



21世纪全国高职高专计算机教育“十一五”规划教材
丛书主编 全国高等学校计算机教育研究会课程与教材建设委员会主任 李大友

数据结构

主 编 王学军 方建文
副主编 孙 杰 胥道强
秦本涛 徐义峰



21世纪全国高职高专计算机教育“十一五”规划教材

数 据 结 构

本书编委会 编著

中 国 计 划 出 版 社

图书在版编目(CIP)数据

数据结构 / 《数据结构》编委会编著. —北京: 中国计划出版社, 2007. 8

21世纪全国高职高专计算机教育“十一五”规划教材
ISBN 978-7-80177-918-2

I. 数… II. 数… III. 数据结构—高等学校: 技术学校—教材 IV. TP311.12

中国版本图书馆CIP数据核字(2007)第070255号

内 容 简 介

本书针对高等职业技术教育的特点, 以培养技术应用人才为目标, 内容基本上涵盖了数据结构的主要内容。全书共8章, 循序渐进地引导读者学习了线性表、栈和队列、串、数组、广义表、树、图等不同的数据结构, 并系统地介绍了查找和排序的各种实现方法。对每一种数据结构, 除了详细阐述其基本概念和具体实现外, 还尽可能地对每种操作给出算法描述; 并且对查找和排序的各种算法进行了时间和空间性能的分析 and 比较。全书采用C语言作为数据结构和算法的描述语言。

本书既可作为高职高专院校计算机专业相关课程的教材, 也可作为计算机学习者的技术参考书。

21世纪全国高职高专计算机教育“十一五”规划教材

数 据 结 构

本书编委会 编著

☆

中国计划出版社出版

(地址: 北京市西城区木樨地北里甲11号国宏大厦C座4层)

(邮政编码: 100038 电话: 63906433 63906381)

新华书店北京发行所发行

北京市艺辉印刷有限公司印刷

787×1092毫米 1/16 19.75印张 480千字

2007年8月第一版 2007年8月第一次印刷

印数1—5000册

☆

ISBN 978-7-80177-918-2

定价: 28.00元

丛书编委会

主 任：李大友

副主任：王行言 郑 莉 傅连仲

委 员：（按音序排列）

蔡 莉	成安霞	东朝晖	范双南	韩小祥
黄国雄	黄志刚	将星军	李国安	李 红
李金祥	李亚平	李寅虎	李玉虹	刘 钢
刘灿勋	刘长生	刘立军	刘文涛	刘晓魁
刘占文	罗文华	孟繁增	商信华	邵 杰
舒大松	万雅静	王德奎	王宏基	文其知
吴 博	吴国经	吴 玉	武嘉平	谢书玉
阳若宁	杨邦荣	杨学全	袁学松	曾凡文
钟新文	周承华	周少华	朱元忠	朱志伯

本书编委会

主 编：王学军 方建文

副主编：孙 杰 胥道强 秦本涛 徐义峰

参 编：王俊平 王晓燕 曾丽娟 杨战旗

丛 书 序

编写背景和目的

高等职业教育是现代国民教育体系的重要组成部分，在实施科教兴国战略和人才强国战略中具有特殊的重要地位。现在，我国就业和经济发展正面临着两个大的变化，即：社会劳动力就业需要加强技能培训，产业结构优化升级需要培养更多的高级技术人才。温家宝总理在 2005 年 11 月 7 日的全国职业教育工作会议上指出，高等职业教育的发展仍然是薄弱环节，不适应经济社会发展的需要；大力发展高等职业教育，既是当务之急，又是长远大计。《国家教育事业发展规划“十一五”规划纲要》中提出，要以培养高素质劳动者和技能型人才为重点，提高学生创新精神和实践能力，大力发展职业教育；扩大高等职业教育招生规模，到 2010 年，使高等职业教育招生规模占高等教育招生规模的一半以上。在以上背景下，我国已进入了新一轮高等职业教育改革的高潮，目前高职院校的学校规模、专业设置、办学条件和招生数量，都超过了历史上任何一个时期。

