

高等学校教材

数据结构与算法

徐绪松



高等教育出版社

高等学校教材

数据结构与算法

徐绪松

高等教育出版社

内 容 提 要

本书全面系统地阐述了各种数据结构,以及应用于各种结构的算法及实例,同时介绍了基本的算法分析方法和算法设计技术。全书渗透了作者多年的教学经验和研究心得。

作者将数据结构与算法紧密结合,以“方法——实施方法的数据结构——施加在数据结构上的算法——程序”的模式展开。章末的习题可以起到概括内容、掌握概念、巩固知识的作用。书中各算法的 C 语言源程序代码和书中内容的 ppt 课件将上载到高等教育出版社计算机教学资源网(<http://cs.hep.com.cn>)上,供读者学习参考。

本书可以作为高等院校计算机应用专业、信息与计算科学专业、电子信息类专业以及信息管理、工程管理、应用数学等专业的教材和参考书,也可供从事上述专业的工作人员自学参考。

图书在版编目(CIP)数据

数据结构与算法/徐绪松. —北京:高等教育出版社,2004.7

ISBN 7-04-014618-5

I. 数... II. 徐... III. ①数据结构-高等学校-教材②算法分析-高等学校-教材 IV. TP311.12

中国版本图书馆 CIP 数据核字(2004)第 061978 号

策划编辑 倪文慧 责任编辑 武林晓 市场策划 陈 振
封面设计 于文燕 责任印制 孔 源

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100011
总 机 010-82028899

购书热线 010-64054588
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>

经 销 新华书店北京发行所
印 刷 北京东光印刷厂

开 本 787×1092 1/16
印 张 24.75
字 数 520 000

版 次 2004 年 7 月第 1 版
印 次 2004 年 7 月第 1 次印刷
定 价 27.00 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

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人给予严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话：(010) 58581897/58581896/58581879

传 真：(010) 82086060

E - mail：dd@hep.com.cn

通信地址：北京市西城区德外大街4号

高等教育出版社打击盗版办公室

邮 编：100011

购书请拨打电话：(010)64014089 64054601 64054588

序

“数据结构”是计算机科学的一门核心课程,随着人类社会进入信息化时代,计算机更加广泛地应用于各个领域,包括科学计算、信息系统等,很多专业都开设了“数据结构”课程。从知识范畴的视角,该课程给出了许许多多的计算机编程技巧;从培养能力的角度,该课程培养了学生的逻辑思维、形象思维和结构化思维。它是大学本科一门重要的基础课。

在诸多的同类教材中,徐绪松教授的《数据结构与算法》一书具有自己的独特之处,为此,本人谨向读者推荐这本书。概括起来,可将本书的特点列举如下:

1. 结构严谨。在整体结构上,作者依据数据类型的性质及相关性,将全书分为5篇16章,以数据的逻辑结构为脉络,对数据进行了全局性组合,章节关联体现了很强的逻辑体系。

2. 将数据结构与算法结合。数据结构是描述非数值问题的数学模型,算法是施加在数据结构上解决某一类问题的办法,作者将两者结合,提出了一个研究非数值问题的研究模式,并贯穿于全书。

3. 形象思维。全书运用图示法,将问题的计算机存储展示在读者面前,引导学生想像计算机实现算法的进程,了解数据的操作,从形象思维获得解决问题的自然思维,再用接近人的自然思维的自然语言进行描述,使很多问题迎刃而解。

4. 结构化思维。书中的每个算法均按模块化、结构化进行设计,不仅设计的算法结构清晰,且使读者在算法设计中培养了结构化的思维,这是难能可贵的。

5. 案例丰富。全书搜集了大量经济管理的应用实例,并进行了系统的布局,有很强的实用性。这些案例用来说明数据结构的属性,也用来说明施加在数据结构上的算法,体现了“程序=数据结构+算法”。这样做使得本教材更加生动而有吸引力,不仅引起学生学习的兴趣和探索的欲望,还达到了学以致用目的。

6. 有创见性的研究成果。本书正式开辟了公式分组排序、公式分组查找的章节,这是作者有创见性的研究成果,它们分别将当前排序、查找的计算复杂度由 $O(n\log n)$ 降到 $O(n)$ 。归结以上这些特点,该书立于一个新的高度,具有先进性、科学性、创新性和实用性。以本书作为教材,将获得良好的教学效果,讲授生动有趣,学生易于理解和接受,并能在潜移默化中开发学生智力,培养学生的创造性思维。

