

“数据结构”教学与教材研究

张乃孝

摘 要：本文回顾了著者三十多年来从事“数据结构”教学与研究的主要经历，重点介绍了对该课程的教材建设方面的主要工作，指出与时俱进、精益求精地编写教材是提高教学水平的基础和关键。

关键词：数据结构；算法；程序设计语言；程序设计方法；教材

一、前言

1946 年 2 月 14 日，世界上第一台电子数字计算机 ENIAC 在美国宾夕法尼亚大学诞生。早期计算机主要用于数值计算，处理的对象是“无结构”的数据（例如整数和浮点数），它们和处理这些数据的程序（根据计算机指令系统编写的代码）都采用二进制表示形式存储在计算机的存储器中。20 世纪 50 年代开始的“程序设计语言”研究，改变了原始的使用机器语言编程的方式，语言的“使用手册”给计算机的使用者提供了一个非常高级的“虚拟机”，使得程序员可以方便快捷地描述需要的数据和处理数据的程序，然后通过语言的“编译器”把它们成功地转换为计算机内部的二进制代码。高级语言的研究成果，打破了计算机只能进行科学计算的限制。“语言编译系统”通过计算机成功地完成从高级语言的模型到计算机硬件语言模型的转换，打开了计算机系统软件研究的大门；同时也提出许多相对比较复杂的结构化数据的需求（例如栈、散列表和二叉树等），促进了数据结构的研究和发展。

“数据结构”的概念最早是由 C. A. R. Hoare 和 N. Wirth 在 1966 年提出。大量关于程序设计理论的研究表明：为了系统而科学地构造大型复杂的程序，必须对这些程序中所包含的数据结构进行深入研究。

1968 年，美国教授 D.E.Knuth 在他的名著《计算机程序设计技巧》（第 1 卷 基本算法 第二章信息结构）中首次系统地研究并整理了当时经常使用的主要数据结构与相关的算法，为数据结构课程的开设提供了丰富的素材（他本人也因此书的成就，在 1974 年获得计算机界最高科学成就奖“图灵奖”）。

自 20 世纪 70 年代起，“数据结构”在西方国家的大学生中，被普遍列为计算机本科的必修课程。

二、不同时期的教材

1978 年著者已有 10 年从事系统软件开发的丰富经验，参加了北京大学计算机系的筹备和创建。在担任数据库教研组组长期间，按系主任张士龙教授的安排，负责“数据结构”等新课的建设。从此围绕数据结构开展的工作，包括学习与研究、讲课与编写教材等，三十多年一直没有停息。其中花费时间和精力最多的是根据教学和科研的需要编写了下面 4 本教材（以 8 种不同版本出版）。

1. 第一本教材：《数据结构》

1979 年教育部在南京大学召开了第一次全国计算机系教学大纲研讨会，著者带着起草的“数据结构教学大纲”和“数据库教学大纲”与系主任一起参加了会议。会上充分肯定了我们的工作，并建议我们分工负责编写数据结构教材。根据这个大纲的修改稿，著者组织教研组内的老师共同编写了第一本《数据结构》讲义。1980 年起，这本讲义在校内外（包括南京大学、中山大学等国内著名高校）广泛试用，三易其稿。1987 年由高等教育出版社正式出版（此书 1992 年获国家教委颁发的全国优秀教材奖）。

2. 第二本教材：《数据结构基础》

1985 年，“北京市自学考试委员会”开设计算机专业。作为数据结构课程的考试委员，著者邀请杨冬青和邵维忠两位老师共同编写了这本自学考试教材。1991 年由北京大学出版社出版，第二年台湾儒林出版公司用繁体字出版。

