普通高等教育“十一五”国家级规划教材

高等学校计算机科学与技术系列教材

数据结构（C语言版）

胡学钢　主编

高等教育出版社
Higher Education Press

TP311.12/160
2008

普通高等教育“十一五”国家级规划教材

高等学校计算机科学与技术系列教材

数据结构(C语言版)

胡学钢　主编

高等教育出版社

内容提要

“数据结构”是计算机类各专业重要的专业基础课程，是提高软件设计水平以及学习后续课程所必需的基础。课程中涉及软件设计中常见的几种数据结构及其在计算机内存中的表示（即存储）形式和各种操作的实现，以及软件设计中常用的排序和查找运算。

本书是针对应用型本科层次计算机类相关专业所编写的，主要内容包括概述、线性表、串、栈、队列、数组、树和二叉树、图、查找、排序和文件等，并配有相关的习题。作者按照实用性、模块化、通俗性的要求组织教材体系并编写各部分内容，加强了算法和程序设计方法的分析，从而避免了概念和理论讲述的平铺直叙，容易激发学生的学习兴趣，能够达到较好的学习效果。

与本书配套使用的实验教程即将推出，主要内容包括实验指导、课程设计指导、典型习题分析讲解、自测试卷及其解析等，可培养学生解决实际问题的能力，达到学以致用的效果。

本书也可作为其他相关专业学生学习“数据结构”课程的教材或参考书。

图书在版编目 (CIP) 数据

数据结构：C 语言版 / 胡学钢主编. —北京：高等教育出版社，2008.1

ISBN 978 - 7 - 04 - 022547 - 1

Ⅰ. 数… Ⅱ. 胡… Ⅲ. ①数据结构 - 高等学校 - 教材②C 语言 - 程序设计 - 高等学校 - 教材 Ⅳ. TP311.12 TP312

中国版本图书馆 CIP 数据核字（2007）第 195955 号

出版发行	高等教育出版社	购书热线	010 - 58581118
社　　址	北京市西城区德外大街 4 号	免费咨询	800 - 810 - 0598
邮政编码	100011	网　　址	http://www.hep.edu.cn
总　　机	010 - 58581000		http://www.hep.com.cn
		网上订购	http://www.landraco.com
经　　销	蓝色畅想图书发行有限公司		http://www.landraco.com.cn
印　　刷	天津新华二印刷有限公司	畅想教育	http://www.widedu.com
开　　本	787×1 092　1/16	版　　次	2008 年 1 月第 1 版
印　　张	14.75	印　　次	2008 年 1 月第 1 次印刷
字　　数	330 000	定　　价	20.30 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

版权所有　侵权必究

物料号　22547 - 00

序

计算机和通信技术的迅猛发展，不仅形成了融合度最高、潜力最大、增长最快的信息产业，而且成为推动全球经济快速增长和全面变革的关键因素。进入21世纪，我国的信息产业虽然已取得了长足的发展，但与发达国家相比，还有不小的差距。国家信息化的发展和信息产业国际竞争能力的提高，迫切需要高素质、创新型的计算机专业人才。

高素质计算机专业人才的培养离不开高质量的计算机教育。我们的专业虽然机会多，处于非常有利的条件，但是我们同样面临着一件事，就是从规模发展向质量提高的转变。怎么提高质量？专业素质的教育和应用素质的训练非常重要。尤其是我国高等教育进入大众化发展阶段，社会对计算机专业人才呈现出了多样化的需求。而与此同时，计算机学科的发展已极大地突破了原有的学科体系框架，形成了在“计算机科学与技术”之下向多个专业方向发展的新格局。在这种背景下，教育部高等学校计算机科学与技术教学指导委员会编制了《高等学校计算机科学与技术专业发展战略研究报告暨专业规范（试行）》（以下简称“专业规范”）。专业规范按照“培养规格分类”的指导思想，提出了三种类型、四个方向，即科学型（计算机科学方向）、工程型（计算机工程方向、软件工程方向）、应用型（信息技术方向）的计算机专业发展建议，体现了社会对不同人才类型的需求，对于指导我国计算机教学改革与建设，规范计算机教学工作，促进计算机教学质量的提高都具有重要的意义。