刘经南教授
武汉大学校长
中国工程院院士
2004年5月

前 言

随着计算机的发展,计算机已广泛应用于各个领域。这些年来兴起了管理信息系统、智能化的决策支持系统以及复杂科学与管理等,又开创了计算机应用的新领域。它们不再拘泥于科学计算,而更多的是数据处理和大型离散问题的程序设计。这种非数值问题,不是一个简单的编程,它将涉及到数据之间的结构关系、计算机内的存储表示以及在各种结构上的非数值运算(如插入、删除、排序、查找等)。非数值问题占用了90%以上的计算机时间,这就需要研究数据结构及其相应的算法。

数据结构与算法之间有着密切的联系,不了解施加于数据上的算法就无法决定数据结构;反过来,算法的结构和选择在很大程度上依赖于作为基础的数据结构。可以说,数据结构为算法提供了工具,而算法则是运用这些工具来实施解决问题的最优方案。1987年,作者首次将数据结构与算法有机结合,编著了《数据结构与算法》一书,由武汉大学出版社出版(以下简称《数》书)。之后一直致力于数据结构与算法的研究,得到了许多新的启示,尤其是体会到数据结构与算法无论是对计算机科学本身,还是对以计算机为基础的信息系统都起着十分重要的作用。于是站在信息系统这一高度,在已出版的《数》书的基础上,1996年重新编著了《数据结构与算法导论》,由电子工业出版社出版(以下简称《导》书)。今天,从启发人、教育人、培养人的角度,作者认为“数据结构与算法”这门课程不仅是使学生获得知识,更重要的是获得一种思维,包括逻辑思维、形象思维、结构化思维,这些思维将影响学生一生,于是再次修改《导》书,从培养人才的视角,编著了《数据结构与算法》。感谢高等教育出版社,使作者有机会再次将她奉献给读者。

作为一本新书,不仅是在计算机广泛应用于经济管理、系统工程、情报检索等领域的新形势下编著的,而且是在我国提出人才强国的新形势下编著的,是在作者作为一名博士生导师,培养了一批复合型人才之后的今天,是在作者作为武汉大学教学督导团团长研究了本科的创新教育的今天重新修改的。她具有如下特点:①在指导思想,站在启发人、教育人、培养人的高度;②在体系结构上,全书按数据的逻辑结构、数据的物理结构和数据的运算展开,赋予了数据结构与算法新的体系;③在学术研究上,建立了数据结构与算法相结合的研究模式;④在内容上,写进了作者的学术研究、项目研究的研究成果以及知识创造者的智力活动;⑤在编排上,遵循教学规律,融入教学经验,图文并茂,实例丰富,集逻辑思维、形象思维、结构化思维于其中,使之能启发思维、培养思维。

全书共分为16章,由5个部分组成。第一篇概论(第一章~第三章)。阐述了数据结构、算法的基本概念,提出了计算机科学是数据结构与算法的学问,建立了将数据结构与算法结合研究非数值问题的研究模式。第二篇几种逻辑结构(第四章~第九章)。介绍了几

种逻辑结构——线性表、栈和队、串、数组、树、图及计算机存储方式,运算和典型应用实例。第三篇 数据运算(第十章~第十二章)。讨论了排序、查找、集合操作等数据处理中的重要问题,给出了算法及其复杂性分析。第四篇 文件(第十三章~第十四章)。介绍了几种文件组织方式、文件排序和文件应用实例。第五篇 算法设计与分析(第十五章~第十六章)。提出了算法复杂性分析的基本方法。算法设计的6个基本策略及在经济管理中的应用实例。除第一篇外,其余各章均附有丰富的习题供读者选用。书中各算法的C语言源程序代码和书中内容的ppt课件将上载到高等教育出版社计算机教学资源网上,供读者学习参考。

本书在算法描述上选择了自然语言。这样做的目的是使得学生将学习的重点放在思维的训练上,教师将教学的重点放在思维的培养上,把算法的思路分析清楚。思路清楚了,就很容易用自然语言将其描述成算法,因为自然语言十分贴近人的自然思维。此外,用自然语言描述的算法,读者还可以很容易地根据需要将其翻译成高级语言程序上机运行。