3. 第三本教材：《数据结构——C++与面向对象的途径》

20 世纪 90 年代，面向对象的语言和方法开始流行。

张乃孝，北京大学数学学院教授。

根据教学和科研的需要, 著者与裘宗燕老师合作编写了该教材。1998 年作为国家“九五”重点教材由高等教育出版社出版, 2001 年出修订版。

4. 第四本教材:《算法与数据结构》

1998 年, 著者由北京大学校长聘任, 主持全校理科主干基础课“算法与数据结构”, 考虑到不同专业的需要, 组织理科教师共同编写了一本适合理科各专业通用的新教材。该书列为“面向 21 世纪教材”, 2002 年由高等教育出版社出版, 获评“北京市高等教育精品教材”; 2006 年出第二版, 列为“十一五”国家级规划教材, 获评教育部“普通高等教育精品教材”; 2012 年再次修改并附著者教学光盘, 出第三版。

回顾三十多年来围绕数据结构教学方面的工作, 深深体会到与时俱进、精益求精地编写教材是提高教学水平的基础和关键。

三、数据结构教材需要与时俱进

计算机科学是一门高速发展的新兴科学, 它的研究内容和研究方法都在不断发展。“结构”可以解释为: (1) 把某些成分(成员、元素、原子等)按一定的规律或方式组织在一起的实体; 或(2) 把某些成分组织在一起的方式。“数据结构”从字面上可以理解为就是以数据为成员的结构。在早期关于数据结构的论文中, 一个数据结构多数情况下是指一个“实体”, 而不是指“方式”。用通俗的程序语言的术语来讲, 一个数据结构就可以看成一个结构化的数据。然而, 计算机科学研究数据结构的目的是为了在计算机中有效地表示和处理客观世界的各种不同对象。所以我们关心的是这些数据如何存储在计算机中, 并且能有效地完成各种需要的操作。随着计算机科学的发展, 对于数据结构的教學与研究也逐步从“实体”, 提高到“方式”, 直到“抽象数据类型的实现”。

1. 教材应该正确反映计算机科学的发展水平

前面提到的第一本教材, 基本反映了 20 世纪 70 年代的认识水平。当时数据结构的许多概念还十分模糊, 即使像“栈”和“队列”这些最基本的结构, 它们的操作定义都不完全统一。许多教材对于什么是“数据结构”都没有解释。我们考察了 20 世纪 70 年代有影响的几本著作。其中 H. A. Maurer 用一个二元组 $B = \langle K, R \rangle$ 来形式地定义一个数据结构 B, 其中 K 是结点有限集合, 而 R 是 K 上的关系的有限集合。C. C. Gotlieb 和 L. R. Gotlieb 则将数据结构的定义扩充成一个五元组: $\langle V, O, G, M, S \rangle$ 。其中 V 是所讨论的结构中成员取值的集合, O 是结构中成员可执行的运算的集合, G 是两个构成名字的文法, M 是结

构中各成员存放位置的集合, S 是 $L(G) \Rightarrow M$ 的映射。

根据数据结构研究的目的是和应用的需要, 我们认为提到一种数据结构离不开以下三个方面: (1) 构成数据结构的成员之间固有的逻辑关系; (2) 将数据存储在计算机中的表示方法; (3) 在计算机中对数据结构进行的运算或处理。将这三方面分别简称为数据的逻辑结构、存储结构和运算, 所以在第一本教材中我们明确采用这三者的统一(三位一体)来非形式地定义“数据结构”的概念。

第二本《数据结构基础》以上一本教材内容为基础, 根据 N. Wirth 教授提出的“算法+数据结构=程序”关系, 把程序理解为在数据的某些特定的结构和表示的基础上对于算法的描述。算法与数据结构是程序设计中相辅相成、不可分割的两个方面。为了适合于自学考试大纲的要求, 参考了 A. V. Aho 教授 20 世纪 80 年代的教材, 采用以数据结构为主线、算法为辅线的结构编写, 使得内容更加紧凑、重点更加突出。

第三本教材《数据结构——C++与面向对象的途径》是在面向对象的语言和方法开始流行的 20 世纪 90 年代, 采用面向对象的设计方法讲解数据结构的内容。参考 Budd 的工作, 由简到繁、从易到难, 系统地引入各种抽象数据类型概念和实现, 并在全书最后, 用类图方式总结了各种经典的抽象数据类型在教材中的相互关系。