高水平的教材是一流教育质量的重要保证。为了配合专业规范的试行，便于广大高等学校教师按照新的专业规范组织实施教学，高等教育出版社在大力支持专业规范研究与起草工作的同时，还邀请规范起草小组的有关专家成立“高等学校计算机科学与技术系列教材编审委员会”，组织规划了结合计算机专业规范、面向全国高等学校计算机专业本科生的“高等学校计算机科学与技术系列教材”。令人高兴的是，一批有创新、改革

精神，且有丰富教学经验的高等学校教师投身到新体系计算机专业教材的编写中来，他们用自己创造性的思维、辛勤的汗水诠释专业规范的思想，把新的课程体系和教学内容生动地传达给师生，并进行着有意义的教学实践。

“高等学校计算机科学与技术系列教材”以专业规范和 CC2001－CC2005 有关教程为依据，以强化基础、突出实践、注重创新为原则，体现了学科课程体系和教学内容改革的新成果。此外，这一系列教材还配有丰富的教学辅助资源，并与现代教育技术手段相结合，充分发挥网络平台的作用，使教材更有利于广大教师和学生使用。目前，这一系列教材有不少选题已列入普通高等教育“十一五”国家级规划教材，希望这些教材的出版能够对新形势下我国高等学校计算机专业课程改革与建设起到积极的推动作用，使我国高等学校的计算机专业教学质量再上一个台阶。

中国科学院院士

2006—2010 年教育部高等学校计算机科学与技术教学指导委员会主任

2007 年 11 月

试读结束：需要全本请在线购买： www.ertongbook.com

前　言

“数据结构”是计算机科学与技术专业及相关专业重要的专业基础课，也是提高软件设计水平，学习后续课程所必需的基础。课程介绍软件设计中常见的线性表、串、栈和队列、数组、树和二叉树、图、文件等数据结构及其在计算机中的存储结构和各种操作的实现，介绍软件设计中常用的排序和查找方法，并讨论有关运算的性能。通过对这些内容的学习，将使学生熟练掌握各种常用结构的特性、各种运算的实现方法及其性能，并能在实际应用中根据具体问题的要求设计出合理的数据结构和运算。

数据结构课程包含的内容多，许多内容较抽象，特别是其中大量的算法设计与分析以及递归技术的应用，使学习本课程的难度加大。为此，将本课程内容划分为基础和实践两部分，分别对应《数据结构(C 语言版)》和《数据结构实验教程》两本教材。本书力求以通俗易懂的方式介绍数据结构有关知识和技术，并给出必要的例题及其分析，以展示课程中知识的应用及其学习方法，希望能够激发学生的学习兴趣，收到良好的学习效果。

全书主要内容如下：

第 1 章介绍数据结构课程的研究内容、特点以及一些基础知识和学习方法，对整个课程的学习起到引导作用。

第 2 章介绍线性表的逻辑结构和运算，重点讨论顺序表和链表这两种存储结构及其基本运算实现，给出了大量例题并进行分析、讲解。

第 3 章介绍栈、队列和数组这 3 种数据结构的逻辑结构和运算，顺序栈、链栈、顺序(循环)队列和链队列及其基本运算实现以及数组的有关内容。

第 4 章介绍树和森林的逻辑结构、存储结构和运算，介绍了二叉树的概念、性质、存储结构，重点讨论二叉树的遍历算法及其应用，讨论线索二叉树的概念、应用及相关算法，并给出了相应的例题及分析，另外还介绍了哈夫曼树的有关知识及其应用。

第5章介绍图的有关概念、运算和应用，介绍图的两种常用存储结构，重点讨论了两种遍历算法的实现和应用，最后以通俗的方式介绍了最小生成树、拓扑排序和最短路径等几个图的典型问题及算法。

第6章所介绍的查找是软件设计中常用的基本运算，讨论了在顺序表、树表和散列表等几种结构上的查找方法及其性能。

第7章重点讨论了各种排序的方法、算法及其性能分析。

第8章简要介绍有关文件方面的基础知识。

书中打“*”号的内容和习题有一定难度，可根据学校具体情况选学或选做。

与本书配套的《数据结构实验教程》包括实验指导、课程设计指导、典型习题分析讲解、自测试卷及其解析等。按照课程的具体要求，实验指导中安排了多个实验，分别侧重于线性表、栈和队列、二叉树、树、图、查找、排序等知识点；课程设计是为有条件和有能力的学生所提供的综合性设计习题，通过运用所学知识来实现给定问题的求解，锻炼学生解决实际问题的能力，以收到学以致用的效果；典型习题讲解让学生能从技术、方法上加深对所学知识的认识；自测试卷及其解析部分能让学生随时检测所学知识的效果。