程序设计是从事计算机科学及其应用领域工作的基本功。作者将数据结构与算法结合,以“方法——实施方法的数据结构——施加在数据结构上的算法——程序”的模式展开,用数据结构与算法结合的观念指导编程。本书提出的这一编程模式,不仅使程序结构性强,易理解,而且对读者编程能力的培养会起到激发引入的效果。有了深厚的编程功底,而后将大展宏图。

全书内容丰富,在叙述上深入浅出。对每个算法从问题入手,用图解法模拟计算机的存储及其操作,分析操作的规律性,写出通式,最后导出算法,使程序设计水到渠成,这样做对培养学生的逻辑思维、形象思维、结构化思维大有裨益。尤其结合经济管理的应用实例,不仅激发读者的学习兴趣,更加加深对算法的理解,且使之便于自学。

本书站在数据结构与算法的前沿,写进了最新研究成果,它们是作者多年来在排序、查找、最小生成树和工件排序等非数值问题方面所取得的一些可喜成果。如:和周建钦同志合作的公式分组排序、公式分组查找等算法,分别将当前世界上排序的计算复杂度由 $O(n\log n)$ 降到 $O(n)$,查找的计算复杂度由 $O(n\log n)$ 降到 $O(n)$ 。同时最小生成树算法、工序问题的动态规划算法等,均将当前同类问题的计算复杂度降低了一个数量级。读者不仅可以从中学习到科学的研究方法,还可将它应用于计算机科学本身及现代化的管理中。

凡是学过一门高级语言程序设计的人都能看懂此书,并可从中学到许许多多的编程技巧。本书有一部分内容具有一定深度(*号处的章节),仅供阅读、参考,而不作教学要求。另外,作者认为应十分注重算法分析,因为不仅要培养读者的编程能力,还要给读者树立时间和空间的观点,设计高效的优化算法,节省计算机的时间和空间资源。由于这是一门创造性的思维活动,有些算法具有相当的难度和深度,故这些部分也仅供研究人员参考,而不作教学要求。

书中所有算法及其应用实例都已编写成C语言程序并上机调试通过,书中的内容均制作成ppt课件。这些工作是由王晓杰、殷圣平、周铭、张鹏、姚艳、曾学工、周琳、柳建民、王丽、

孙靓、王祎、江颖、程瑾、张振国、何慧婷、孙辉伟、汪志波、夏岩、张巧龄等研究生做的。

《数据结构与算法》一书可以说经过了 20 余年的锤炼,是我个人最青睐的著作之一,她是我在顶层 12 m² 小房挥汗如雨孕育出来的处女作;是我今天拥有 300 万字论著的基础;记载了我在本科讲台上默默耕耘的 20 余个春秋;她反映了我乐于创新、勇于攀登科学高峰的精神。如今,我已是博士生导师,两个专业博士点学科(管理科学与工程、技术经济及管理)的学科带头人,但我仍然觉得本科教学是根本,在人才强国的今天,本科教学尤为重要,对学生,优秀人才出自于本科的学习;对教师,大师从本科教学中走出来。而《数据结构与算法》对培养人的逻辑思维、形象思维、结构化思维起着十分重要的作用,我有今天的成就,感谢她给予我的基本功,因此我要将她奉献给本科教学,这是我今生的追求——对社会、对人才培养做出贡献。

《数据结构与算法》这部著作的再次问世,感谢我所有教过的和我正在教的学生。敬请专家、读者批评、指正。

徐绪松

2004 年 5 月于武汉大学

目 录

第一篇 概论

第一章 数据结构	(5)
1.1 什么是数据结构	(5)
1.2 数据的逻辑结构	(6)
1.3 数据的物理结构	(8)
第二章 算法	(10)
2.1 什么是算法	(10)
2.2 算法的评价	(10)
2.3 算法的描述	(12)
第三章 数据结构与算法	(14)
3.1 数据结构与算法	(14)
3.2 算法 + 数据结构 = 程序	(15)

第二篇 几种逻辑结构

第四章 线性表	(19)
4.1 线性表	(19)
4.2 线性表的顺序存储结构及其运算	(20)
4.2.1 向量	(20)
4.2.2 对向量的插入和删除运算	(21)
4.2.3 线性表的应用实例 1—— 仓库管理系统(一)	(22)
4.3 线性表的链式存储结构及其运算	(25)
4.3.1 单链表	(25)
4.3.2 对单链表的插入和删除运算	(26)
4.3.3 线性表的应用实例 2—— 仓库管理系统(二)	(31)
4.3.4 循环链表	(34)
4.3.5 线性表的应用实例 3—— 多项式的算法运算	(35)
4.3.6 双向链表	(40)
4.3.7 对双向链表的插入和	