随着信息社会的到来，灵活应用计算机知识、解决各自领域的实际问题成了当代人必须掌握的技能，为此，高职院校面向不同专业的学生开设了相关的计算机课程。然而，作为高职院校改革核心之一的教材建设大大滞后于高等职业教育发展和社会需求的步伐，尤其是多数计算机应用教材，或显得陈旧，或显得过于偏重理论而忽视应用。以致于一些通过 3 年学习的高职院校学生毕业后，所掌握的技能不能胜任用人单位的需求。

鉴于此，中国计划出版社与全国高等学校计算机教育研究会课程与教材建设委员会联合在全国 1105 所高职高专中做了广泛的市场调查，并成立了《21 世纪全国高职高专计算机教育“十一五”规划教材》编委会，由全国高等学校计算机教育研究会课程与教材建设委员会主任委员、北京工业大学李大友教授担任编委会主任。编委会进行了大量调查研究，通过借鉴国内外最新的、适用于高职高专教学的计算机技术经验成果，推出了切合当前高职教育改革需要、面向就业的系列职业技术型计算机教材。

系列教材

本计算机系列教材主要涵盖了当前较为热门的以下就业领域：

- 计算机基础及其应用
- 计算机网络技术
- 计算机图形图像处理和多媒体
- 计算机程序设计
- 计算机数据库

- 电子商务
- 计算机硬件技术
- 计算机辅助设计

教材特点

本套教材的目标是全面提高学生的计算机技术实践能力和职业技术素质，为此，中国计划出版社与全国高等学校计算机教育研究会课程与教材建设委员会合作，邀请了来自全国各类高等职业学校的骨干教师（其中很多为主管教学的院长或系主任）作为编委会成员外，还特聘了多位具有丰富实践经验的一线计算机各应用领域工程师参加教材的技术指导和编审工作，以期达到教学理论和实际应用紧密结合的效果。

同时，为配合各学校的精品课程建设工程，本套教材以国家级精品课程指标为指引方向，借鉴其他兄弟出版社的先进经验和成功案例，提出了建设“立体化教学资源平台”的概念，其内容包括教材、教学辅导资料、教学资源包、网络平台等内容，并将在后续培训、论文发表等多方面满足教师与精品课程建设的需求。

本系列教材的特点如下：

（1）面向就业。本系列教材的编写完全从满足社会对技术人才需求和适应高等职业教育改革的角度出发，教材所涉及的内容是目前高职院校学生最迫切需要掌握的基本就业技能。

（2）强调实践。高职高专自身教育的特点是强调实践能力，计算机技术本身也是实践性很强的学科，本系列教材紧扣提高学生实践能力这一目标，在讲解基本知识的同时配套了大量相关的上机指导、实训案例和习题。

（3）资源丰富。本系列教材注重教材的拓展配套，辅助教学资源丰富。除了由本书作为主干教材外，还配有电子课件、实训光盘、习题集和资源网站等辅助教学资源。

读者定位

本计算机应用系列教材完全针对职业教育，主要面向全国的高职高专院校。本系列教材还可作为同等学历的职业教育和继续教育的教学用书或自学参考书。

本系列教材的出版是高职教育在新形势下发展的产物。我们相信，通过精心的组织和编写，这套教材将不仅能得到广大高职院校师生的认可，还会成为一套具有时代鲜明特色、易教易学的高质量计算机系列教材。我们与时俱进，紧密配合高职院校的办学机制和运行体制改革，在后期的组织推广及未来的修订出版中不断汲取最新的教学改革经验和教师学生及用人单位的反馈意见，为国家高等职业教育奉献我们的力量。

丛书编委会

前 言

数据结构是计算机应用及相关专业一门重要的专业基础课,也是一门必修的核心课程。数据结构是计算机程序设计的重要理论技术基础,在计算机科学的各个领域都要用到各种不同的数据结构。例如,编译系统中要使用栈、散列表和语法树,操作系统中要使用队列,数据库系统中要使用线性表、链表,人工智能中要使用有向图。此外,在面向对象的程序设计、计算机图形学、软件工程、多媒体技术和计算机辅助设计等领域,都会用到许多数据结构的知识。

