program state) 是在程序的执行过程中，某给定点上所有变量值的集合。它是计算机头前状态的一个快照。

刚刚讨论了一种跟踪程序状态的方法：自己模拟计算机逐步执行程序。但是，如果程序中有 10000 次循环，这种方法恐怕行不通。不过，可以跟踪一小部分循环，看看程序是否按照预期的方式执行。另外，还要考虑一种情况：很可能按照自己所想去执行程序，而不是根据实际写出来的代码去执行。因此，要尽量忠实代码来模拟。

定位主义错误的另一种方法是：在程序中的关键点插入窗外的 `printf()` 语句，以监视制定变量值的变化。通过查看值的变化可以了解程序的执行情况。对程序的执行满意后，便可删除额外的 `printf()` 语句，然后重新编译。

检测程序状态的第 3 种方法是使用调试器。调试器 (debugger) 是一种程序，让你一步一步运行另一个程序，并检查该唾弃变量的值。调试器有不同的使用难度和复杂度。较高级的调试器会显示正在执行的源代码行号。这在检查有多条执行路径的程序时很方便，因为很容易知道正在执行哪条路径。如果你的编译器自带调试器，现在可以花点时间学会怎么使用它。例如，试着调试一下程序清单 2.4。

2.8 关键字和保留标识符

关键字是 C 语言的词汇。它们以 C 而言比较特殊，不能用它们作为标识符 (如，变量名)。许多关键字用于指定不同的类型，如 `int`。还有一些关键字 (如，`if`) 用于控制程序中语句的执行顺序。在表 2.2 中所列的 C 语言关键字中，粗体表示是 C90 标准新增的关键字，斜体表示 C99 标准新增

的关键字，粗斜体表示的是 C11 标准新增的关键字。表 2.2 ISO C 关键字

auto	extern	short	while
break	float	signed	<i>__Alignas</i>
case	fort	sizeof	<i>__Alignof</i>
char	goto	static	<i>__Atomic</i>
const	if	struct	<i>__Bool</i>
continue	<i>inline</i>	switch	<i>__Complex</i>
default	int	typedef	<i>__Generic</i>
do	long	union	<i>__Imaginary</i>
double	register	unsigned	<i>__Noreturn</i>
else	restrict	void	<i>__Static_assert</i>
enum	return	volatile	<i>__Thread_local</i>

如果使用关键字不当（如，用关键字作为变量名），编译器会将其视为语法错误。还有一些保留标识符（reserved identifier），C 语言已经指定了它们的用途或保留它们的使用权，如果你使用这些标识符来表示其他意思会导致一些问题。因此，尽管它们也是有效的名称，不会引起语法错误，也不能随便使用。保留标识符包括那些以下划线字符开头的标识符和标准库函数名，如 printf()。

2.9 关键概念

编程是一件富有挑战性的事情。程序员要具备抽象和逻辑的思维，并谨慎地处理细节问题（编译器会强迫你注意细节问题）。平时和朋友交流时，可能用错几个单词，犯一两个语法错误，或者说几句不完整的句子，但是对方能明白你想说什么。而编译器不允许这样，对它而言，几乎正确仍然是错误。

编译器不会在下面擀到的概念性问题上帮助你。因此，在这一章中介绍一些关键概念帮助读者弥补这部分的内容。

在本章中，读者的目标应该是理解什么是 C 程序。可以把程序看作是你希望计算机如何完成任务的描述。编译器负责处理一些细节工作，例如把你要计算机完成的任务转换成底层的机器语言（如果从量化方面来解释编译器所做的工作，它可以把 1KB 的源文件创建成 60KB 的可执行文件；即使是一个很简单的 C 程序民要用大量的机器语言来表示）。由于编译器不具有真正的智能，所以你必须用编译器能理解的术语表达你的意图，这些术语就是 C 语言标准规定的形式规则（尽管有些约束，但总比直接用机器语言方便得多）。

编译器希望接收到特定格式的指令，在本章已经介绍过。作为程序员的任务是，在符合 C 标准的编译器框架中，表达你希望程序应该如何完成任务的想法。

2.10 本章小结

C 程序由一个或多个 C 函数组成。每个 C 程序必须包含一个 `main()` 函数，这是 C 程序要调用的第 1 个函数。简单的函数由函数头和后面的一对花括号组成，花括号中是由声明、语句组成的函数体。

在 C 语言中，大部分语句都以分号结尾。声明为变量创建变量名和标识该变量中储存的数据类型。变量名是一种标识符。赋值表达式语句把赋值给变量，或者更一般地说，把值赋给存储空间。函数表达式语句用于调用指定的已命名函数。调用函数执行完毕后，程序会返回到函数调用后面的语句继续执行。

`printf()` 函数用于输出想要表达的内容和变量的值。

一门语言的语法是一套规则，用于管理语言中各有效语句组合在一直的方式。语句的主义是语句要表达的意思。编译器可以检测出语法错误，但是程序里的主义错误只有在编译完之后才能从程序的行为中表现出来。检查程序是否有主义错误要跟踪程序的状态，即程序每执行一步后所有变量的值。

最后，关键字是 C 语言的词汇。

2.11 复习题

复习题的参考答案在附录 A 中。

2.11.1 C 语言的基本模块是什么？（函数）

2.11.2 什么是语法错误？写出一个英语例子和 C 语言例子。

不遵循 C 语言的规则，把有效的 C 语言符号放在错误的地方。Bugs frustrate be can. Bugs can be frustrate. `int a;b;c;` correct: `int a,b,c;` or `int a; int b; int c;`

3. 什么是语义错误？

2.11.3 写出一个英语例子和 C 语言例子。

遵循了 C 语言的规则，但结果不正确，即为语义错误。如：A tiger is a strong man. 求立方：//语义错误作业

```
#include <stdio.h>
int main(int argc, const char*argv[])
{
    int n, n2, n3;
    n=5;
    n2=n*n;
    n3=n2*n2;
    printf("n 为%d, n的平方是: %d, n的立方是: %d\n", n, n2, n3);
}
```

2.11.4 Indiana Sloth 编写了下面的程序，并征求你的意见。请帮助他评定。

```
include studio.h
//correct:
#include<stdio.h>
int main{void}/*该程序打印一年有多少周/*
int main(void)//
(
{//correct
int s
int s;//correct
s:=56;
s=56;//correct
print(There are s weeks in a year.);
printf("There are %d weeks in a year.\n",s);//correct
return 0;

)
}//correct
```

10 多处错误

2.11.5 假设下面 4 例子都是完整程序中的一部分，它们都输出什么结果？

- a. printf("Baa Baa Black Sheep.");

printf("Have you any wool?");

Baa Baa Black Sheep.Have you any wool?

b.printf("Begone!creature of lard!"); Begone! O creature of lard!

```
C.printf("What?\nfish?"); What? No\nfish?
```

```
d. int num; num=2; printf("%d+%d=%d",num,num,num+num); 2+2=4
```

2.11.6 在 main、int、function、char、= 中，哪些是 C 语言的关键字？

int,char

2.11.7 如何以下面的格式输出变量 words 和 lines 的值 (这里，3020 和 350 代表两个变量的值)？

There were 3020 words and 350 lines.

```
printf("There are %d words and %d",words,lines);
```

2.11.8 考虑下面的程序：

```
#include<stdio.h>
int main(void)
{
    int a,b;
    a=5;
    b=2; //第7行
    b=a; //第8行
    a=b; //第9行
    printf( "%d %d\n",b,a);
    return 0;
}
```

请问，在执行完第 7、第 8、第 9 行后，程序的状态分别是什么？

执行第 7 行 a=5 b=2

执行第 8 行: $a=5$ $b=5$

执行第 9 行 $b=5$ $a=5$

a 与 b 相等皆为 5.

2.11.9 考虑下面的程序:

```
#include<stdio.h>
int main(void)
{
    int x,y;
    x=10;
    y=5; //第7行
    y=x+y; //第8行
    x=x*y; //第9行
    printf("%d %d\n",x,y);
    return 0;
}
```

请问, 在执行完第 7、8、9 行后, 程序的状态分别是什么?

第 7 行 $x=10$ $y=5$

第 8 行 $x=10$ $y=15$

第 9 行 $y=15$ $x=150$

2.12 编程练习

纸上得来终觉浅, 绝知此事要躬行。读者应该试着编写一两个简单程序, 体会一下编写程序是否和阅读本章介绍的这样轻松。题目中会给出一些建议, 但是应该尽量自己思考这些问题。一些编程答案练习的答案可在出版商网站获取。

2.12.1 1.

编写一个程序，调用一次 printf() 函数，把你的姓名打印在一行。再调用一次 printf() 函数，把你的姓名打印在一行。输出应如下所示（当然要把示例的内容换成你的姓名）

```
#include<stdio.h>
int main(int argc,const char*argv[])
{
    printf("Xia le zhai\n");
    printf("Xia\n");
    printf("le zhai\n");
    printf("Xia le zhai\n");
    return 0;
}
```

2.12.2 2

编写一个程序，打印你的姓名和地址。

```
#include<stdio.h>
int main(int argc,const char*argv[])
{
    printf("My name is: Xia le zhai\n");
    printf("My address :Taian Road 281#\n");
    return 0;
}
```

2.12.3 3

编写一个程序把你的年龄转换成天数，并显示这两个值。这里不用考虑闰年的问题。

```
#include<stdio.h>
int main(int argc,const char*argv[])
{
    int age,days;
    age=55;
    days=age*365;
    printf("Your age is %d,and you have spend %d days!\n",age,days);
    return 0;
}
```

2.12.4 4

编写一个程序，生成以下输出：For he's a jolly good fellow!

For he's a jolly good fellow!

For he's a jolly good fellow!

Which nobody can deny!

除了 main() 函数以外，该程序还要调用两个自定义函数：一个名为 jolly(), 用于打印前 3 条消息，调用一次打印一条；另一个函数名为 deny(), 打印最后一条消息。

```
#include<stdio.h>
void jolly(void); //声明自定义函数
void deny(void);
int main(int argc,const char*argv[])
{
    jolly();
    jolly();
    jolly();
    deny();
    return 0;
}
```

```
}  
void jolly(void)//函数定义开始  
{  
    printf("For he's a jolly good fellow!\n");  
}  
  
void deny(void)  
{  
    printf("Which nobody can deny!\n");  
}
```

2.12.5 5

编写一个程序，生成以下输出：

Brazil, Russia, India, China

India, China,

Brazil, Russia

除了 main() 以外，该程序还要调用两个自定义函数：一个名为 br()，调用一次打印一次“Brazil, Russia”；另一个名为 ic()，调用一次打印一次“India, China”。其他内容在 main() 函数中完成。

```
#include<stdio.h>  
void br(void);//声明函数  
void ic(void);  
int main(int argc,const char*argv[])  
{  
    br();  
    printf(",");  
    ic;  
    printf("\n");  
}
```



```
ic();
printf( "\n" );
br();
printf( "\n" );
return 0;
}

void br(void)
{
printf( "Brazil , Russia" );
}

void ic(void)
{
printf( "India , China" );
}
```

2.12.6 6

编写一个程序，创建一个整型变量 `toes`，并将 `toes` 设置为 10。程序中还要计算 `toes` 的两倍和 `toes` 的平方。该程序应打印 3 个值，并分别描述以示区分。

```
#include<stdio.h>
int main(int argc, const char*argv[])
{
int toes;
toes=10;
printf( "Toes is %d, it's double number is:%d, it's squared is %d.\n",toes,2*toes,toes*toes);
return 0;
}
```

2.12.7 7

许多研究表明，微笑益处多多。编写一个程序，生成以下格式的输出：

Smile!Smile!Smile! Smile!Smile! Smile!