本书由多位作者合作完成，编写分工如下：

合肥工业大学胡学钢负责全书的策划与组织，具体编写第1章、第4章和第5章，并对全书作了统稿与校对；合肥工业大学张玉红编写第2章和第8章，张晶编写第3章、阙夏编写第6章、周红鹃编写第7章。

本书可作为计算机及相关专业本科“数据结构”课程的教材和有关人员的教学参考书。

南京大学陈道蓄教授和南京信息工程大学顾韵华教授对本书进行了仔细的审阅，并提出了宝贵的建议，在此，特别向他们表示衷心的感谢。

由于编者水平及时间所限，书中错误和不足之处在所难免，欢迎读者批评指正。

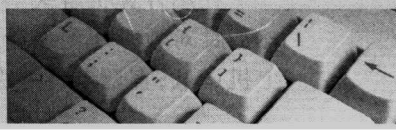
胡学钢

2007年11月于合肥工业大学

目　录

第1章

概　论



本章导读

数据结构是计算机专业重要的专业基础课程，重点讨论程序设计中常用的基本技术，因此，本课程的学习效果直接关系到后续课程的学习和软件设计水平的提高。本章首先通过实例来说明“数据结构”课程在软件设计中的作用以及在计算机专业课程中的地位，然后介绍与整个课程有关的概念、术语、算法及其描述语言和算法分析方法。

试读结束：需要全本请在线购买：　www.ertongbook.com

1.1 “数据结构”的研究内容

“数据结构”这门课程在计算机专业中有什么作用？下面从运用计算机解决实际问题的过程来讨论本课程在软件设计中的作用。

1.1.1 用计算机解决实际问题的过程

在用计算机解决实际问题时，一般要经过以下几个步骤：首先，对具体问题抽象出数学模型，然后针对数学模型设计出求解算法，选择或设计合理的数据结构存储相关数据，最后编写程序并上机调试，直至最终解决问题。

1. 解决实际问题中各环节的内容

（1）建立模型

实际应用问题可能会较复杂，例如工资表的处理问题、学生成绩管理问题、电话号码查询问题等。这些问题无论是所涉及的数据还是其操作要求都可能存在一定的差异。尽管如此，许多应用问题之间还是具有一定的相似之处的。虽然工资表和学生成绩表的具体信息（栏目）不同，但如果将两个表中的每个人的工资信息和成绩信息分别看做一个整体，则这两个表结构之间就有了某些共性。从操作方面来看，虽然对这两种表的操作存在差异，但基本操作都是相同或相似的。例如，查询一个人的工资信息和成绩信息，修改有关信息等。

正因为许多不同的问题之间存在着某些共性，使我们可以将一个具体的问题用这些共性的形式描述出来，这就是通常所说的**建立模型**。建立问题的模型通常包括所描述问题的数据对象及其关系的描述、问题求解的要求及方法等方面。建立问题模型的好处是：通过建立模型，可以将一个具体的问题转换为熟悉的模型，然后借助这一模型来实现求解。“数据结构”、“离散数学”及许多数学课程中就介绍了许多模型。例如，要描述一个群体中个体之间的关系时，可以采用“数据结构”和“离散数学”中所介绍的图的结构；要描述一个工程中的关系或进展情况，可以采用“数据结构”中所介绍的 AOV 网或 AOE 网等。即使所建立的模型没有现成的求解方法，借助于已有模型的适当组合，也比较易于构造求解方法。

数值计算类问题的数学模型一般可由数学方程或数学公式来描述。然而，对于非数值计算问题，如图书资料的检索、职工档案管理、博弈等问题，它们的数学模型是无法用数学方程或数学公式来描述的。在这类问题的处理对象中，各分量不再是单纯的数值型数据，更多的是字符、字符串以及其他用编码表示的信息。因此，首要的问题是把处理对象中的各种信息按照其逻辑特性组织起来，再存储到计算机中，然后设计出求解算法，并编写出相应的程序。