最后一本《算法与数据结构》参考了 Kurt Mehlhorn 等人的观点, 把“数据结构”定义为“抽象数据类型的物理实现”。提出“物理实现”的意图是强调本课程关心的“实现”应具体到可以用计算机的两个最重要的物理量(主机的运行时间和内存的存储空间)来权衡。这一观点突出了抽象数据类型在数据结构教学中的地位, 包含了数据结构与面向对象技术的内在联系。使读者可以从更高的层次理解数据结构与算法的关系, 也容易解释数据的逻辑结构、存储结构与运算的三者关系。

后两本教材都反映了 20 世纪 90 年代的理解水平。其共同之处是: 都强调了数据结构是“抽象数据类型的实现”; 前一本使用的是面向对象的实现方法, 而后一本为了突出讲解实现的物理效率, 没有采用面向对象的方法。

2. 教材内容既要相对稳定又要逐步更新

需要指出的是, 尽管计算机科学的发展使得数据结构的地位和作用产生了许多变化, 但是数据结构学习的目的并没有大的改变。所以教材的内容是基本稳定的。

第一本教材按“逻辑结构、存储结构、运算和应用”四个层次的结构组架架构。全书共 18 章, 除第一章概论外分为四大部分: 第一部分是线性结构, 包括顺序表、链表与动态存储管理、串、内排序和线性表的检索等五章; 第二部分是树形结构, 包括树形结构的概念、树形结构的存

储、二叉树周游算法、树目录和树形结构的其他应用等五章；第三部分是复杂结构，包括图和多维数组与广义表两章；第四部分是文件结构，包括顺序文件、散列文件、索引顺序文件、倒排文件和外排序等五章。全书概念清楚、内容丰富、体系完整。

第二本作为自学考试教材，内容在第一本的基础上加以精简，并增加集合与字典结构，把检索归入集合的基本运算。在结构上更加强调基础、突出重点、适合自学。全书共分 8 章。第一章通过分析一个实际问题的求解过程，引入抽象数据类型、数据结构和算法等重要概念作为全书的引论；第二章到第五章分别讨论了表、树、集合和图等常见的各种数据结构，一般均以抽象数据类型引路，重点讨论抽象数据类型在计算机中各种不同的实现方法；第六章对链接表示所需要的动态存储管理问题作了系统的阐述；第七章综述了外存上数据结构的各种组织方式；第八章给出内排序和外排序的各种算法。

第三本由于采用了面向对象的方法，在内容上做了较大调整。增加了面向对象的方法入门和优先队列。全书共分 12 章：第一章，绪论；第二章，C++与面向对象初步；第三章，字符串，本章定义了一种更安全可靠的字符串类型，同时也以字符串做例子，讨论数据抽象和封装的有关问题；第四章，向量，本章建立了一种安全可靠的向量数据类型，还给出了几个主要的向量排序算法；第五章，动态数据结构——链表，主要讨论了各种常用的链表结构及其实现方法；第六章，栈和队列，介绍了栈和队列的抽象概念、具体实现及其应用；第七章，树和二叉树，介绍了树和二叉树的概念，重点介绍二叉树的实现及树结构用于快速检索的一些技术；第八章，优先队列，主要介绍了堆和斜堆的概念以及通过它们实现优先队列的方法；第九章，集合和字典；第十章，散列表；第十一章，图；第十二章，文件。在附录中用类图方式给出本书介绍的主要抽象数据类型及其相互关系图。

第四本书作为北京大学主干课“算法与数据结构”的通用教材。全书共分以下 10 章。第一章绪论；第二章线性表；第三章字符串；第四章栈与队列；第五章二叉树与树；第六章集合与字典；第七章高级字典结构；第八章排序，第九章图；第十章算法分析与设计，主要给读者概括地介绍算法的分析和设计的主要技术。本书在编写中注意到知识模块的独立性和相关性，不同专业的学生可以根据不同的需要进行组合使用。

在我们后编写的两本教材中，都大幅度减少了存储管理和文件系统的内容，其主要原因是它们应该分别属于“操作系统”和“数据库”的教学范围。“文件系统”又称“物理数据库”，主要讲解数据库的物理实现。在我们新编的教材中，只给出了与“字典类型”紧密相关的“索引文件”