该程序要定义一个函数，该函数被调用一次打印一次“Smile!”，根据程序的需要使用该函数。

```
#include<stdio.h>
void smile(void);
int main(int argc, const char*argv[])
{
    smile();
    smile();
    smile();
    printf( "\n" );
    smile();
    smile();
    printf( "\n" );
    smile();
    printf( "\n" );
    return 0;
}
void smile(void)
{
    printf( "Smile!" );
}
```

2.12.8 8

在 C 语言中，函数可以调用另一个函数。编写一个程序，调用一个名为 `onethree()` 的函数。该函数在一行打印单词“one”，再调用第 2 个函数 `two()`，

然后在另一行打印单词“three”。two() 函数在一行显示单词“two”。main() 函数在调用 one_three() 函数前要打印短语“starting now:”，并在调用完毕后显示短语“done!”。因此，该程序的输出应如下所示：

starting now:

one

two

three

done!

```
#include<stdio.h>
void one_three(void); // 声明函数
void two(void);
int main(int argc, const char*argv[])
{
    printf("starting now:\n");
    one_three();
    printf("done!\n");
    return 0;
}
void two(void)
{
    printf("two\n");
}
void one_three(void)
{
    printf("one\n");
    two();
    printf("three\n");
}
```


第三章 数据和 C

本章介绍以下内容：

关键字：int short long unsigned char float double `_Bool` `_Complex`
`_Imaginary`

运算符：sizeof()

函数：scanf()

整数类型的浮点型常数，如何声明这些类型的变量

如何使用 printf() 和 scanf() 函数读写不同类型的值。

程序离不开数据。把数字、字母和文字输入计算机，利用这些数据完成某些任务。如，需要计算一份利息或显示一份葡萄酒商的排序列表。本章除了介绍如何读取数据外，还将教会读者如何操控数据。

C 语言提供两大系列的多种数据类型。本章详细介绍两大数据类型：整数类型和浮点类型，讲解这些数据是什么、如何声明它们、如何以及何时使用它们。除此之外，还将介绍常量和变量的区别。读者很快就能看到第 1 个交互式程序。

3.1 示例程序

本章仍从一个简单的程序开始。如果发现有不熟悉的内容，别担心，稍后会详细解释。该程序的意图比较明了，请着编译并运行程序清单 3.1 中的源代码。为了节省时间，在输入源代码时省略注释。

程序清单 3.1platinum.c 程序

```
// platinum.c — your weight in platinum
#include<stdio.h>
int main(int argc, const char*argv[])
{
    float weight; //你的体重
    float value; //相等重量的白金价值
    printf("Are you worth your weight in platinum?\n");
    printf("Let's check it out.\n");
    printf("Please enter your weight in pounds: ");
    //获取用户的输入
    scanf("%f",&weight);
    //假设白金的价格是每盎司转换为金衡盎司
    value=1700.0*weight*14.5833;
    printf("Your weight in platinum is worth $%.2f.\n",value);
    printf("You are easily worth that! If platinum prices drop,\n");
    printf("eat more to maintain your value.\n");
    return 0;
}
```

提示错误与警告

如果输入程序时打错（如漏了一个分号），编译器会报告语法错误消息。然而，即使输入正确无误，编译器也可能给出一些警告，如“警告：从 double 类型转换成 float 类型可能会丢失数据”。错误消息表明程序中有错，不能进行编译。而警告则表明，尽管编写的代码有效，但可能还是程序员想要的。警告并不终止编译。特殊的警告与 C 如何处理 1700.0 这样的值有关。本例不必理会这个问题，本章稍后会进一步说明。

输入该程序时，可以把 1700.0 改成贵金属白金当前的市价，但是还要改动 14.5833，该数是 1 英镑的金衡盎司数（金衡盎司用于衡量贵金属，而英镑常衡盎司用于衡量人的体重）。

注意，“enter your weight”的意思是输入你的体重，然后按下 Enter 或

Return 键（还要键入体重后就一直等着）。按下 Enter 键是告知计算机，你已完成输入数据。该唾弃需要你输入一个数字（如 155），而不是单词（如，too mush）。如果输入字母而不是数字，会导致程序出问题。这个问题要用 if 语句来解决（详见第 7 章），因此请先输入数字。下面是程序的输出示例：

```
Are you worth your weight in platinum?  
Let's check it out.  
Please enter your weight in pounds: 156  
Your weight in platinum is worth $3867491.25.  
You are easily worth that! If platinum preces drop,  
eat more to maintain your value.
```

程序调整

即使用第 2 章介绍的方法，在程序中添加下面一行代码：

```
getchar ();
```

程序的输出是否依旧在屏幕上一闪而过？本例需要调用两次 `getchar ()` 函数：

```
getchar (); getchar ();
```

`getchar ()` 函数读取下一个输入字符，因此程序会等待用户输入。在这种情况下，键入 156 并按下 Enter（或 Return 键）（发送一个换行符），然后 `scanf ()` 读取键入的数字，第 1 个 `getchar ()` 读取换行符，第 2 个 `getchar ()` 让程序暂停，等待输入。

3.1.1 程序中的新元素

程序清单 3.1 中包含 C 语言的一些新元素。

注意，代码中使用了一种新的变量声明。前面的例子只使用了整数类型的变量（`int`），但是本使用了浮点数类型（`float`）的变量，以便处理更大范围的数据。`float` 类型可以储存带小数的数字。

程序中演示了常量的几种新写法。现在可以使用带小数点的数了。

为了打印新类型的变量，在 `printf ()` 中使用 `%f` 来处理浮点值。`%.2f` 中的 `.2` 用于精确控制输出，指定输出的浮点数只显示小数点后面两位。

`scanf()` 函数用于读取键盘的输入。`%f` 说明 `scanf()` 要读取用户从键盘输入的浮点数, `&weight` 告诉 `scanf()` 把输入的值给名为 `weight` 的变量。`scanf()` 函数使用 `&` 符号表明找到 `weight` 变量的地点。下一章将详细讨论 `&`。就目前而言, 请按照这样写。

也许本程序最突出的新特点是它的交互性。计算机向用户询问信息, 然后用户输入数字。与交互式程序相比, 交互式程序用起来更有趣。更重要的是, 交互式使得程序更加灵活。例如, 示例程序可以使用任何合理的体重, 而不只是 156 磅。不必重写程序, 就可以根据不同体重进行计算。`scanf()` 和 `printf()` 函数用于实现这种交互。`scanf()` 函数读取用户从键盘输入的数据, 并把两个函数结合起来, 就可以建立人机双向的通信 (见图 3.1) 这让使用计算机更加有趣味。

本章着重解释上述新特性中的前两项: 各种数据类型的变量和常量。第 4 章将介绍后 3 项。

3.2 变量与常量数据

在程序的指导下, 计算机可以做许多事情, 如数值计算、名字排序、执行语言或视频命令、计算彗星轨道、准备邮件列表、拨电话号码、画画、做决策或其他你能所能想到的事情。要完成这些任务, 程序需要使用数据, 即承载信息的数字和字符。有些数据类型在程序使用之前已经预告设定好了, 在整个程序的运行过程中没有变化, 这些称为常量 (constant)。其他数据类型在程序运行期间可能会改变或被赋值, 这些称为变量 (variable)。在示例程序中, `weight` 是一个变量, 14.5833 是一个常量。那么, 1700.0 是常量还是变量? 在现实生活中, 白金的价格不会是常量, 但是在程序中, 像 1700.0 这样的价格被视为常量。

3.2.1 常量 (constant)

数据类型在程序使用之前预先设定好了, 在整个程序运行过程中没有变化, 这些被称为常量 (constant)。

3.2.2 变量 (variable)

数据类型在程序运行期间会改变或被赋值，其值是变化的，称为变量 (variable)。

3.3 数据：数据类型关键字

不仅变量和常量不同，不同的数据类型之间也有差异。一些数据类型表示数字，一些数据类型表示字母（更普遍地说是字符）。C 通过识别一些基本的数据类型来区分和使用这些不同的数据类型。如果数据是常量，编译器一般通过用户书写的形式来识别类型（如，42 是整数，42.100 是浮点数）。但是，对变量而言，要在声明时指定其类型。稍后会详细介绍如何声明变量。现在，先来了解一下 C 语言的基本类型关键字。K&C 给出了 7 个与类型相关的关键字。C90 标准添加了 2 个关键字，C99 标准又添加了 3 个关键字（见表 3.1）。

表 3.1 C 语言的数据类型关键字

最初 K&R 给出的关键字	C90 标准添加的关键字	C99 标准添加的关键字
int	signed	_Bool
long	void	_Complex
short		
unsigned		
char		
float		
double		

在 C 语言中，用 int 关键字来表示基本的整数类型。后 3 个关键字 (long、short 和 unsigned) 和 C90 新增的 signed 用于提供基本整数类型的变式，例如 unsigned short int 和 long long int。char 关键字用于指定字母和其他字符（如，#、\$、% 和 *）。另外，char 类型也可以表示较小的整

数。float、double 和 long double 表示带小数点的数。Bool 类型表示布尔值 (true 或 false)，Complex 和 Imaginary 分别表示得数和虚数。

通过这些关键字创建的类型，按计算机的储存方式可分为两大基本类型：整数类型和浮点数类型。

3.3.1 位、字节和字

位、字节和字是描述计算机数据单元或存储单元的术语。这里主要指存储单元。

最小的存储单元是位 (bit)，可以储存 0 或 1 (或者说，位用于设置“开”或“关”)。虽然 1 位储存的信息有限，但是计算机中位的数量十分庞大。位是计算机内存的基本构建块。

字节 (byte) 是常用的计算机存储单位。

对于几乎所有的机器，1 字节均为 8 位。这是字节的标准定义，至少在衡量存储单位时是这样 (但是，C 语言对此有不同的定义，请参阅本章 3.4.3 节)。既然 1 位可以表示 0 或 1，那么 8 位字节就有 256 (2 的 8 次方) 种可能的 0、1 的组合。通过二进制编码 (仅用 0 和 1 便可表示数字)，便可表示 0-255 的整数或一组字符 (第 15 章将详细讨论二进制编码，如果感兴趣可以现在浏览一下该章的内容)。

字 (word) 是设计计算机时给定的自然存储单位。

对于 8 位的微型计算机 (如，最初的苹果机)，1 个字长只有 8 位，从那以后，个人计算机字长增至 16 位、32 位，直到目前的 64 位。计算机的字长越大，其数据转移越快，允许的内存访问也更多。

3.3.2 整数和浮点数

整数类型？浮点数类型？如果觉得这些术语非常陌生，别担心，下面先简述它们的含义。如果不熟悉位、字节和字的概念，请阅读上面方框中的内容。刚开始学习时，不必了解所有的细节，就像学习开车之前不必详细了解汽车内部引擎的原理一样。但是，了解一些计算机或汽车引擎内部的原理会对你有所帮助。

对我们而言，整数和浮点数的区别是它们的书写方式不同。对计算机而言，它们的区别是储存方式不同。下面详细介绍整数和浮点数。

3.3.3 整数

和数学的概念一样，在 C 语言中，整数是没有小数部分的数。例如，2、-23 和 2456 都是整数。而 3.14、0.22 和 2.000 都不是整数。计算机以二进制数字储存整数，例如，整数 7 以二进制写成 111。因此，要在 8 位字节中储存该数字，需要把前 5 位都置成 0，后 3 位设置成 1（如图 3.2 所示）。

0	0	0	0	0	1	1	1
---	---	---	---	---	---	---	---

 字长 8 位

$$2^2 \ 2^1 \ 2^0$$

$$4 + 2 + 1 = 7 \text{ 整数 } 7 \text{ 图 3.2 使用二进制编码储存整数 } 7$$

3.3.4 浮点数

浮点数与数学中的实数的概念差不多。2.75、3.16E7、7.00 和 2e-8 都是浮点数。注意，在一个值的后面加上一个小数点，该值就成为一个浮点值。所以，7 是整数，7.00 是浮点数。显然，书写浮点数有多种形式。稍后将详细介绍 e 记数法，这里先做简要介绍：3.16E7 表示 3.16×10^7 （3.16 乘以 10 的 7 次方）。其中， $10^7=10000000$ ，7 被称为 10 的指数。