删除运算	(40)
4.3.8 线性表的应用实例 4—— 动态存储管理	(42)
4.3.9 线性表的应用实例 5—— 自动预订飞机票系统	(52)
习题	(57)
第五章 栈和队列	(60)
5.1 栈	(60)
5.2 栈的应用实例	(62)
5.2.1 学生业务档案系统	(62)
5.2.2 计算表达式	(63)
5.3 队	(67)
* 5.4 队的应用实例	(70)
5.4.1 排队模拟	(70)
5.4.2 售票处的服务系统	(76)
5.5 栈和队的应用实例—— 停车场管理	(82)
习题	(85)
第六章 串	(87)
6.1 串的基本概念	(87)
6.2 串的运算及其实现	(87)
6.3 串的存储结构	(90)
* 6.4 字符串的应用实例—— 模式匹配	(95)
习题	(101)
第七章 数组和广义表	(102)
7.1 数组及其存储结构	(102)
7.2 稀疏矩阵的存储压缩及其运算	(103)
7.2.1 求转置矩阵	(104)
* 7.2.2 求矩阵的乘积	(108)
7.3 稀疏矩阵的十字链表结构	(111)
7.4 广义表和多重链表	(114)
习题	(115)

第十二章 集合操作	(254)	14.2 磁盘排序	(311)
12.1 对集合的基本操作	(254)	14.2.1 k 路归并	(311)
12.2 顺序搜索与链表结构	(254)	14.2.2 并行操作的缓冲区处理	(314)
12.3 二元搜索与二元搜索树	(255)	14.2.3 初始归并段的产生	(321)
* 12.4 最佳二元搜索树	(256)	14.3 磁带排序	(327)
* 12.5 UNION - FIND 操作	(261)	14.3.1 平衡归并排序	(327)
12.6 字典和优先队	(266)	14.3.2 多步归并排序	(329)
12.7 Hash(杂凑)技术	(271)	习题	(330)
12.7.1 Hash 函数的构造方法	(273)		
12.7.2 冲突的处理	(275)		
12.8 集合运算的运用——最小生成树算法	(279)		
习题	(282)		
		* 第五篇 算法设计与分析	
		第十五章 算法分析技术	(335)
		15.1 算法分析简介	(335)
		15.2 循环程序的分析	(336)
		15.3 递归算法的分析	(337)
		15.3.1 递归方程	(337)
		15.3.2 递归算法的分析	(339)
		习题	(342)
		第十六章 算法设计的技术	(344)
		16.1 分割求解法	(344)
		求集合的最大、最小元	(344)
		16.2 动态规划	(347)
		16.2.1 单源路径问题	(347)
		16.2.2 项目群投资决策问题	(350)
		16.2.3 工件排序问题	(356)
		16.3 子目标法	(361)
		吉普车问题	(361)
		16.4 探索法	(363)
		旅游花费问题	(363)
		16.5 回溯法	(365)
		组合锁问题	(365)
		16.6 分枝与限界	(367)
		旅游路线问题	(367)
		习题	(379)
		参考文献	(380)
第十三章 文件组织	(285)		
13.1 基本知识	(285)		
13.2 顺序文件	(289)		
13.2.1 顺序文件的查找	(289)		
13.2.2 顺序文件的更新	(290)		
13.3 随机处理文件	(290)		
13.3.1 索引文件	(290)		
13.3.2 索引顺序文件	(291)		
13.4 直接存取文件	(294)		
13.4.1 按桶散列	(295)		
* 13.4.2 可扩充散列	(296)		
13.5 索引链接文件	(298)		
* 13.6 多关键字文件	(299)		
13.6.1 多重链表文件	(300)		
13.6.2 倒排文件	(301)		
* 13.7 文件的应用实例——旅馆管理系统	(302)		
习题	(309)		
* 第十四章 文件排序	(310)		
14.1 文件排序的主要过程	(310)		