及“散列文件”介绍，也没有给出实现代码。

3. 算法描述语言要配合教学的需要

数据结构的教学内容原本独立于任何一种特定的程序设计语言，但是这门课程的教与学又离不开程序设计语言的支持。在这里语言是一种教学的工具。工具的选择应该有利于表达数据结构的基本思想与算法的设计方法，有利于算法的分析与设计，并且简单明了，便于老师讲学和学生学习理解。前面讲的四本教材，因为在不同的时期创作，根据不同的学术观点，针对不同的读者，所以也选择了不同的算法描述语言。

在编写第一本教材时，由于国内当时主要流行的程序设计语言是 Basic、Fortran 和 Algol60，它们不适用于描述数据结构中的许多算法，为此我们在书的附录中给出“关于书写算法的若干规定”，除了过程语言允许的基本的控制语句外，还为描述链表操作和存储管理引进了一些专用过程，以便于描述动态的存储分配和内存空间的管理功能。

第二本教材根据“算法+数据结构=程序”的观点编写，当时 N. Wirth 教授提出的 Pascal 语言已经在国内流行。它的指针可以方便地描述链表的操作，但是在描述存储管理时显得不够灵活。所以该书对其进行了简单的扩充并加入汉字的注释，称为伪 Pascal 语言。

第三本教材采用面向对象的设计方法讲解数据结构的内容。所以首选当前国内最流行的面向对象的程序设计语言 C++来描述。学生在学习数据结构的同时，又加深了对于面向对象方法的理解，提高了使用 C++语言编程的能力。

使用了良好的面向对象的 C++描述，程序表面的可读性很好，但内涵十分丰富，例如各种构造函数和析构函数的自动选择和运行，各种继承和多态功能的动态处理等，所以要具体分析一个独立算法的时间和空间的代价往往比较困难。而这些内容恰恰是学习数据结构的一个重要目标，也是许多专业学生学习计算机的主要因素。加上教学计划安排的课程顺序、学时要求等因素，所以我们在编写第四本教材时选择了 C 语言描述。

C 语言虽然是一个小语言，但具有丰富的表达能力，这使它简单、易学，又能满足基本的教学需求。另外，C 语言是一个过程语言，用它描述的算法语义清晰、确定可行。特别重要的是，C 比较低级，使用 C 描述的算法，其时间和空间代价分析最直观、准确。

四、精品教材应该精益求精

如前所述，我们希望所编的教材与时俱进，跟上计算机科学的发展步伐，使得每本教材独具特色、体系完整、结构合理。与此同时，著者对教材的每个重要环节，

包括概念的定义、思想的陈述、难点的分解、算法的设计与分析方法以及教学的方法甚至书面排版的格式等,都经过认真的考虑和细心的安排。下面举几个简单的例子,从中可见一斑。

1. 概念准确,语言流畅

对于基础课的教材,概念准确是十分重要的。然而由于计算机科学十分年轻,发展又快,使得许多概念在文献中没有统一的定义。例如“文件”和“记录”这两个概念在外存的讨论和内排序中都使用,但是意义不同;另外关于二叉树的“高度”和“深度”的定义,不同的数据结构教材可能不同;关于“满二叉树”的定义,不同教材的差别可能很大,等等。

为了尽可能给出所有概念的准确定义,我们花费了大量时间,反复查阅了多种不同的文献,包括数据结构和离散数学的各种教材,经过认真分析、比较,给出我们的理解,在教材最后整理了所有名词的索引。并且在第一次出现的位置注释出可能不同的定义。

语言流畅是提高教材可读性的基础。我们的几本教材多数都是从校内试用讲义开始,反复修改,出版后也再版甚至三版,对于局部小修改有的就利用加印的机会进行。其中修改最多的是语言文字,包括标点符号的修改,通过润色使其更加准确、流畅。应该承认,广大学生的参与为本书文字的加工作出了很大的贡献。