（2）构造求解算法

在建立了模型之后，一个具体的问题就转变成了一个用模型所描述的抽象问题。借助于这一模型以及已有的知识，就可以相对容易地描述出原问题的求解方法，即算法。从某种意义上说，该算法不仅能实现原问题的求解，而且还可能实现许多类似具体问题的求解，尽管这些具体问题的背景及其描述形式可能存在较大的差异。

（3）选择存储结构

在构造出求解算法之后，就需要考虑在计算机上实现求解了。首先是要选择合适的存储结构，以便将问题所涉及的数据（包括数据中的基本对象及对象之间的关系）存储到计算机中。通过本书后续各章节的学习可知，不同的存储形式对问题的求解实现有较大的影响，所占用的存储空间也可能有较大的差异。

（4）编写程序

在选择了存储结构之后，就可以编写程序了。存储形式和问题要求决定了编写程序的方法。

（5）测试

在编写出完整的程序之后，需要经过测试才能交付使用。

例如，要编写一个计算机程序以查询某市或某单位内的私人电话，要求对任意给定的一个姓名进行判断。若该人装有电话，则要求迅速找到其电话号码；否则指出该人没有装电话。

为解决这一问题，首先要构造一张电话号码登记表，表中每人登记两项信息：姓名和电话号码。在将众多的登记项合成一个数据表时，有多种不同的组织形式，查找的速度取决于表的结构及存储方式。

最简单的方式是把表中的信息按照某种次序（如登记的次序）依次存储在计算机内一组连续的存储单元中。用高级语言表述，就是把整个表作为一个数组，每个人的信息（即一个人的姓名和电话号码）是数组的一个元素。查找时从表的第一项开始，依次查对姓名，直到找出指定的姓名或是确定表中没有要找的姓名为止。这种查找方法对于一个规模不大的单位或许是可行的，但对于一个有几十万乃至几百万私人电话的城市就不适用了。因此，一种常用的做法是把这张表按姓氏或姓氏笔画排列，并另造一张姓名索引表，这类似于汉语字典的组织形式。对这样的表的查找过程可以先在索引表中查对姓氏，然后根据索引表中的地址到登记表中核查姓名，这样查找登记表时就无需查找其他姓氏的名字了。因此，在这种新的结构上产生的查找方法就会更有效。这两张表便是我们为解决电话号码查询问题而建立的数学模型。这类模型的主要操作是按照某个特定要求（如给定姓名）去对登记表进行查询。诸如此类的还有人事档案管理、图书资料管理等。在这类文档管理的数学模型中，计算机处理的对象之间通常存在一种简单的线性关系，故称这类数学模型为线性的数据结构。

2. “数据结构”课程涉及求解过程的主要步骤

（1）与建立模型的关系

数据结构课程中介绍了许多基本的数据结构模型及其运算实现。例如，线性表、栈和队列、树和二叉树、图、二叉排序树、堆等。通过学习，不仅应掌握这些基本内容及其应用，还应能根据

实际问题选择合适的模型。

(2) 与算法设计的关系

课程中对每种结构都讨论了相应的基本运算的实现,并且其中的一些算法是非常经典的,掌握这些基本运算的实现方法有助于进行更为复杂的算法设计。

(3) 与选择存储结构的关系

课程中对每种结构都讨论了其具体存储结构及其对运算实现的影响。例如,在第2章中所介绍的对顺序表作插入和删除操作,平均需要移动表中一半的元素,而采用链表结构则不需要移动元素。通过对这些内容的学习和比较,使学生在实际应用时能根据问题的要求熟练地选择乃至设计合理的存储结构。

(4) 与编程之间的关系

在实现各结构的算法中涉及许多具有代表性的设计方法。通过对这些方法的学习有助于编程技术的提高。

综上所述,"数据结构"课程对提高软件设计人员的软件设计水平有较大的影响。也正因为如此,这一课程在计算机专业课程中具有极其重要的地位,绝大多数学校和研究单位的计算机专业研究生入学考试都将这一课程定为必考课程之一。

1.1.2 学习"数据结构"的意义

"数据结构"作为一门独立的课程在美国是从1968年开始的。在这之前,它的某些内容曾在其他课程中有所涉及。1968年,美国一些大学计算机系的教学计划中,虽然把"数据结构"规定为一门课程,但对其内容并未作明确规定。随后,"数据结构"这个概念被扩充到包括网络、集合代数、格、关系等理论中,从而演变成现在的"离散结构"的内容。然而,由于数据必须在计算机中进行处理,因此,不仅要考虑数据本身的数学特性,而且还要考虑数据的存储结构,这就进一步扩大了数据结构的内容。