第四篇 文件

第一篇

概 论

电子计算机已广泛地应用于系统工程、运筹学、经济管理、金融、商务、政务等各个领域，它们将数据处理向前推进了一步。看几个例子。

例 1 高校教师的教学管理

高校教务处使用计算机对全校教师的教学情况作统一管理。它要了解各门课程由哪些教师承担、每门课程的教学时数、使用教材及各位教师承担的课程、教师的基本情况、教学对象、教学效果等。将这个问题抽象成一张表，如表 1 所示。

表 1 高校教师教学情况表

姓名	性别	出生年月	职务	工作单位	专业	学历	教龄	所授课程	学时	教学对象	教材	效果
----	----	------	----	------	----	----	----	------	----	------	----	----

每个教师的教学情况按表存入计算机内，根据需要对这张表进行查找。求解这个问题的数学模型是信息表，运算是和信息表的查找操作。

例 2 工厂的组织管理

某工厂的组织机构如图 1 所示。

厂长要通过计算机了解各个部门工作情况及车间生产情况。将这个问题抽象成如图 2 所示的一棵树。

各部门及各车间情况按图 2 所示的树以一定的方式存入计算机内，对这棵树进行遍历便能了解厂内的整个情况。求解这个问题的数学模型是树，运算是树的遍历。

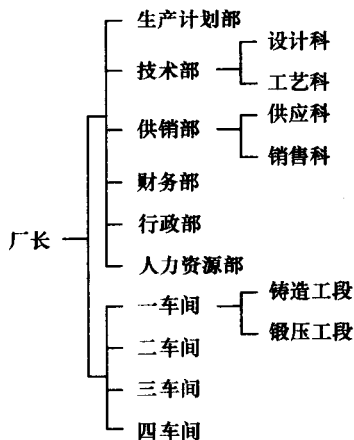


图 1 某工厂的组织机构图

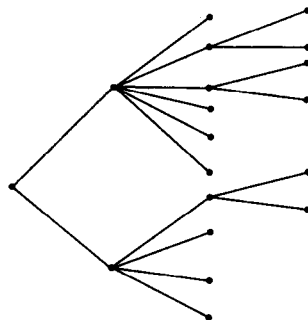


图 2 抽象后的某工厂组织机构图

例 3 最短路径问题

从油田铺设管道，把原油运到加工厂。求使管道总长最短的铺设方案。将这个问题抽象成如图 3 所示的有向网路。其中 v_1 为油田， v_9 为原油加工厂， $v_2 \sim v_8$ 是问题要求管道必须按给定的道路铺设所经过的地点。每条边旁的数字是这条道路的长度。用计算机求解最

短长度的铺设方案,首先要将图3的有向网络按一定方式存入计算机内,然后对这个有向网络设计一种算法(并非简单的数值计算),在各种铺设方案中选一种总长度最短的铺设方案。求解这个问题的数学模型是图,运算是图的一种较为复杂的非数值计算——图的算法。

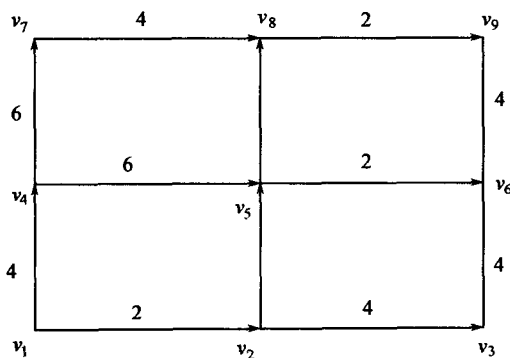


图3 从油田到原油加工厂的有向图

从以上3个例子可以看到,首先描述这类问题的数学模型不再只是数值方程,而是诸如表、树和图的非数值性的数据结构;其次,求解这类问题不再只是数值计算,而是要对一些信息表进行插入、删除、排序、查找;对树进行遍历;对图做较为复杂的非数值计算等。总的来说,当前计算机面临大量的非数值性程序设计问题。当计算机面临这些非数值问题时,操作的对象——数据,将具有一定的结构关系,有些甚至具有很复杂的结构关系。因此,对非数值性程序设计需解决如下问题:数据间的结构关系如何表示;数据在计算机内如何存储;处理这些数据(或叫数据运算)有哪些技巧。这就是本书要研究的问题。