2. 重点突出,难点讲透

数据结构的内容十分丰富,难度也比较大。著者的主要工作是准确讲解该讲的内容,把重点要讲透彻,把难点加以分解,使得读者容易学习和理解。数据结构教学的重点是讲解经典抽象数据类型的各种实现方法;难点则是算法的设计、实现和分析比较。

为此,我们在教材中每章的最后都编写了“小结”,明确指出本章的难点和重点。为了便于教学,我们的教材特别注意将讲解知识和培养能力并行:不但讲解一个抽象数据类型可以采取的常用表示形式,同时还指出不同表示的特点和各种表示使用的环境;不但讲解如何在选定的存储表示上正确实现抽象数据类型要求的各种操作,同时还强调不同存储表示对算法效率的影响;不但讲解单个算法的设计和实现方法,同时也强调同类算法之间的共性,并且在教材最后对学到的主要算法进行总结,以提高学生利用学到的知识去解决实际问题的能力。

3. 算法逐步求精,程序简明可读

设计一个算法时,如何从思想出发,逐步细化直到变成程序语言的代码?阅读一个程序代码时,如何分解

一个算法程序,正确理解这个程序完成的算法功能?这是学习数据结构课程的学生的共同问题,也是一般数据结构教材难以书写清楚、教师难以讲解明白的问题。

为了帮助老师和学生解决上述问题,我们在教材的第一章,都用一个实际问题生动地介绍了使用计算机求解的全过程,还增加了从算法思想到代码实现逐步求精的具体例子。对于教材中比较难理解的算法,一般都是先提出问题,然后对实例进行分析,给出解决的方法,再整理出算法的思路,最后通过逐步求精得到程序代码;而在讲解这个程序代码时,利用电子课件的灵活方式,结合算法的思想,自顶向下进行分解。为此,我们对教材的代码的结构如何与算法思想更加一致,语句(特别是存在多种不同循环语句时)的选择尽可能简单明了,如何引进局部变量使得程序更为简短等,都进行反复改进。力求书中提供的程序具有高度的简明性和可读性。

我们知道程序是没有“标准答案”的,但是高水平的程序员能够根据语言的风格和编程的艺术,编写出更美、更能体现算法魅力的艺术珍品。我们为此尽力而为。

4. 利用多种方式,提高学习兴趣和教学质量

除了上述的工作以外,著者还编写出版了《数据结构与算法学习辅导及习题详解》和《算法与数据结构(第2版)学习指导与习题解析》;作为《算法与数据结构——C语言描述(第三版)》的附件,整理并出版了《“算法与数据结构”课程录像》的完整光盘和全部课件;还完成了一个《数据结构、算法和问题求解》三维的网络课件;设计了一个《数据结构中算法的可视化演示》软件框架等。

在从事数据结构教学的同时,著者还发表了十多篇与数据结构相关的研究论文,其中关于“三叉树结构的设计和实现”研究成果获得国家专利。指导学生提出一种基于对象的数据结构教学小语言——ALL,设计了数据结构中主要算法的ALL语言程序,并且实现了ALL语言程序到目前流行语言(JAVA、C++、C等)的自动转换。

为帮助学生深入理解算法与数据结构的精髓、提高学生的学习兴趣,我们还在北京大学教学网上开辟专栏,引导学生进行广泛讨论,受到学生的热烈欢迎。通过这些讨论,大大丰富了学生的知识范围,激发了学生自主学习的热情,也弥补了书面作业和上机作业的局限。

光阴似箭,日月如梭,著者围绕“数据结构”教学和研究的三十多年正是数据结构在我国从引介到发展的三十多年。在此前,著者曾有十年“系统软件”开发经验,因而一接触“数据结构”就产生了极大兴趣,能够很快体会到其中的真谛;加上个人长期从事“程序设计