这里关键要理解浮点数和整数的储存方案不同。计算机把浮点数分成小数部分和指数部分来表示，而且分开储存这两部分。因此，虽然 7.00 和

7 在数值上相同，但是它们的存在方式不同。在十进制下，可以把 7.0 写成 0.7E1。这里，0.7 是小数部分，1 是指数部分。图 3.3 演示了一个储存浮点数的例子。当然，计算机在内部使用二进制和 2 的幂进行储存，而不是 10 的幂。第 15 章将详述相关内容。现在，我们着重讲解这两种类型的实际区别。

整数没有小数部分，浮点数有小数部分。

浮点数可以表示范围比整数大。参见本章末的表 3.3。

对于一些自述运算（如两个很大的数相减），浮点数损失的精度更多。

+	.314159	1
---	---------	---

符号小数指数

• .314159 (E10¹ 3.14159

图 3.3 以浮点格式（十进制）储存 π 的值

因为在任何区间内（如，1.0 到 2.0 之间）都存在无穷多个实数，所以计算机的浮点数不能表示区间内所有的值。浮点数通常只是实际值的近似值。例如，7.0 可能被储存为浮点值 6.9999。稍后会讨论更多精度方面的内容。

过去，浮点运算比整数运算慢。不过，现在许多 CPU 都包含浮点处理器，缩小了速度上的差距。

表 3.1: 这里是表头

+	.314159	1
---	---------	---

3.4 C 语言基本数据类型

本节将详细介绍 C 语言的基本数据类型，包括如何声明变量、如何表示字面值常量（如，5 或 2.78），以及典型的用法。一些老式的 C 语言编译

表 3.2: 图 3.3 以浮点格式（十进制）储存

	x1	x2	x3	x4
A	3	4	2	1
B	2	5	4	3
C	5	4	3	1
D	3	3	2	4

器无法支持这里提到的所有类型，请查阅你使用的编译器文档，了解可以使用哪些类型。

3.4.1 int 类型

C 语言提供了许多整数类型，为什么一种类型不够用？因为 C 语言让程序员针对不同情况选择不同的类型。特别是，C 语言中的整数类型可表示不同的取值范围和正负值。一般情况使用 `int` 类型即可，但是为满足特定任务和机器的要求，还可以选择其他类型。

`int` 类型是有符号整型，即 `int` 类型的值必须是整数，可以是正整数、负整数或零。其取值范围依计算机系统而异。一般而言，储存一个 `int` 要占一个机器字长。因此，早期的 16 位 IBM PC 兼容机使用 16 位来储存一个 `int` 值，其取值范围（即 `int` 值的取值范围）是 -32768~32767。目前的个人计算机一般是 32 位，因此用 32 位储存一个 `int` 值。现在，个人计算机产业正逐步向着 64 位处理器发展，自然能储存更大的整数。ISO C 规定 `int` 的取值范围最小为 -32768~32767。一般而言，系统用一个特殊位的值表示有符号整数的正负号。第 15 章将介绍常用的方法。

声明 `int` 变量

第 2 章中已经用 `int` 声明过基本整形变量。先写上 `int`，然后写变量名，最后加上一个分号。要声明多个变量，可以单独声明每个变量，也可以在 `int` 后面列出多个变量名，变量名之间用逗号分隔。下面都是有效的声明：

```
int erns;  
int hogs, cows, goats;
```

可以分别在 4 条声明中声明各变量，也可以在一条声明中声明 4 个变量。两种方法的效果相同，都为 4 个 int 大小的变量赋予名称并分配内存空间。

以上声明创建了变量，但是并没有给它们提供值。变量如何获得值？前面介绍过在程序中获取值的两种途径。第 1 种途径是赋值：

```
cows=112;
```

第 2 种途径是，通过函数（如，scanf ()）获得值。接下来，我们着重介绍第 3 种途径。

初始化变量

初始化（initialize）变量就是为变量赋一个初始值。在 C 语言中，初始化可以直接在声明中完成。只需在变量名后面加上赋值运算符（=）和待赋给变量的值即可。如下所示：

```
int hogs=21;  
int cows=32, goats=14;  
int dogs, cats=94; //有效，但是这种格式很糟糕
```

以上示例的最后一行，只初始化了 cats，并未初始化 dogs。这种写法很容易让人误认为 dogs 也被初始化为 94，所以最好还要把初始化的变量和未初始化的变量放在同一条声明中。

简而言之，声明为变量创建和标记存储空间，并为其指定初始值（如图 3.4 所示）。

```
int sows; |||  
创建内存空间  
int boars=2; || 2 || boars 创建内存空间并为其赋值
```

图 3.4 定义并初始化变量

int 类型常量

上面示例中出现的整数（21、32、14 和 94）都是整形常量或整形字面量。C 语言把不含小数点和指数的数作为整数。因此，22 和-44 都是整形常量，但是 22.0 和 2.2E1 则不是。C 语言把大多数整形常量视为 int 类型，但是非常大的整数除外。详见后面“long”常量和 long long 常量”小节对 long int 类型的讨论。

打印 int 值

可以使用 printf () 函数打印 int 类型的值。第 2 章介绍过，%d 指明了在一行中打印整数的位置。%d 称为转换说明，它指定了 printf () 应使用什么格式来显示一个值。格式化字符串中的每个%d 都待打印变量列表中相应的 int 值匹配。这个值可以是 int 类型的变量、int 类型的常量与待打印值的数量相同，编译器不会捕获这类型的错误。程序清单 3.2 演示了一个简单的程序，程序中初始化了一个变量，并打印该变量的值、一个常量值和一个简单表达式的值。另外，程序还演示了如果粗心犯错会导致什么结果。

程序清单 3.2 print1.c 程序

```
//PRINT1.C 演示 printf () 的一些特性
#include<stdio.h>
int main(void)
{
    int ten=10;
    int two=2;
    printf("Doing it right: ");
    printf("%d minus %d\ is %d\n",ten,2,ten-two);
    printf("Doing it wrong: ");
    printf("%d minus %d is %d\n",ten); //遗漏2个参数
    return 0;
```

```
}
```

编译并运行该程序，输出如下：

```
Doing it right: 10 minus 2 is 8
```

```
Doing it wrong: 10 minus 16 is 1650287143
```

在第一行输出中，第 1 个%d 对应 int 类型变量 `ten`；第 2 个%d 对应 int 类型常量 2；第 3 个%d 对应 int 类型表达式 `ten-two` 的值。在第二行输出中，第 1 个%d 对应 `ten` 的值，但是由于没有给后面两个%d 提供任何值，所以打印出的值是内存中的任意值（读者在运行该程序时显示的这两个数值会与输出示例中的数值不同），因为内存中储存的数据不同，而且编译器管理内存的位置也不同。

你可能会抱怨编译器为何不能捕获这种明显的错误，但实际上问题出在 `printf()` 不寻常的设计。大部分函数都需要指定数目的参数，编译器会检查参数的数目是否正确。但是，`printf()` 函数的参数数目不定，可以有 1 个、2 个、3 个或更多，编译器也爱莫能助。记住，使用 `printf()` 函数时，要确保转换说明的数量与打印值的数量相等。

八进制和十六进制

通常，C 语言都是假定整数型常量是十六进制数。然而，许多程序员很喜欢使用八进制和十六进制数。因为 8 和 16 都是 2 的幂，而 10 却不是。显然，八进制和十六进制记数系统在表达与计算机 `dhudr` 值时很方便。例如，二进制数 65536 经常出现在 16 位机中，用十六进制表示正好是 10000。另外，十六进制数的每一位的数恰好由 4 位二进制数表示。例如，十六进制数 3 是 0011，十六进制数 5 是 0101。因此，十六进制数 35 的位组合 (bit pattern) 是 00110101，十六进制数 53 的位组合是 01010011。这种对应关系使得十六进制和二进制的转换非常方便。但是，计算机如何知道 10000 是十进制、十六进制还是二进制？在 C 语言中，用特定的前缀表示使用哪种进制。0x 或 0X 前缀表示十六进制值。所以十进制数 16 表示成十六进制是 0x10 或 0X10。与此类似，o 前缀表示八进制。例如，十进制数 16 表示成八进制是 o20。第 15 章将更全面地介绍进制相关的内容。

要清楚，使用不同的进制数是为了方便，不会影响数被储存的方式。也就是说，无论把数字写成 16、020 或 0x10，储存该数的方式都相同。因为计算机内部都以二进制进行编码。

显示八进制和十六进制

在 C 语序中，既可以使用和显示不同进制的数。不同的进制要使用不同的转换说明。以十进制显示数字，使用 %d；以八进制显示数字，使用 %o；以十六进制显示数字，使用 %X。另外，要显示各进制的前缀 o、ox 和 oX，必须分别使用 %#o、%#x、%#X。程序清单 3.3 演示了一个小程序。回忆一下，在某些集成开发环境（IDE）下编写的代码中插入 getchar ()；语句，程序在执行完毕后不会立即关闭执行窗口。

程序清单 3.3 bases.c 程序

```
//bases.c 以十进制、八进制、十六进制打印十进制数100
#include<stdio.h>
int main(int argc, const char*argv[])
{
    int x=100;
    printf("dec=%d; octal=%o; hex=%x\n", x, x, x);
    printf("dec=%d; octal=%#o; hex=%#x\n", x, x, x);
    return 0;
}
```

编译并运行该程序，输出如下：dec=100;octal=144;hex=64 dec=100;octal=0144;hex=ox64

该程序以 3 种不同记数系统显示同一个值。printf () 函数做了相应的转换。注意，如果要在八进制和十六进制前显示 o 和 ox 前缀，要分别在转换说明中加入 #。

3.4.2 其他整数类型

初学 C 语言时，int 类型应该能满足大多数程序的整数类型需求。尽管如此，还应了解一下整形的其他形式。当然，也可以略过本章跳至 3.4.3 节阅读 char 类型的相关内容，以后有需要时再阅读本节。

C 语言提供了 3 个附属关键字修饰基本整数类型：short、long 和 unsigned。应记住以下几点。

short int 类型（或者简写为 short）占用的存储空间可能比 int 类型少，常用于较小数值的场合以节省空间。与 int 类似，short 是有符号类型。

long int 或 long 占用的存储空间可能比 int 多，适用于较大数值的场合。与 int 类似，long 是有符号类型。

long int 或 long 占用的存储空间可能比 int 多，适用于较大数值场合。与 int 类似，long 是有符号类型。

long long int 或 long long（C99 标准加入）占用的存储空间可能比 long 多，适用于更大数值的场合。该类型至少占 64 位。与 int 类似，long long 是有符号类型。

unsigned int 或 unsigned 只用于非负值的场合。这种类型与有符号类型表示的范围不同。例如，16 位 unsigned int 允许的取值范围是 0~65535，而不是-32768~32767。用于表示正负号的位现在用于表示另一个二进制位，所以无符号整形可以表示更大的数。

在 C90 标准中，添加了 unsigned long int 或 unsigned long 和 unsigned int 或 unsigned short 类型。C99 标准还添加了 unsigned long long int 或 unsigned long long。

在任何有符号类型前添加关键字 signed，可强调使用有符号类型的意图。例如，short、short int、signed short、signed short int 都表示同一种类型。

声明其他整数类型

其他整数类型的声明方式与 `int` 类型相同，下面列出了一些例子。还是所有的 C 编译器都能识别最后 3 条声明，最后一个例子所有的类型是 C99 标准新增的。

```
long int estine;  
long johns;  
short ribs;  
unsigned int s_count;  
unsigned players;  
unsigned long headcount;  
unsigned short yearvotes;  
long long ago;
```

使用多种整数类型的原因

为什么说 `short` 类型“可能”比 `int` 类型战胜的空间少，`long` 类型“可能”比 `int` 类型占用的空间多？因为 C 语言只规定了 `short` 占用的存储空间不能多于 `int`，`long` 占用的存储空间不能少于 `int`。这样规定是为了适应不同的机器。例如，过去的一台运行 Windows 3 的机器上，`int` 类型和 `short` 类型都占 16 位，`long` 类型占 32 位。后来，Windows 和苹果系统都使用 16 位储存 `short` 类型，32 位储存 `int` 类型和 `long` 类型（使用 32 位可以表示的整数数值超过 20 亿）。现在，计算机普遍使用 64 位的整数，才引入了 `long long` 类型。