"数据结构"是计算机科学中一门综合性的专业基础课程。数据结构的研究不仅涉及计算机硬件(特别是编码理论、存取方法等)的研究范围,而且和计算机软件的研究有着密切的关系,无论是编译程序还是操作系统都涉及如何组织数据,使检索和存取数据更为方便的问题。因此,可以认为"数据结构"是介于数学、计算机硬件和软件三者之间的一门核心课程。在计算机科学中,"数据结构"不仅是一般程序设计的基础,而且是设计和实现编译程序、操作系统、数据库系统及其他系统程序和大型应用程序的重要基础。

目前,"数据结构"不仅是大学计算机专业的核心课程之一,而且还是一些非计算机专业的主要选修课程之一。

随着计算机应用领域的扩大和软、硬件技术的发展,"非数值性问题"显得越来越重要。据统计,当今处理非数值性问题占用了90%以上的机器时间。从前面的例子可以看到,解决此类问题的关键已不再是数学方法,而是设计出合适的数据结构。瑞士计算机科学家沃斯(N. Wirth)曾以"算法 + 数据结构 = 程序"作为他的一部著作的名字。可见,程序设计的实质是针

对实际问题选择或设计好的数据结构和好的算法。因此，若仅仅掌握几种计算机语言和程序设计方法，而缺乏数据结构的知识，则难以应对众多复杂的课题，且不能有效地利用计算机解决实际问题。

1.2 基本术语

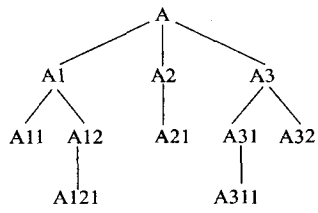
数据是信息的载体，是指能够输入到计算机中，并被计算机识别、存储和处理的符号的集合。数据的形式较多，例如前面所述的工资表、学生成绩表、一个家族关系、一个群体中个体及其之间的关系等，如图 1.1 所示。

编号	姓名	基本工资	奖金	…	…

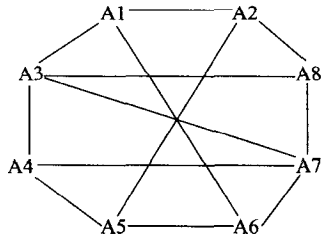
(a) 工资表示例

序号	学号	姓名	成绩	备注

(b) 成绩表示例



(c) 家族关系示例



(d) 群体间关系示例(连线表示相互认识关系)

图 1.1 数据示例

虽然这些数据的形式及运算存在较大的差异，但可以找出其中的共性：都是由具有独立意义的若干个体所组成的，个体之间存在着某些关系。对这些数据的运算也有某些相似之处。例如，在家族关系中，组成数据的基本个体是个人信息，其中的个人与个人之间存在着多种关系，例如父子关系、兄弟关系、祖先－后代关系等，其中有些关系是直接表示出来的，还有一些关系则是隐含的。对家族关系数据，通常要涉及查询特定个体间的关系、插入和删除个体等操作。

数据结构课程中主要讨论这些具有共性的内容。为便于讨论，先给出有关概念。

1. 数据元素

数据中具有独立意义的个体。例如工资表中的个人工资信息、成绩表中的学生成绩信息、家族关系中的个人基本信息等，在有些场合下，也称为**元素、记录、结点、顶点**等。

2. 字段(域)

虽然将具有独立意义的个体用元素来表示,但在许多情况下还需要特定个体的具体信息,因而涉及元素的字段信息。字段是对元素的详细描述,通常情况下,元素可能包含多个字段。例如,在图 1.1 所示的成绩表中,每个元素包括序号、学号、姓名、成绩、备注 5 个字段,分别描述了每个学生的有关信息。