这一部分将介绍有关数据结构、算法的基本概念,阐述数据结构与算法之间的本质联系,建立一个非数值问题程序设计的模式。本部分的内容将在以后各章节逐步深化、展开。

第一章 数据结构

1.1 什么是数据结构

什么是数据结构？

首先了解一下什么是数据。直观地说，数据是描述客观事物的数字、字母和符号，是计算机程序使用和加工的“原料”。数据的基本单位是数据元素，性质相同的数据元素的集合叫数据对象。数据对象中的元素彼此之间存在的相互关系叫做结构。

为了进一步理解什么是数据结构，先来看一个具体例子。图书馆是大家所熟悉的，那么图书馆的结构是什么呢？从物理上来看，图书馆主要由装书的书架及书籍组成，但另一方面，还存在一个图书馆的编目表，即图书馆藏书的索引。图书馆从两个方面管理图书：物理的藏书和逻辑的编目表。这就是图书馆的结构。和图书馆一样，计算机管理数据也有两个方面：即物理的存储和逻辑的关系。下面从这两个方面来回答什么是数据结构。

数据结构指的是数据之间的结构关系。具体来说，它包括数据的逻辑结构和数据的物理结构。

数据的逻辑结构——仅考虑数据元素之间的逻辑关系。它包括：**线性结构**，如线性表、栈、队；**非线性结构**，如树、二叉树和图。

数据的物理结构——指数据元素在计算机存储器中的表示，即存储结构，比如向量、链表。

一种逻辑结构通过映像便得到它相应的存储结构。同一种逻辑结构可以映像成不同的内部存储结构。反过来，数据的存储结构一定要反映数据之间的逻辑关系。

为了更具体一些，举一个例子。

有一叠扑克牌，希望在计算机中表示这一叠扑克牌的内容（也就是这一组信息）。

在计算机内一组信息是由一组节点组成的。这里用一个节点表示一张牌，为了说明这张扑克牌的内容，必须将它的花色（梅花、方块、红心、黑桃）、点数、正反面、名称表示出来，同时还要将这张牌的下一张牌表示出来，为此，用5个域组成一个节点，如图1.1所示。

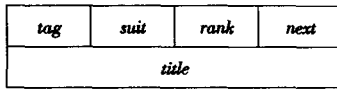


图 1.1 节点的格式

其中 tag 表示牌的正、反面(用 0、1 表示), suit 表示花色(用 1、2、3、4 分别表示梅花、方块、红心、黑桃), rank 表示点数, next 表示下一节点的地址(即组成该节点的那些存储单元的首地址), title 表示这张牌的名称,用 5 个字符表示。

如图 1.2 所示的一叠扑克牌,它的逻辑结构是线性表:(方块 2, 梅花 3, 黑桃 10(反)),在计算机中相应的存储结构用链表表示,如图 1.3 所示(也可以用另外的存储方式)。

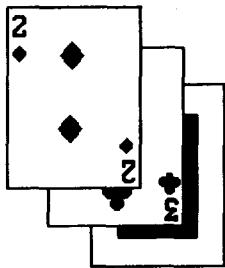


图 1.2 扑克牌的逻辑结构

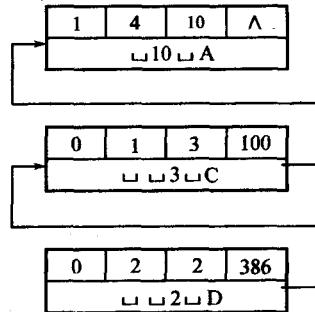


图 1.3 扑克牌的存储结构

1.2 数据的逻辑结构

数据的逻辑结构是独立于计算机的、对数据元素之间的逻辑关系的描述。从集合的观点,它可以形式地用一个二元组 $B = (D, R)$ 来表示,其中 D 是数据元素的集合, R 是 D 上关系的集合。

按集合的观点,数据的逻辑结构有两个要素:一是数据元素;二是关系。

数据元素是独立的信息。它可以是一个单独的符号,如英文字母表(A, B, ..., Z),数据元素是一个字母;也可以由若干个数据项组成,如一个企业单位的全体职工档案登记表(如表 1.1 所示),数据元素由姓名、职工号、性别、职务、工资 5 个数据项组成。

关系是指数据元素间的逻辑关系。可以是线性的——对数据元素而言,只有一个前驱和一个后继;也可以是非线性的——对数据元素而言,或者有一个前驱、多个后继,或者有多个前驱、多个后继。