现在，个人计算机上最常见的设置是，`long long` 占 64 位，`long` 占 32 位，`short` 占 16 位或 32 位（依计算机的自然字长而定）。原则上，这 4 种类型代表 4 种不同的大小，但是在实际使用中，有些类型之间通常有重叠。

C 标准对基本数据类型只规定了允许的最小大小。对于 16 位机，`short` 和 `int` 的最小取值范围是 **【-32767, 32767】**；对于 32 位机，`long` 的最小取值范围是 **【-214783647, 214783647】**。对于 `unsigned short` 和 `unsigned int`,

最小取值范围是【0, 65535】；对于 unsigned long，最小取值范围是【0, 4294967295】。long long 类型是为了支持 64 位的需求，最小取值范围是【-1223372036854775807, 9223372036854775807】；unsigned long long 的最小取值范围是【0, 18446744073709551615】。如果要开支票，这个数是一千八百亿亿（兆）六千七百四十四成亿零九百五十五万一千六百一十五。但是，谁会去数？

int 类型那么多，应该如何选择？首先，考虑 unsigned 类型。这种类型的数常用于计数，因为计数不用负数。而且，unsigned 类型可以表示更大的正数。

如果一个数超出了 int 类型的取值范围，且在 long 类型的取值范围内时，使用 long 类型。然而，对于那些 long 占用的空间比 int 大的系统，使用 long 类型会减慢运算速度。因此，如非必要，请不要使用 long 类型。另外要注意一点：如果在 long 类型和 int 类型战胜空间相同的机器上编写代码，当确实需要 32 位整数时，应使用 long 类型而不是 int 类型，以便把程序移植到 16 位机后仍然可以正常工作。类似地，如果确实需要 64 位的整数，应使用 long long 类型。

如果在 int 设置为 32 位系统中要使用 16 位的值，应使用 short 类型以节省存储空间。通常，只有当程序使用相对于系统可用内存较大的整形数组时，才需要重点考虑节省空间的问题。使用 short 类型的另一个原因是，计算机中某些组件使用的硬件寄存器是 16 位。

long 常量和 long long 常量

通常，程序代码中使用的数字（如，2345）都被储存为 int 类型。如果使用 10000000 这样在大数字，超出了 int 类型能表示的范围，编译器会将其视为 long int 类型（假设这种类型可以表示该数字）。如果数字超出 long 可表示的最大值，编译器则将其视为 unsigned long 类型。如果还不够大，编译器则将其视为 long long 或 unsigned long long 类型（前提上编译器能识别这些类型）。

八进制和十六进制常量被视为 int 类型。如果值太大，编译器会尝试使

用 unsigned int。如果还不够大，编译器依次使用 long、unsigned long、long long 和 unsigned long long 类型。

有些情况下，需要编译器以 long 类型储存一个小数字。例如，编程时要显式使用 IBM PC 上的内存地址是。另外，一些 C 标准函数也要求使用 long 类型的值。要把一个较小的常量作为 long 类型对待，可以在值的末尾加上 l（小写的 l）或 L 后缀。使用 L 后缀更好，因为 l 看上去和数字 1 很像。因此，在 int 为 16 位、long 为 32 位的系统中，会把 7 作为 16 位储存，把 7L 作为 32 位储存。l 或 L 后缀也可用于八进制和十六进制整数，如 020L 和 0x10L。

类似地，在支持 long long 类型的系统中，也可以使用 ll 和 LL 后缀来表示 long long 类型的值，如 3LL。另外，u 或 U 后缀表示 unsigned long long，如 5ull、10LLU、6LLU 或 9ULL。

整数溢出

如果整数走出了相应类型的取值范围会怎样？下面分别将有符号类型的无符号类型的整数设置为比最大值略大，看看会发生什么（printf () 函数使用 %u 说明显示 unsigned int 类型的值）。

```
//toobig.c 走出系统允许的最大 int 值
#include<stdio.h>
int main(int argc, const char*argv[])
{
    int i=2147483647;
    unsigned int j=4294967295;
    printf("%d %d %d\n", i, i+1, i+2);
    printf("%u %u %u\n", j, j+1, j+2);
    return 0;
}
```

在系统下输出的结果是：

```
2147483647 -2147483648 -2147483647
4294967295 0 1
```

可以把无符号整数 *j* 看作是汽车的里程表。当达到能表示的最大值时，会重新从起始点开始。整数 *i* 也是类似的情况。主要区别是，在超过最大值时，`unsigned int` 类型的变量 *j* 从 0 开始。注意，当 *i* 超出（溢出）其相应类型所能表示的最大值时，系统并未通知用户。因此，在编程时必须自己注意这类问题。

溢出行为是未定义的行为，C 标准并未定义有符号类型的溢出规则。以上描述的溢出行为比较有代表性，但是也可能会出现其他情况。

打印 `short`、`long`、`long long` 和 `unsigned` 类型

打印 `unsigned int` 类型的值，使用 `%u` 转换说明；打印 `long` 类型的值，使用 `%ld` 转换说明。如果系统中 `int` 和 `long` 的大小相同，使用 `%d` 就行。但是，这样的程序被移植到其他系统（`int` 和 `long` 类型的大小不同）中会无法正常工作。在 `x` 和 `o` 前面可以使用 `l` 前缀，`%lx` 表示以十六进制格式打印 `long` 类型整数。`%lo` 表示以八进制格式打印 `long` 类型整数。注意，虽然 C 允许使用大写或小写的常量后缀，但是在转换说明中只能用小写。

C 语言有多种 `printf()` 格式。对于 `short` 类型，可以使用 `h` 前缀。`%hd` 表示以二进制显示 `short` 类型的整数，`%ho` 表示以八进制显示 `short` 类型的整数。`h` 和 `l` 前缀都可以和 `u` 一起使用，用于表示无符号类型。例如，`%lu` 表示打印 `unsigned long` 类型的值。程序清单 3.4 演示了一些例子。对于支持 `long long` 类型的系统，`%lld` 和 `%llu` 分别表示有符号和无符号类型。第 4 章将详细介绍转换说明。

```
//程序清单 3.4 print2.c 程序
#include <stdio.h>
int main(int argc, const char*argv[])
{
    unsigned int un=3000000000; //int 为 32 位和 short 为 16 位的系统
    short end=200;
    long big=65537;
```

```
long long verybig=12345678908642;
printf( "un=%u and not %d\n",un,un );
printf( "end=%hd and %d\n",end,end );
printf( "big=%ld and not %hd\n",big,big );
printf( "verybig=%lld and not %ld\n",verybig,verybig );
return 0;
}
```

在特定的系统中输出如下（输出的结果可能不同）：

un=3000000000 and not -1294967296 end=200 and 200 big=65537 and
not 1 verybig=12345678908642 and not 12345678908642

该例表明，使用错误的转换说明会得到意想不到的结果。第 1 行输出，对于无符号变量 un，使用 %d 会生成负值！其原因是，无符号值 3000000000 和有符号值 -1294967296 在系统内存中的内部表示完全相同（详见第 15 章）。因此，如果告诉 print () 该数是无符号数，它打印一个值；如果告诉它该数是有符号数，它将打印另一个值。在待打印的值大于有符号值的最大值时，会发生这种情况。对于较小的正数（如 96），有符号和符号类型的存储、显示都相同。

第 2 行输出，对于 short 类型的变量 end，在 printf () 中无论指定以 short 类型 (%hd) 还是 int 类型 (%d) 打印，打印出来的值都相同。这是因为在给函数传递参数时，C 编译器把 short 类型的值自动转换成 int 类型的值。你可能会抽出疑问：为什么要进行转换？h 修饰符有什么用？第 1 个问题的答案是，int 类型被认为是计算机处理整数类型时最高效的类型。因此，在 short 和 int 类型的大小不同的计算机中，用 int 类型的参数传递速度更快。第 2 个问题的答案中，使用 h 修饰符可以显示较大整数被截断成 short 类型值的情况。第 3 行输出就演示了这种情况。把 65537 以二进制格式写成一个 32 位数是 000000000000000100000000000001。使用 %hd, printf () 只会查看后 16 位所以显示的值是 1。与此类似，输出的最后一行先显示了 verybig 的完整值，然后由于使用了 %ld, print () 只显示了储存在后 32 位的值。

本章前面介绍过，程序员必须保证转换说明的数量和待打印值的数量相同。以上内容也提醒读者，程序员还必须根据待打印值的类型使用正确的转换说明。

提示匹配 `printf()` 函数时，切记检查每个待打印值都有对应的转换说明，还要检查转换说明的类型是否与待打印值的类型相匹配。

3.4.3 使用字符：char 类型

`char` 类型用于储存字符（如，字母或标点符号），但是从技术层面看，`char` 是整数类型。因为 `char` 类型实际上储存的是整数而不是字符。计算机使用数字编码来处理字符，即用特定的整数表示特定的字符。美国最常用的编码是 ASCII 编码，此也使用此编码。例如，在 ASCII 码中，整数 65 代表大写字母 A。因此，储存字母 A 实际上储存的是整数 65（许多 IBM 的大型主机使用另一种编码-EBCDIC，其原理相同。另外，其他国家的计算机系统可能使用完全不同的编码）。

标准 ASCII 码的范围是 0~127，只需 7 位二进制数即可表示。通常，`char` 类型被定义为 8 位的存储单元，因此容纳标准 ASCII 码绰绰有余。许多其他系统（如 IBM PC 和苹果 Macs）还提供扩展 ASCII 码，也在 8 位表示范围之内。一般而言，C 语言会保证 `char` 类型足够大，以储存系统（实现 C 语言的系统）的基本字符集。

许多字符集都超过了 127，甚至多于 255。例如，日本汉字（kanji）字符集。商用的统一码（Unicode）创建了一个能表示世界范围内多种字符集的系统，目前包含的字符已超过 110000 个。国际标准化组织（ISO）和国际电工技术委员会（IEC）为字符集开发了 ISO/IEC10646 标准。统一码标准也与 ISO/IEC10646 标准兼容。

C 语言把 1 字节定义为 `char` 类型占用的位（bit）数，因此无论是 16 位还是 32 位系统，都可以使用 `char` 类型。

声明 char 类型变量

char 类型变量的声明方式与其他类型变量的声明方式相同。下面是一些例子：

```
char response;
```

1. response response 英 [rspns] 美 [rspns] n. (口头的或书面的) 回答, 答复; 反应; 响应; (礼拜仪式中的) 答唱咏, 启应经; [例句]There has been no response to his remarks from the government 政府尚未对他的言论作出回应。[其他] 复数: responses

```
char itable, Iatan;
```

2. itable

合适的

3. Iatan 国际航空协会

以上声明创建了 3 个 char 类型的变量: response、itable 和 Iatan。

字符常量和初始化

如果要把一个字符常量初始化为字母 A, 不必背下 ASCII 码, 用计算机语言很容易做到。通过以下初始化把字母 A 赋给 grade 即可:

```
char grade='A';
```

在 C 语言中, 用单引号括起来的单个字符被称为字符常量 (character constant)。编译器一发现 'A', 就会将其转换成相应的代码值。单引号必不可少。下面还有一些其他的例子:

```
char broiled; //声明一个 char 类型的变量
```

```
broiled='T'; //为其赋值, 正确
```

```
broiled=T; //错误! 此时 T 是一个变量
```

```
broiled="T"; //错误! 此时"T"是一个字符串
```

如上所示，如果省略单引号，编译器认为 T 是一个变量名：如果把 T 用双引号括起来，编译器则认为“T”是一个字符串。字符串的内容将在第 4 章中介绍。

实际上，字符是以数值形式储存的，所以也可使用数字代码值来赋值：

```
char grade=65;//对于 ASCII，这样做没问题，但这是一种不好的编程风格
```

在本例中，虽然 65 是 int 类型，但是它在 char 类型能表示的范围内，所以将其赋值给 grade 没问题。由于 65 是字母 A 对应的 ASCII 码，因此本例是把 A 赋给 grade。注意，能这样做的前提是系统使用 ASCII 码。其实，用‘A’代替 65 才是较为妥当的做法，这样在任何系统中都不会出问题。因此，最好使用字符常量，而不是数字代码值。