语言”和“软件理论”的研究，教学与科研相辅相成、互相促进。在辛勤耕耘的同时，个人也享受着授业解惑的快乐，体会到人生的价值。

参考文献：

- [1] 许卓群, 张乃孝, 杨冬青等. 数据结构[M]. 北京: 高等教育出版社, 1987.
- [2] 张乃孝等. 数据结构基础[M]. 北京: 北京大学出版社, 1991; 台湾儒林出版公司(繁体版), 1992.
- [3] 张乃孝, 裴宗燕. 数据结构——C++与面向对象的途径[M]. 北京: 高等教育出版社, 1998 第一版, 2001 修订版.
- [4] 张乃孝等. 算法与数据结构——C 语言描述[M]. 北京: 高等教育出版社, 2002 第一版, 2006 第二版, 2012 第三版(附教学光盘).
- [5] 张乃孝. 算法与数据结构(第2版)学习指导与习题解析[M]. 北京: 高等教育出版社, 2009.
- [6] Knuth D E. The Art of Computer Programming[M]. Addison Wesley Publishing Company, 1973.
- [7] Horowitz E, Sahni S. 数据结构基础[M]. 程惟宁译. 北京: 国防工业出版社, 1983.
- [8] Wirth N. Algorithms + Data Structures = Programs[M]. Prentice Hall, 1976.
- [9] Maurer H. A. Data Structures and Programming Techniques[M]. Prentice Hall, 1977.
- [10] Gotlieb C C, Gotlieb L R. Data Types and Structures[M].

Prentice Hall, 1978.

- [11] Aho A V, Hopcroft J E, Ullman J D. Data Structures and Algorithms[M]. Addison Wesley Publishing Company, 1983.
- [12] Mehlhorn K, Tsakalidis A. Data Structures .Handbook of Theoretical Computer Science [M]. Elsevier, 1990.
- [13] Budd T A. Classic Data Structures in C++[M]. Addison Wesley Publishing Company, 1994.
- [14] 张乃孝, 于晓迪. 有关 C 语法形式化中若干问题的探讨[J]. 计算机工程与应用, 1985(2): 1-5.
- [15] 张乃孝. 数据结构体系分析[J]. 计算机研究与发展, 1988, 25(5): 36-40.
- [16] 张乃孝. 知识结构的三叉树表示及逻辑推理的实现[J]. 计算机学报, 1990, 13(1): 32-41.
- [17] 张乃孝. 三叉树结构及其实现[J]. 计算机研究与发展, 1993, 30(1): 50-54.
- [18] 张乃孝, 蒋凌霄. ALL——算法与数据结构教学小语言[J]. 计算机科学, 2003, 30(11): 178-180.
- [19] 孙玉方, 张乃孝. 实用 C 语言程序设计[M]. 北京: 北京大学出版社, 1989; 台湾儒林出版公司(繁体版), 1992.
- [20] 张乃孝. 数据结构与算法学习辅导及习题详解[M]. 北京: 电子工业出版社, 2004.

[本文得到国家自然科学基金 61202069 项目资助]

[责任编辑：余大品]

(上接第 94 页)教育思想观念, 改革人才培养模式, 为社会输送大量高素质创新创业型人才以适应社会发展的需要。

参考文献：

- [1] 胡敏, 陈立俊. 基于 SWOT 分析的大学生创业现状及创业教育对策研究[J]. 教育探索, 2008(11).
- [2] 刘帆, 陆跃祥. 中美两国高校创业教育发展比较研究——基于统一标准[J]. 中国青年研究, 2008(5).
- [3] 岳晓东. 大学生创新能力培养之我见[J]. 高等教育研究, 2004(1).
- [4] 曹胜利, 雷家骕. 中国大学生创新创业教育发展报告[M].

沈阳: 万卷出版公司, 2009.

- [5] 吴金秋. 高校推进创新创业教育的理念定位[N]. 中国教育报, 2010-08-14.
- [6] Powers, J. and McDougall, P. University startup formation and technology licensing with firms that go public: a resource based view of academic entrepreneurship. Journal of Business Venturing, 20, 3, 2009
- [7] Lockett, A. and Wright, M. Resources, capabilities, risk capital and the creation of university spinout companies, Research Policy, 34, 7, 2005

[责任编辑：李文玲]