3. 数据结构

在图 1.1 所示的几个数据示例中,元素之间各自具有一定的构成形式,这些构成形式被称为**数据结构**。**数据结构**是指组成数据的元素之间的结构关系。在图 1.1 中,图(a)和图(b)中的元素依次排列,构成**线性关系**;图(c)中的元素按树的形式构成**树状结构**;图(d)中的元素构成**图结构**。线性结构、树状结构和图结构是数据结构中的几类常见的数据结构形式。如果数据中的元素之间没有关系,则构成**集合**,这也是一种结构。

通常这几类结构称为**逻辑结构**,因为只考虑了元素之间的逻辑关系,而没有考虑到其在计算机中的具体实现。

在用计算机解决数据结构问题时,需要完成如下任务:

① 数据结构的**存储**。为所涉及的数据结构选择一种存储形式,并将其存储到计算机中,这样就得到了数据结构在内存中的**存储结构**(也称为**物理结构**)。一种逻辑结构可能会有多种存储结构。例如,可以采用顺序存储,也可采用链表形式存储。不同存储结构上实现的运算的性能可能有一定的差异。

② 数据结构**运算**的实现。在选择了数据结构的存储结构之后,就可以实现所给出的运算了。在本课程中,以 C 语言作为算法描述语言,因而算法就以 C 语言中的函数形式给出。

由此可见,对一种数据结构,需要涉及其**逻辑结构、存储结构**和**运算三个方面**。也就是说,对每种结构都要注意这 3 方面的联系。

由于不同的存储形式对算法的时间性能、空间性能等有较大影响,即使是相同的存储结构也可能存在不同的算法实现,因此,需要研究和解决这样的问题:何种存储结构更为合适?什么算法更有效?为此,需要对算法进行分析,通过分析,可以知道所实现的算法的性能及所选择的存储结构是否符合要求。有关算法分析的内容将在本章的后面部分讨论。

1.3 算法描述及分析

1.3.1 算法描述语言概述

什么是算法?简单地说,算法就是某类问题的求解方法。下面给出另一个关于算法的描述:**算法**就是一段程序,该程序段对给定的输入可在有限的时间内产生出确定的输出结果。

算法可采用多种描述语言来描述，例如自然语言、计算机语言或某些伪语言。各种描述语言在对问题的描述能力方面存在一定的差异。例如，自然语言较为灵活，但不够严谨，而计算机语言虽然严格，但由于语法方面的限制，显得灵活性不足。因此，许多教材中采用的是以一种计算机语言为基础，适当添加某些功能或放宽某些限制而得到的一种语言，这些语言既具有计算机语言的严格性，又具有某些灵活性，同时也容易上机实现，因而被广泛接受。目前的许多数据结构教材采用类 Pascal 语言、类 C ++ 或类 C 语言作为算法描述语言。

本教材中选用的是以 C 语言为主体的算法描述语言，为了便于实现，算法主要以 C ++ 环境中的 C 语言来描述，并适当扩充一些功能或放宽某些限制。所涉及的算法以 C 语言的函数形式给出。

考虑到读者已经学过 C 语言，因而不再对 C 语言部分作详细描述。下面补充说明所扩充的一些功能。

1. 输入和输出语句

由于 C 语言中的输入和输出语句的形式对数据的类型有一定的限制，因此，本书中采用 C ++ 中的独立于数据类型的输入和输出语句。

(1) 输入

```
cin >> x;
```

其功能是读入从键盘输入的一个数，并赋值给相同类型的变量 x。其中变量 x 的类型可以是整型、浮点型、字符型等不同类型。

该语句可用下面的形式依次输入多个不同类型的变量：

```
cin >> x1 >> x2 >> x3 >> x4 >> x5;
```

(2) 输出

```
cout << exp;
```

其功能是将表达式 exp 的值输出到屏幕上。其中表达式 exp 的类型也可以是整型、浮点型、字符型等不同类型。

该语句也可用下面的形式依次输出多个不同类型的表达式的值：

```
cout << exp1 << exp2 << exp3 << exp4 << exp5;
```

2. 最小和最大值函数 min( ) 和 max( )

```
datatype min(datatype exp1, datatype exp2,..., datatype expn)
datatype max(datatype exp1, datatype exp2,..., datatype expn)
```

分别返回表达式 expi(i = 1,2,…,n) 中的最大(小)的值。其中元素类型 datatype 可以是整型、浮点型、字符型等各种可比较的类型。

3. 交换变量的值

```
x1↔x2;
```

试读结束：需要全本请在线购买： www.ertongbook.com