奇怪的是，C 语言将字符常量视为 int 类型而非 char 类型。例如，在 int 为 32 位、char 为 8 位的 ASCII 系统中，有下面的代码：

```
char grade='B';
```

本来‘B’对应的数值 66 储存在 32 位的存储单元中，现在却可以储存在 8 位的存储单元中（grage）。利用字符常量的这种特性，可以定义一个字符常量‘FATE’，即把 4 个独立的 8 位 ASCII 码储存在一个 32 位存储单元中。如果把这样的字符常量赋给 char 类型变量 grade，只有最后 8 位有效。因此，grade 的值是‘E’。

[2020-05-06 Wed 15:55]

非打印字符

单引号只适用于字符、数字和标点符号，浏览 ASCII 表会发现，有些 ASCII 字符打印不出来。例如，一些代表行为的字符（如，退格、换行、终端响铃或蜂鸣）。C 语言提供了 3 种方法表示这些字符。

第 1 种方法前面介绍过-使用 ASCII 码。例如，蜂鸣字符的 ASCII 值是 7，因此可以这样写：

```
char beep=7;
```

第 2 种方法是，使用特殊的符号序列表示一些特殊的字符。这些符号序列叫作转义序列 (escape sequence)。表 3.2 列出了转义序列及其含义。

把转义序列赋给字符变量时，必须用单引号把转义序列括起来。例如，假设有下面一行代码：

```
char nerf='°';
```

稍后打印变量 nerf 的效果是，在打印机或屏幕上另起一行。

表 3.2 转义序列

现在，我们来仔细分析一下转义序列。使用 C90 新增的警报字符（是否能产生听到的或看到的警报，取决于计算机的硬件，蜂鸣是最常见的警报（在一些系统中，警报字符不起作用）。C 标准规定警报字符不得改变活跃位置。标准中的活跃位置 (active position) 指的是显示设备（屏幕、电传打字机、打印机等）中下一个字符将出现的位置。简而言之，平时常说的屏幕光标位置就是活跃位置。在程序中把警报字符输出在屏幕上的效果是，发出一声蜂鸣，但不会移动屏幕光标。

接下来的转义字符、\、°~常用的输出设备控制字符。了解它们最好的方式是查看它们对活跃位置的影响。换页符 () 把活跃位置移至下一页的开始处；换行符 () 把活跃位置移至下一行的开始处；回车符 (° 把活跃位置移动到当前行的开始处；水平制表符 (将活跃位置移至下一个水平制表点（通常是第 1 个、第 9 个、第 17 个、第 25 个等字符位置）；垂直制表符 (~ 把活跃位置移到下一个垂直制表点。

这些转义序列字符不一定在所有的显示设备上都起作用。例如，换页符和垂直制表符在 pc 屏幕上会生存奇怪的符号，光标并不会移动。只有将其输出到打印机上时才会产生前面描述的效果。

接下来的 3 个转义序列 (\\、\'、\") 用于打印 \、\'、\" 字符（由于这些字符用于定义字符常量），是 printf () 函数的一部分，若直接使用它们会造成混乱)。如果打印下面一行内容：

```
Gramps sez,\"a\\ is a backslash."
```

应这样编写代码：

```
printf ("Gramps sez,\"a \\ is a backslash.\"") ;
```

表 3.2 中的最后两个转义序列 (`\0oo` 和 `\x`) 是 ASCII 码的特殊表示。如果要用八进制 ASCII 码表示一个字符，可以在编码值前面加一个反斜杠 (`\`) 并用单引号括起来。例如，如果编译器不识别警报字符 (`\a`)，可以使用 ASCII 码来代替：

```
beep='7';
```

可以省略前面的 `o`，`'\07'` 甚至 `'\7'` 都可以。即使没有前缀 `o`，编译器在处理这种写法时，仍会解释为八进制。

从 C90 开始，不仅可以用十进制、八进制形式表示字符常量，C 语言还提供了第 3 种选择-用十六进制形式表示字符常量，即反斜杠后面跟一个 `x` 或 `X`，再加上 1~3 位十六进制数字。例如，`Ctrl+P` 字符的 ASCII 十六进制码是 `10`（相当于十进制的 `16`），可表示为 `'10'` 或 `'010'`。图 3.5 列出了一些整数类型的不同进制形式。

整形常量的例子			
类型	十六进制	八进制	十进制
char	41	0101	N.N.
int	0x41	0101	65
unsigned int	0x41u	0101u	65u
long	0x41L	0101L	65L
unsigned long	0x41UL	0101UL	65UL
long long	0x41LL	0101LL	65LL
unsigned long long	0x41ULL	0101ULL	65ULL

图 3.5int 系列类型的常量写法示例

使用 ASCII 码时，注意数字和数字字符的区别。例如，字符 `4` 对应的 ASCII 码是 `52`。`'4'`表示字符 `4`，而不是数值 `4`。

关于转义序列，读者可能有下面 3 个问题。

上面最后一个例子 (`printf ("Gramps sez, \a \a backslash \n")`)，为何没有用单引号把转义序列括起来？无论是普通字符还是转义序列，只要是双引号括起来的字符集合，就无需单引号括起来。双引号中的字符集

合叫作字符串（详见第 4 章）。注意，该例中的其他字符（G、r、a、m、p、s 等）都没有用单引号括起来。与此类似，`printf("Hello!\007");` 将打印 Hello! 并发出一声蜂鸣，而 `printf("Hello! 7");` 则打印 Hello! 7。不是转义序列中的数字将作为普通字符被打印出来。

何时使用 ASCII 码？何时使用转义序列？如果要在转义序列（假设使用”）和 ASCII 码（’\014’）之间选择，请选择前者（即’）。这样的写法不仅更好记，而且可移植性更高。”在不使用 ASCII 码的系统中，仍然有效。

如果要使用 ASCII 码，为何要写成’\032’而不是 032？首先，’\032’能更清晰地表达程序员使用字符编码的意图。其次，类似 \032 这样的转义序列可以嵌入 C 的字符串中，如 `printf("Hello!\007");` 中就嵌入了 \007。

打印字符

`printf()` 函数用 %C 指明待打印的字符。前面介绍过，一个字符变量实际上被储存为 1 字节的整数值。因此，如果用 %d 转换说明打印 char 类型变量的值，打印的是一个整数。而 %c 转换说明告诉 `printf()` 打印该整数值对应的字符。程序清单 3.5 演示了打印 char 类型变量的两种方式。

程序清单 3.5 charcode.c 程序

```
//charcode.c - 显示字符的代码编号
#include<stdio.h>
int main(int argc, const char*argv[])
{
    char ch;
    printf("Please enter a character.\n");
    scanf("%c",&ch); //用户输入字符
    printf("The code for %c is %d.\n",ch,ch);
    return 0;
}
```

运行该程序后，输出示例如下：

Please enter a character.

c

The code for c is 67.

运行该程序时，在输入字母后还要忘记按下 Enter 或 Return 键。随后，scanf () 函数会读取用户输入的字符，& 符号表示把输入的字符赋给变量 ch。接着，printf () 函数打印 ch 的值两次，第 1 次打印一个字符（对应代码中的%c），第 2 次打印一个十进制整数值（对应代码中的%d）。注意，printf () 函数中的转换说明决定了数据的显示方式，而不是数据的储存方式（见图 3.6）。

ch |0|1|0|0|0|0|1|1| 存储（ASCII 码）

“%c”“%d”代码

c 67 显示

图 3.6 数据显示和数据存储

有符号还是无符号

有些 c 编译器把 char 实现为有符号类型，这意味着 char 可表示的范围是-128~127。而有些 C 编译器把 char 实现为无符号类型，那么 char 可表示的范围是 0~255。请查阅相应的编译器手册，确定正在使用的编译器如何实现 char 类型。或者，可以查阅 limits.h 头文件。下一章将详细介绍头文件的内容。

根据 C90 标准，C 语言允许在关键字 char 前面使用 signed 或 unsigned。这样，无论编译器默认 char 是什么类型，signed char 表示有符号类型，而 unsigned char 表示无符号类型。这在用 char 类型处理小整数时很有用。如果只用 char 处理字符，那么 char 前面无需使用任何修饰符。

3.4.4 __Bool 类型

C99 标准添加了_Bool 类型，对于表示布尔值，即逻辑值 true 和 false。因为 C 语言用值 1 表示 true，所以_Bool 类型实际上也是一种整数类型。但

原则上它仅占用 1 位存储空间，因为对 0 和 1 而言，1 位的存储空间足够了。

程序通过布尔值可选择执行哪部分代码。我们将在第 6 章和第 7 章中详述相关内容。

3.4.5 可移植类型：stdint.h 和 inttypes.h

C 语言提供了许多有用的整数类型。但是，某些类型名在不同系统中的功能不一样。C99 新增了两个头文件 `stdint.h` 和 `inttypes.h`，以确保 C 语言的类型在各系统中的功能相同。

C 语言为现有类型创建了更多类型名。这些新的类型名定义在 `stdint.h` 头文件中。例如，`int32_t` 表示 32 位的有符号整数类型。在使用 32 位 `int` 的系统中，头文件会把 `int32_t` 作为 `int` 的别名。不同的系统也可以定义相同的类型名。例如，`int` 为 16 位、`long` 为 32 位的系统会把 `int32_t` 作为 `long` 的别名。然后，使用 `int32_t` 类型编写程序，并包含 `stdint.h` 头文件时，编译器会把 `int` 或 `long` 替换成当前系统匹配的类型。

上面讨论的类型别名是精确宽度整数类型（exact-width integer type）的示例。`int32_t` 表示整数类型的宽度正好是 32 位。但是，计算机的底层系统可能不支持。因此，精确宽度整数类型是可选项。

如果系统不支持精确宽度整数类型怎么办？C99 和 C11 提供了第 2 类别名集合。一些类型名保证所表示的类型一定是至少有指定宽度的最小整数类型。这组集合被称为最小宽度类型（minimum width type）。例如，`int_least8_t` 是可容纳 8 位有符号整数值类型中宽度最小的类型的一个别名。如果某系统的最小整数类型是 16 位，可能不会定义 `int8_t` 类型。尽管如此，该系统仍可使用 `int_least8_t` 类型，但可能该类型实现为 16 位的整数类型。

当然，一些程序员更关心速度而非空间。为此，C99 和 C11 定义了一组可使计算达到最快的类型集合。这组类型集合被称为最快最小宽度类型（fastest minimum width type）。例如，`int_fast8_t` 被定义为系统中对 8 位有符号值而言最快的整数类型的别名。

另外，有些程序员需要系统最大整数类型。为此，C99 定义了最大的有

符号整数类型 `intmax_t`，可储存任何有效的有符号整数值。类似地，`uintmax_t` 表示最大的无符号整数类型。顺带一提，这些类型有可能比 `long long` 和 `unsigned long` 类型更大，因为 C 编译器除了实现标准规定的类型以外，还可以利用 C 语言实现其他类型。例如，一些编译器在标准引入 `long long` 类型之前，已提前实现了该类型。

C99 和 C11 不仅提供可移植的类型名，还提供相应的输入和输出。例如，`printf()` 打印特定类型时要求与相应的转换说明匹配。如果要打印 `int32_t` 类型的值，有些定义使用 `%d`，而有些定义使用 `%ld`，怎么办？C 标准针对这一情况，提供了一些字符串宏（第 4 章中详细介绍）来显示可移植类型。例如，`inttypes.h` 头文件中定义了 `PRId32` 字符串宏，代表打印 32 位有符号值的合适转换说明（如 `d` 或 `l`）。程序清单 3.6 演示了一种可移植类型和相应转换说明的用法。

程序清单 3.6 `altnames.c` 程序

```
//altnames.c 可移植整数类型名
#include <stdio.h>
#include <inttypes.h> //支持可移植类型
int main(int argc, const char*argv[])
{
    int32_t me32; //me32是一个32位有符号整形变量
    me32=45933945;
    printf("First, assume int32_t is int: ");
    printf("me32 = %d\n",me32);
    printf("Next, let's not make any assumptions.\n");
    printf("Instead, use a \"macro\" from inttypes.h: ");
    printf("me32=%" PRId32 "\n",me32);
    return 0;
}
```

该程序最后一个 `printf()` 中，参数 `PRId32` 被定义在 `inttypes.h` 中的“`d`”替换，因而这条语句等价于：


```
printf ("me16=%" "d" ",me16) ;
```

在 C 语言中，可以把多个连续的字符串组合成一个字符串，所以这条语句又等价于：

```
printf ("me16=%d",me16) ;
```

下面是程序的输出，注意，程序中使用了 \" 转义序列来显示双引号：

```
First, assume int32_t is int: me32=45933945
```

```
Next, let's not make any assumptions.
```

```
Instead, use a "macro" from inttypes.h: me32=45933945
```

篇幅有限，无法介绍扩展的所有整数类型。本节主要是为了让读者知道，在需要时可进行这种级别的控制。附录 B 中的参考燃料 VI“扩展的整数类型”介绍了完整的 `inttypes.h` 和 `stdint.h` 头文件。

注意对 C99/C11 的支持

C 语言发展至今，虽然 ISO 已发布了 C11 标准，但是编译器供应商对 C99 的实现程度却各不相同。在本书第 6 版的编写过程中，一些编译器仍未实现 `inttypes.h` 头文件及其相关功能。

3.4.6 float、double 和 long double

各种整数类型对大多数软件开发项目而言够用了。然而，面向金融和数学的程序经常使用浮点数。C 语言中的浮点类型有 `float`、`double` 和 `long double` 类型。它们与 FORTRAN 和 Pascal 中的 `real` 类型一致。前面提到过，浮点类型能表示包括小数在内更大范围的数。浮点数的表示类似于科学记数法（即用小乘乘以 10 的幂来表示数字）。该记数系统常用于表示非常大或非常小的数。表 3.3 列出了一些示例。

表 3.3 记数法示例

数字	科学记数法	指数记数法
10000000000	1.0×10^{10}	1.0e9
123000	1.23×10^5	1.23e5
322.56	3.2256×10^2	3.2256e2
0.000056	5.6×10^{-5}	5.6e-5

第 1 列是一般记数法：第 2 列是科学记数法：第 3 列是指数记数法（或称为 e 记数法），这是科学记数法在计算机中的写法，e 后面的数字代表 10 的指数。图 3.7 演示了更多的浮点数写法。

C 标准规定，float 类型必须至少能表示 6 位有效数字，且取值范围至少是 $10^{-37} \sim 10^{37}$ 。前一项规定指 float 类型必须至少精确表示小数点后的 6 位有效数字，如 33.333333。后一项规定用于方便地表示诸如太阳质量（ 2.0×10^{30} 千克）、一个质子的电荷量（ 1.6×10^{-19} 库仑）或国家债务之类的数字。通常，系统储存一个浮点数要占用 32 位。其中 8 位用于表示指数的值和符号，剩下 24 位用于表示非指数部分（也叫作尾数或有效数）及其符号。

C 语言提供的另一种浮点类型是 double（意为双精度）。double 类型 and float 类型的最小取值范围相同，但至少必须能表示 10 位有效数字。一般情况下，double 占用 64 位而不是 32 位。一些系统将多出的 32 位全部用来表示非指数部分，这不仅增加了有效数字的倍数（即提高了精度），而且还减少了舍入误差。另一些系统把其中的一些位分配给指数部分，以容纳更大的指数，从而增加了可表示数的范围。无论哪种方法，double 类型的值至少有 13 位有效数字，超过了标准的最低位数规定。

C 语言的第 3 种浮点类型是 long double，以满足比 double 类型更高的精度要求。不过，C 只保证 long double 类型至少与 double 类型的精度相同。

1. 声明浮点型变量

浮点型变量的声明和初始化方式与整型变量相同，下面是一些例子：

```
float noah, jonah;
```

```
double trouble;
```

```
float planck=6.63e-34;
```

```
long double gnp;
```

2. 浮点型常量

在代码中，可以用多种形式书写浮点型常量。浮点型常量的基本形式是：有符号的数字（包括小数点），后面紧跟 e 或 E，最后是一个有符号数表示 10 的指数。下面是两个有效的浮点型常量：

```
-1.56E+123000
```

```
2.87e-3
```

正号可以省略。可以没有小数点（如，2E5）或整数部分（如，.45E-6），但是不能同时省略两者。下面是更多的有效浮点型常量示例：

```
3.14159
```

```
.2
```

```
4e16
```

```
.8E-5
```

```
100
```

不要在浮点型常量中间加空格：1.56 E+12(错误！)

默认情况下，编译器假定浮点型常量是 double 类型的精度。例如，假设 some 是 float 类型的

变量，编写下面的语句：

```
some=4.0×2.0;
```

参演，4.0 和 2.0 被储存为 64 位的 double 类型，使用双精度进行乘法运算，然后将乘积截断成 float 类型的宽度。这样做虽然计算精度更高，但是会减慢程序的运行速度。

在浮点数后面加上 f 或 F 后缀可覆盖默认设置，编译器会将浮点型常量看作 float 类型，如 2.3f 和 9.11E9F。使用 l 或 L 后缀使得数字成为 long double 类型，如 54.31 和 4.32L。注意，建议使用 L 后缀，因为字母 l 和数字 1 很容易混淆。没有后缀的浮点型常量是 double 类型。

C99 标准添加了一种新的浮点型常量格式-用十六进制表示浮点型常量，即在十六进制数前加上十六进制前缀 (0x 或 0X)，用 p 和 P 分别代替 e 和 E，用 2 的幂代替 10 的幂（即，p 计数法）。如下所示：

0xa.1fp10

十六进制 a 等于十进制 10, .1f 是 1/16 加上 15/256 (十六进制 f 等于十进制 15), p10 是 2^{10} 或 1024。0xa.1fp10 表示的值是 $(10+1/16+15/256) \times 1024$ (即, 十进制 10364.0)。

注意, 并非所有的编译器都支持 C99 的这一特性。

3. 打印浮点值

printf() 函数使用 %f 转换说明打印十进制记数法的 float 和 double 类型浮点数, 用 % 打印指数记数法的浮点数。如果系统支持十六进制格式的浮点数, 可用 a 和 A 分别代替 e 和 E。打印 long double 类型要使用 %Lf、%Le 或 %La 转换说明。给那些未在函数原型中显式说明参数类型的函数 (如, printf()) 传递参数时, C 编译器会把 float 类型的值自动转换成 double 类型。程序清单 3.7 演示了这些特性。

程序清单 3.7 showf—pt.c 程序

```
//showf_pt.c—以两种方式显示 float 类型的值
#include<stdio.h>
int main(void)
{
    float  aboat=32000.0;
    double abet=2.14e9;
    long double dip=5.32e-5;
    printf("%f can be written %e\n",aboat,aboat);
    //下一行要求编译器支持 C99 或其中相关特性
    printf("And it's %a in hexadecimal, powers of 2 notation\n",aboat);
    printf("%f can be written %e\n",abet, abet);
    printf("%Lf can be written %Le\n",dip dip);
    return 0;
}
```

该程序示例演示了默认的输出效果。下一章将介绍如何通过设置字段宽度和小数位数来控制输出格式。

4. 浮点数的上溢和下溢

假设系统的最大 float 类型值是 3.4E38，编写如下代码：

```
float toobig=3.4E38*100.0f;
printf ("%e",toobig) ;
```

会发生什么？这是一个上溢（overflow）的示例。当计算导致数字过大，超过当前类型能表达的范围时，就会发生上溢。这种行为在过去是未定义的，不过现在 C 语言规定，在这种情况下会给 toobig 赋一个表示无穷大的特定值，而且 printf () 显示该值为 inf 或 infinity（或者具有无穷含义的其他内容）。

当除以一个很小的数时，情况更为复杂。回忆一下，float 类型的数以指数和尾数部分来储存。存在这样一个数，它的指数部分是最小值，即由全部可用位表示的最小尾数值。该数字是 float 类型能用全部精度表示的最小数字。现在把它除以 2。通常，这个操作会减小指数部分，但是假设的情况中，指数已经是最小值了。所以计算机只好把尾数部分的位向右移，突出第 1 个二进制位，并丢弃最后一个二进制数。以十进制为例，把一个有 4 位有效数字的数（如，0.1234E-10）除以 10，得到的结果是 0.123E-10。虽然得到了结果，但是在计算过程中却损失了原末尾有效位上的数字。这种情况叫作下溢（underflow）。C 语言把损失了类型全精度的浮点值称为低于正常的（subnormal）浮点值。因此，把最小的正浮点数除以 2 将得到一个低于正常的值。如果除以一个非常大的值，会导致所有的位都为 0。现在，C 库已提供了用于检查计算是否会产生低于正常值的函数。

还有另一个特殊的浮点值 NaN（not a number 的缩写）。例如，给 asin () 函数传递一个值，该函数将返回一个角度，该角度的正弦就是传入函数的值。但是正弦值不能大于 1，因此，如果传入的参数大于 1，该函数的行为是未定义的。在这种情况下，该函数将返回 NaN 值，printf () 函数可将其显示为 nan、NaN 或其他类似的内容。

浮点数舍入错误

给定一个数，加上 1，再减去原来给定的数结果是多少？你一定认为是 1。但是，下面的浮点运算给出了不同的答案：

//floaterr.c --演示舍入错误

```
#include<stdio.h>
int main(void)
{
    float a,b;
    b=2.0e20+1.0;
    a=b-2.0e20;
    printf( "%f \n",a);
    return 0;
}
```

该程序的输出如下：

0.000000 \leftarrow Linux 系统下的老式 gcc

-13584010575872.000000 \leftarrow Turbo C 1.5

4008175468544.000000 \leftarrow XCode 4.5、Visual Studio 2012、当前版本的 gcc

得出这些奇怪答案的原因是，计算机缺少足够的小数位来完成正确的运算。2.0e20 是 2 后面有 20 个 0。如果把该数加 1，那么发生变化的是第 21 位。要正确运算，程序至少要储存 21 位有效数字。而 float 类型的数字通常只能储存按指数比例缩小或放大的 6 或 7 位有效数字。在这种情况下，计算结果一定是错误的。另一方面，如果把 2.0e20 改成 2.0e4，计算结果就没问题。因为 2.0e4 加 1 只需改变第 5 位上的数字，float 类型的精度足够进行这样的计算。

浮点数的表示法

上一个方框中列出了由于计算机使用的系统不同，一个程序有不同的输出。原因是，根据前面介绍的知识，实现浮点数表示法的方法有多种。为了尽可能地统一实现，电子和电气工程师协会 (IEEE) 为浮点数计算和表示法开发了一套标准。现在，许多硬件浮点单元都采用该标准。2011 年，该标准被 ISO/IEC/IEEE60559: 2011 标准收录。该标准作为 C99 和 C11 的可选项，符号硬件要求的平台可开启。floaterr.c 程序的第 3 个输出示例既

是支持该浮点标准的系统显示的结果。支持 C 标准的编译器还包括含捕获异常问题的工具。详见附录 B.5，参考资料 V。

3.4.7 得数和虚数类型

许多科学和工程计算都要用到得数和虚数。C99 标准支持复数类型和虚数类型，但是有所保留。一些独立实现，如嵌入式处理器的实现，就不需要使用复数和虚数（VCR 芯片就不需要复数）。一般而言，虚数类型都是可选项。C11 标准把整个复数软件包都作为可选项。

简而言之，C 语言有 3 种复数类型：float ——Complex、double ——Complex 和 long double ——Complex。例如 float _Complex 类型的变量应包含两个 float 类型的值，分别表示复数的实部和虚部。类似地，C 语言的 3 种虚数类型是 float _Imaginary、double _Imaginary 和 long double _Imaginary。

如果包含 complex.h 头文件，便可用 complex 代替 _Complex，用 imaginary 代替 _Imaginary，还可以用 I 代替-1 的平方根。

为何 C 标准不直接用 complex 作为关键字来代替Complex，而要添加一个头文件（该头文件中把 complex 定义为Complex）？因为标准委员会考虑到，如果使用新的关键字，会导致以该关键字作为标识的现有代码全部失效。例如，之前的 C99，许多程序员已经使用 struct complex 定义一个结构来表示复数或者心理学程序中的心理状况（关键字 struct 用于定义能储存多个值的结构，详见第 14 章）。让 complex 成为关键字会导致之前的这些代码出现语法错误。但是，使用 struct _Complex 的人很少，特别是标准使用首字母是下划线的标识符作为预留字以后。因此，标准委员会选定Complex作为关键字，在不用考虑名称冲突的情况下可选择使用 complex。

3.4.8 其他类型

现在已经介绍完 C 语言的所有基本数据类型。有些人认为这些类型实在太多了，但有些人觉得还不够用。注意，虽然 C 语言没有字符串类型，但

也能很好地处理字符串。第 4 章将详细介绍相关内容。

C 语言还有一些从基本类型衍生的其他类型。包括数组、指针、结构和联合。尽管后面章节中会详细介绍这些类型，但是本章的程序示例中已经用到了指针【指针 (pointer) 指向变量或其他数据对象位置】。例如，在 `scanf ()` 函数中用到的前缀 `&`，便创建了一个指针，告诉 `scanf ()` 把数据放在何处。

小结：基本数据类型

关键字：

基本数据类型由 11 个关键字组成：`int`、`long`、`short`、`unsigned`、`char`、`float`、`double`、`signed`、`Bool`、`Complex` 和 `Imaginary`。

有符号整形：

有符号整形可用于表示正整数和负整数。

`int` ——系统给定的基本整数类型。C 语言规定 `int` 类型不小于 16 位。

`short` 或 `short int` ——最大的 `short` 类型整数小于或等于最大的 `int` 类型整数。C 语言规定 `short` 类型至少占 16 位。

`long` 或 `long int` ——该类型可表示的整数大于或等于最大的 `int` 类型整数。C 语言规定 `long` 类型至少占 32 位。

`long long` 或 `long long int` ——该类型可表示的整数大于或等于最大的 `long` 类型整数。`Long long` 类型至少占 64 位。

一般而言，`long` 类型占用的内存比 `short` 类型大，`int` 类型的宽度要么和 `long` 类型相同，要么和 `short` 类型相同。例如，旧 DOS 系统的 PC 撮合 16 位的 `short` 和 `int`，以及 32 位 `long`；Windows 95 系统提供 16 位的 `short` 以及 32 位的 `int` 和 `long`。

无符号整形：

无符号整形只能用于表示零和正整数，因此无符号整形可表示的正整数比有符号整形的大。在整形类型前加上关键字 `unsigned` 表明该类型是无符号整形：`unsignedint`、`unsigned long`、`unsigned short`。单独的 `unsigned` 相当于 `unsignedint`。

字符类型：

可打印出来的符号（如 A、& 和 +）都是字符。根据定义，char 类型表示一个字符要占用 1 字节内存。出于历史原因，1 字节通常是 8 位，但是如果要表示基本字符集，也可以是 16 位或更大。

char ——字符类型的关键字。有些编译器使用有符号的 char，而有些则使用无符号的 char。在需要时，可在 char 前面加上关键字 signed 或 unsigned 来指明具体使用哪一种类型。

布尔类型：

布尔值表 true 和 false。C 语言的 1 表示 true，0 表示 false。

_Bool ——布尔类型的关键字。布尔类型是无符号 int 类型，所占用的空间只要能储存 0 或 1 即可。

实浮点类型：

实浮点类型可表示正浮点数和负浮点数。

float ——系统的基本浮点类型，可精确表示至少 6 位有效数字。

double ——储存浮点数的范围（可能）更大，能表示比 float 类型更多的有效数字（至少 10 位，通常会更多）和更大的指数。

long long ——存储浮点数的范围（可能）比 double 更大，能表示比 double 更多的有效数字和更大的指数。

复数和虚数浮点数

虚数类型是可选的类型。复数的实部和虚部类型都基于实浮点类型来构成：

float _Complex

double _Complex

long double _Complex

float _Imaginary

double _Imaginary

long long _Imaginary

小结：如何声明简单变量

1. 选择需要的类型。
2. 使用有效的字符给变量起一个变量名。

3. 按以下格式进行声明：

类型说明符号变量名；(char Fname; int Num;)

类型说明符号由一个或多个关键字组成。下面是一些示例：

int erest;

unsigned short cash;

4. 可以同时声明相同类型的多个变量，用逗号分隔各变量名，如下所示：

char ch, init, ans;

5. 在声明的同时还可以初始化变量：

float mass=6.0E24;

3.4.9 类型大小

如何知道当前系统的指定类型大小是多少？运行程序清单 3.8，会列出当前系统的各类型的大小。

程序清单 3.8 typesize.c 程序

```
//typesize.c —打印类型大小
#include<stdio.h>
int main(void)
{
    //C99为类型大小提供%zd转换说明
    printf("The int has a size of %zd bytes.\n", sizeof(int));
    printf("Type char has a size of %zd bytes.\n", sizeof(char));
    printf("Type long has a size of %zd bytes.\n", sizeof(long));
    printf("Type long long has a size of %zd bytes.\n", sizeof(long long));
    printf("Type double has a size of %zd bytes.\n", sizeof(double));
    printf("Type long double has a size of %zd bytes.\n", sizeof(long double));
    return 0;
}
```

sizeof 是 C 语言的内置运算符，以字节为单位给出指定类型的大小。C99 和 C11 提供 %zd 转换说明匹配 sizeof 的返回类型。一些不支持 C99 和 C11 的编译器可用 %u 或 %lu 代替 %zd。

该程序的输出如下：

Type int has a size of 4 bytes.

Type char has a size of 1 bytes.

Type long has a size of 8 bytes.

Type long long has a size of 8 bytes.

Type double has a size of 8 bytes.

Type long double has a size of 16 bytes.

该程序列出了 6 种类型的大小，你也可以把程序中的类型更换成感兴趣的其他类型。注意，因为 C 语言定义了 char 类型是 1 字节，所以 char 类型的大小一定是 1 字节。而在 char 类型为 16 位、double 类型为 64 位的系统中，sizeof 给出的 double 是 4 字节。在 limits.h 和 float.h 头文件中有类型限制的相关信息（下一章将详细介绍这两个头文件）。

顺带一提，注意该程序最后几行 printf () 语句都被分为两行，只要不在引号内部或一个单词中间断行，就可以这样写。

3.4.10 使用数据类型

编写程序时，应注意合理选择所需的变量及其类型。通常，用 int 或 float 类型表示数字，char 类型表示字符。在使用变量之前必须先声明，并选择有意义的变量名。初始化变量应使用与变量类型匹配的常数类型。例如：

```
int apples=3; //正确
```

```
int oranges=3.0; //不好的形式
```

与 Pascal 相比，C 在检查类型匹配方面不太严格。C 编译器甚至允许二次初始化，但在激活了较高级别警告时，会给出警告。最好不要养成这样的习惯。

把一个类型的数值初始化给不同类型的变量时，编译器会把值转换成与变量匹配的类型，这将导致部分数据丢失。例如，下面的初始化：

```
int cost=12.99; //用 double 类型的值初始化 int 类型的变量
float pi=3.1415926536; //用 double 类型的值初始化 float 类型的变量
```

第 1 个声明，cost 的值是 12。C 编译器把浮点数转换成整数时，会直接丢弃（截断）小数部分，而不进行四舍五入。第 2 个声明会损失一些精度，因为 C 只保证了 float 类型的前 6 位的精度。编译器对这样的初始化可能给出警告。读者在编译程序清单 3.1 时可能就遇到了这样警告。

许多程序员和公司内部都胡系统化的命名约定，在变量名中体现其类型。例如，用 i_ 前缀表示 int 类型，us_前缀表示 unsigned short 类型。这样，一眼就能看出来 i_smart 是 int 类型的变量，us_versmart 是 unsigned short 类型的变量。

3.4.11 参数和陷阱

有必要再次提醒读者注意 printf () 函数的用法。读者应该还记得，传递给函数的信息被称为参数。例如，printf (“Hello, pal.”) 函数调用有一个参数：“Hello,pal.”。双引号中的字符序列（如，“Hello,pal.”）被称为字符串（string），第 4 章将详细讲解相关内容。现在，关键是要理解无论双引号中包含多少个字符和标点符号，一个字符串就是一个参数。

与此类似，scanf(“%d”,&weight)函数调用有两个参数：“%d”和 &weight。C 语言用逗号分隔函数中的参数。printf () 和 scanf () 函数与一般函数不同，它们的参数个数是可变的。例如，前面的程序救命中调用过带一个、两个，甚至三个参数的 printf () 函数。程序要知道函数的参数个数才能正常工作。printf () 和 scanf () 函数用第 1 个参数表明后续 3 多少个参数，即第 1 个字符串中转换说明与后面的有数一一对应。例如，下面的语句有两个 %d 转换说明，说明后面还有两个参数：

```
printf ("%d cats ate %d cans of tuna",cats,cans) ;
```

后面的确还有两个参数：cats 和 cans。

程序员要负责确保转换说明的数量、类型与后面参数的数量、类型相

匹配。现在，C 语言通过函数原型机制检查函数调用时参数的个数和类型是否正确。但是，该机制对 `printf()` 和 `scanf()` 不起作用，因为这两个函数的参数个数可变。如果参数在匹配上有问题，会出现什么情况？假设你编写了程序清单 3.9 中的程序。

程序清单 3.9 badcount.c 程序

```
//badcount.c — 参数错误的情况
#include<stdio.h>
int main(void)
{
    int n=4;
    int m=5;
    float f=7.0f;
    float g=8.0f;
    printf( "%d\n",n,m); // 参数太多
    printf( "%d %d %d\n",n); // 参数太少
    printf( "%d %d\n",f,g); // 值的类型不匹配
    return 0;
}
```

XCode 4.6(OS 10.8) 的输出如下：

```
4
4 1 -706337836
1606414344 1
```

Microsoft Visual Studio Express 2012(Windows 7) 的输出如下：

```
4
4 0 0
0 1075576832
Mac (OS10.15)
4
4 0 -630456095
```

73896 0

编译警告信息:

```
badcount.c:9:21: warning: data argument not used by format string [-Wformat-extra-args] printf("%d",n,m);//参数太多 ~~~~ ^ badcount.c:10:18:
warning: more '%' conversions than data arguments [-Wformat] printf("%d
%d %d",n);//参数太少 ^ badcount.c:11:23: warning: format specifies
type 'int' but the argument has type 'float' [-Wformat] printf("%d
%d\n",f,g);//值的类型不匹配 ~ ^ %f badcount.c:11:25: warning: format
specifies type 'int' but the argument has type 'float' [-Wformat] printf("%d
%d",f,g);//值的类型不匹配 ~~ ^ %f 4 warnings generated.
```

Compilation finished at Thu May 14 10:45:43

注意，用 %d 显示 float 类型的值，其值不会被转换成 int 类型。在不同的平台下，缺少参数或参数类型不匹配导致的结果不同。

所有编译器都能顺利编译并运行该程序，但其中大部分会给出警告。的确，有些编译器会捕获到这类问题，然而 C 标准对此未作要求。因此，计算机在运行时可能不会捕获这类错误。如果程序正常运行，很难觉察出来。如果程序没有打印出期望值或打印出意想不到的值，你才会检查 printf () 函数中的参数个数和类型是否得当。

3.4.12 转义序列示例

再来看一个程序示例，该程序使用一些特殊的转义序列。程序清单 3.10 演示了退格 (_)、水平制表符 (\t) 的工作方式。这些概念在计算机使用电传打字机作为输出设备时就有了，但是它们不一定能与现代的图形接口兼容。例如，程序清单 3.10 在某些 Macintosh 的实现中就无法正常运行。

程序清单 3.10 escape.c 程序

```
//escape.c —使用转义序列
#include<stdio.h>
int main(void)
```

```

{
float salary;
printf( "\aEnter your desired monthly salary: "); //1
printf( "$_____ \b\b\b\b\b\b\b\b"); //2
scanf( "%f",&salary );
printf( "\n\t$%.2f a month is $%.2f a year.", salary, salary*12.0); //3
printf( "\rGee!\n");
return 0;
}

```

程序运行情况

假设在系统中运行的转义序列行为与本章描述的行为一致（实际行为可能不同。例如，XCode 4.6 把和 ° 示为颠倒的问号），下面我们来分析这个程序。

第 1 条 printf () 语句（注释中标为 1）发出一声警报（因为使用了，然后打印下面的内容：

Enter your desired monthly salary:

因为 printf () 中的字符串末尾没有，所以光标停留在冒号后面。

第 2 条 printf () 语句在光标处接着打印，屏幕上显示的内容是：

Enter your desired monthly salary: \$_____

冒号和美元符号之间有一个空格，这是因为第 2 条 printf () 语句中的字符串以一个空格开始。7 个退格字符使得光标左移 7 个位置，即把光标移至 7 个下划线字符的前面，紧跟在美元符号后面。通常，退格不会擦除退回所经过的字符，但胡些实现是擦除的，这和本例不同。

假设键入的数据中 4000.00（并按下 Enter 键），屏幕显示的内容应该是：

Enter your desired monthly salary: \$4000.00

键入的字符替换了下划线字符。按下 Enter 键后，光标移至下一行起始处。水平制表符使光标移至该行的下一个制表点，一般是第 9 列（但不

一定)。然后打印字符串的其他内容。执行完该语句后，此时屏幕显示的内容应该是：

```
Enter your desired monthly salary: $4000.00
```

```
$4000.00 a month is $48000.00 a year.
```

因为这条 `printf()` 语句中没有使用换行字符，所以光标停留在最后的点号后面。

第 4 条 `printf()` 语句以 `^` 始。这使得光标回到当前行的起始处。然后打印 `Gee!`，接着使光标移至下一行的超处。屏幕最后显示的内容应该是：

```
Enter your desired monthly salary: $4000.00
```

```
Gee! $4000.00 a month is $48000.00 a year.
```

刷新输出

`printf()` 何时把输出发送到屏幕上？最初，`printf()` 语句把输出发送到一个叫作缓冲区（buffer）的中间存储区域，然后缓冲区中的内容再不断被发送到屏幕上。C 标准明确规定了何时把缓冲区的内容发送到屏幕：当缓冲区满、遇到换行字符或需要输入的时候（从缓冲区把数据发送到屏幕或文件被称为刷新缓冲区）。例如，前两个 `printf()` 语句既没有填满缓冲区，也没有换行符，但是下一条 `scanf()` 语句要求用户输入，这迫使 `printf()` 的输出被发送到屏幕上。

旧式编译器遇到 `scanf()` 也不会强行刷新缓冲区，程序会停在那里不显示任何提示内容，等等用户输入数据。在这种情况下，可以使用换行字符刷新缓冲区。代码应改为：

```
printf("Enter your desired monthly salary:");
```

```
scanf("%f",&salary);
```

无论接下来的输入是否能刷新缓冲区，代码都会正常运行。这将导致光标移至下一行起始处，用户无法在提示内容一行输入数据。还有一种刷新缓冲区的方法是使用 `fflush()` 函数，详见第 13 章。

3.4.13 关键概念

C 语言提供了大量的数值类型，目的是为程序员提供方便。那以整数类型为例，C 认为一种类型不够，提供了有符号、无符号，以及大小不同的整形，以满足不同程序的需求。

计算机中的浮点数和整数在本质上不同，基存储方式和去处过程有很大区别。即使两个 32 位存储单元储存的位组合完全相同，但是一个解释为 float 类型，另一个解释为 long 类型，这两个相同的位组合表示的值也完全不同。例如，在 PC 中，假设一个位组合表示 float 类型的数 256.0，如果将其解释为 long 类型，得到的值是 113246208。C 语言允许编写混合数据类型的表达式，但是会进行自动类型转换，以便在实际运算时统一使用一种类型。

计算机在内存中用数值编码来表示字符。美国最常用的是 ASCII 码，除此之外 C 也支持其他编码。字符常量是计算机系统使用的数值编码的符号表示，它表示为单引号括起来的字符，如：‘A’。

3.4.14 本章小结

C 有多种的数据类型。基本数据类型分为两大类：整数类型和浮点类型。通过为类型分配的储存量以及是有符号还是无符号，区分不同的整数类型。最小的整数类型是 char，因实现不同，可以是有符号的 char 或无符号的 char，即 unsigned char 或 signed char。但是，通常用 char 类型表示小整数时才这样显示说明。其他整数类型有 short、int、long 和 long long 类型。C 规定，后面的类型不能小于前面的类型。上述都是有符号类型，但也可以使用 unsigned 关键字创建相应的无符号类型：unsigned short、unsigned int、unsigned long 和 unsigned long long。或者，在类型名前加上 signed 修饰符显式表明该类型是有符号类型。最后，_Bool 类型是一种无符号类型，可储存 0 或 1，分别代表 false 和 true。

浮点类型有 3 种：float、double 和 C90 新增的 long double。后面的类型应大于或等于前面的类型。有些实现可选择支持复数类型和虚数类型，

通过关键字 `_Complex` 和 `_Imaginary` 与浮点类型的关键字组合（如，`double _Complex` 类型和 `float _Imaginary` 类型）来表示这些类型。

整数可以表示为十进制、八进制或十六进制。o 前缀表示八进制，ox 或 oX 前缀表示十六进制数。例如，32、o40、ox20 分别以二进制、八进制、十六进制表示同一个值。l 或 L 前缀表明该值是 long 类型，ll 或 LL 前缀表明该值是 long long 类型。

在 C 语言中，直接表示一个字符常量的方法是：把该字符用单引号括起来，如 'Q'、'8' 和 '\$'。C 语言的转义序列（如，'\007'），表示 ASCII 码中的一个字符。

浮点数可写成固定小数点的形式（如，9393.912）或指数形式（如，7.38E10）。C99 和 C11 提供了第 3 种指数表示法，即用十六进制数和 2 的幂来表示（如，ox1.1fp10）。

printf () 函数杜撰转换说明打印各种类型的值。转换说明最简单的形式由一个百分号 (%) 和一个转换字符组成，如 %d 或 %f。

3.4.15 复习题

复习题的参考答案在附录 A 中。

1. 指出下面各种数据使用的合适数据类型（有些可使用多种数据类型）：

- a. East Simpleton 的人口
- b. DVD 影碟的价格
- c. 本章出现次数最多的字母
- d. 本章出现次数最多的字母次数

1. 在什么情况下要用 long 类型的变量代替 int 类型的变量？

3. 使用哪些可移植的数据类型可以获得 32 位有符号整数？选择的理由是什么？

4. 指出下列常量的类型和含义（如果有的话）：

- a. ''

b.1066

c. 99.44

d. oXAA

e. 2.0e30

5.Dottie Cawm 编写一个程序，请打出程序中的错误。

```
#include<stdio.h>
main
(
float g;h;
float tan,rate;
g=e21;
tan=rate*g;
)
```

6. 写出下列常量在声明中使用的数据类型和在 printf () 中对应的转换说明：

常量	类型	转换说明 (% 转换字符)
12		
oX3		
'C'		
2.34E07		
'\040'		
7.0		
6L		
6.0f		
ox5.b6p12		

7. 写出下列常量在声明中使用的数据类型和在 printf () 中对应的转换说明（假设 int 为 16 位）：

常量	类型	转换说明 (% 转换字符)
012		
2.9e05L		
'a'		
100000		
”		
20.0f		
ox44		
-40		

8. 假设程序的开头有下列声明：

```
int imate=2;
long shot=53456;
char grade='A';
float log=2.711828;
```

把下面 printf () 语句中的转换字符补充完整：

```
printf("The odds against the %_ were %_ to 1.",imate,shot);
printf("A score of %_ is not an %_ grade.",log, grade);
```

9. 假设 ch 是 char 类型的变量。分别使用转义序列、十进制值、八进制字符常量和十六进制字符常量把回车字符赋给 ch（假设使用 ASCII 编码值）。

10. 修正下面的程序（在 C 中，/表示除以）。

```
void main(int) / this program is perfect/
{
cows,legs integer;
printf("How many cow legs did you count?);
scanf("%c",legs);
cows=legs/4;
printf("That implies there are %f cows.",cows)
```

}

11. 指出下列转义序列的含义

a.

b.

c. \"

d.

3.4.16 编程练习

1. 通过试验（即编写带有此类问题的程序）观察系统如何处理整数上溢、浮点数上溢和浮点数下溢的情况。

2. 编写一个程序，要求提示输入一个 ASCII 码值（如，66），然后打印输入的字符。

3. 编写一个程序，发出一声警报，然后打印下面的文本：

Startled by the sudden sound, Sally shouted,

"By the Great Pumpkin, what was that!"

4. 编写一个程序，读取一个浮点数，先打印成小数点形式，再打印成指数形式。然后，如果系统支持，再打印成 p 记数法（即十六进制记数法）。按以下格式输出（实际显示的指数位数因系统而异）：

Enter a floating-point value: 64.25

fixed-point notation: 64.250000

exponential notation: 6.4250000e+01

p notation: 0x1.01p+6.4250000e

5. 一个大约有 3.156×10^7 秒。编写一个程序，提示用户输入年龄，然后显示该年龄对应的秒数。

6.1 个水分子的质量约为 3.0×10^{-23} 克。1 夸脱水大约是 950 克。编写一个程序，提示用户输入水的夸脱数，并显示水分子的数量。

7.1 英寸相当于 2.54 厘米。编写一个程序，提示用户输入身高（/英寸），然后以厘米为单位显示身高。

8. 在美国的体积测量系统中, 1 品脱等于 2 杯, 1 杯等于 8 盎司, 1 盎司等于 2 大汤勺, 1 大汤勺等于 3 茶勺。编写一个程序, 提示用户输入杯数, 并以品脱、盎司、汤勺、茶勺为单位显示等价容量。思考对于该程序, 为何使用浮点类型比整数类型更合适?

第四章 总结与思考

4.1 数的类型

4.1.1 整形

- 关键字 `int`
- 声明格式: `int num; int argc, int main (int argc, const char*argv[])`
- 赋值: `num=1; scanf ("%d", &num);`
- 显示与输出: `printf("%d",num)`
- `&`: 因为是数, 所以取地址

4.1.2 浮点型

- 关键字 `float`
- 声明格式: `float weight;`
- 赋值: `weight=3.14;scanf("%f",&weight);`
- 显示与输出: `printf("%.2f",weight);`

```
//  
//newhello.c  
//newhello.c  
//  
//Created by Xia on 2020/04/25  
//Copyright @ 2020 Xia All rights reserved.  
//  
#include<stdio.h>  
#include<stdlib.h>  
#define LEN 15  
int main(int argc,const char*argv[])  
{  
    char name[LEN];  
    printf("Plases enter your name:\n");  
    //scanf("%s",name);  
    fgets(name,LEN,stdin);  
    printf("Hello , world. %s\n",name);  
    //fputs(name,stdout);  
    return 0;  
}
```

4.1.3 字符型

- 声明: `char name[20];`
- 输入: `fgets(name,20,stdin);`

因本身 `name` 为字符型, 即指针地址, 不用 `&`, 直接用 `name`。

- 输出: `printf("Hello, world! %s",name);`

4.2 目标:

- 指导性强、条理清晰而且有用的 C 语言教程。
- 一本友好、方便使用、便于自学的指南。
- 介绍 C 语言细节的同时，讲解编程的概念。假定读者为非专业的程序员。
- 当概念用文字较难解释时，以图表演示以助理解。
- C 语言的主要特征总结在方框中，便于查找及复习。
- 每章末有复习和编程练习，测试和加深
- 为有最佳学习效果，扮演一个积极的角色。仔细阅读，录入运行。
- * 手写与思考为最有效的方法，不要匆忙录入。*