数据结构课程的教学要求是:使学生学会分析研究计算机加工的数据对象的特性,以便在实际应用中选择适当的数据结构和相应的算法,初步掌握算法的时间与空间性能的分析技巧,掌握分析复杂程序设计的能力。通过对本书的学习,读者能够以典型的数据结构和算法为研究对象,学会分析数据对象的特征,掌握数据组织的方法和在计算机中的表示方法,为数据选择适当的逻辑结构、存储结构以及相应的处理算法,初步掌握算法的时间、空间复杂度的分析技巧,培养良好的程序设计风格以及进行复杂程序设计的能力。

全书采用C语言作为数据结构和算法的描述语言,对每一种数据结构,都分别给出了相应的抽象数据类型规范说明和实现方法,并增添了C++语言的引用、调用参数传递方式等,使本书对各种抽象数据类型的定义和实现简明清晰,既不拘泥于C语言的细节,又容易转换成能上机执行的C或C++程序。本书基本上覆盖了数据结构的主要内容,注重培养读者的实际操作能力,力求内容与应用实例相结合。在介绍基本理论时,尽量做到浅显易懂,同时以实例来讲解多种数据结构,叙述上通俗易懂,可加深读者对基本概念的理解,且有利于提高读者分析问题和解决问题的能力。

本书在编写中重视基础,循序渐进,内容精炼,重点突出。本书融入了学科内容和科学理念,反映了计算机技术的发展要求,倡导了理论联系实际和科学的思想方法,并以计算机学科的科学体系为依托,分类组织实施,兼容互补;注重理论与实践并重,强调理论与实践相结合,突出了学科发展特点,体现了学科发展的内在规律;在内容上循序渐进,保证学术深度,减少知识重复,前后相互呼应,内容编排合理,整体结构完整;采取自顶向下的设计方法,突出学科方法论,强调知识体系可扩展的原则。

全书共8章。

第1章引入了数据结构与算法的一些基本概念与术语,并对算法描述及算法分析做了简要说明。

第2章~第4章由浅入深地介绍了线性结构中的线性表、栈、队列、串、数组及广义表的基本定义、基本算法和基本应用。

第5章和第6章介绍了非线性结构的树、二叉树和图。

第7章和第8章介绍了在实际应用中使用非常广泛的查找和排序的基本算法，并进行了简单的时间、空间性能分析。

每章都配有上机实验、本章小结及练习题。

本书由王学军、方建文主编，孙杰、胥道强、秦本涛、徐义峰担任副主编，王俊平、王晓燕、曾丽娟、杨战旗参与编写。

由于时间仓促与编者水平有限，不足与欠妥之处在所难免，恳请广大读者不吝指正。

编者

2007年6月

目 录

第1章 绪论	1
1.1 什么是数据结构	1
1.1.1 数据的逻辑结构	2
1.1.2 数据的存储结构	3
1.1.3 数据的运算	6
1.2 基本概念	6
1.3 抽象数据类型的表示与实现	7
1.3.1 C/C++语言的基本数据类型	7
1.3.2 C/C++语言的指针类型	7
1.3.3 C/C++语言的数组类型	8
1.3.4 C/C++语言中的结构体类型	8
1.3.5 C/C++语言中的共用体类型	8
1.3.6 C/C++语言中的自定义类型	9
1.4 算法和算法分析	9
1.4.1 算法	10
1.4.2 算法描述	10
1.4.3 算法设计的要求	12
1.4.4 算法效率的度量	13
1.4.5 算法的存储空间需求	14
1.5 上机实验	14
1.5.1 实验内容	14
1.5.2 实验指导	15
1.6 本章小结	16
1.7 练习题	17
第2章 线性表	21
2.1 线性表的类型定义	21
2.2 线性表的顺序表示和实现	22
2.2.1 顺序表的定义	22
2.2.2 顺序表的基本运算	23
2.2.3 顺序表实现算法的分析	27
2.3 线性表的链式表示和实现	28
2.3.1 线性链表	28

2.3.2	循环链表	36
2.3.3	双向链表	40
2.4	一元多项式的表示及相加	46
2.5	上机实验	51
2.5.1	实验内容	51
2.5.2	实验指导	51
2.6	本章小结	53
2.7	练习题	54
第3章	栈和队列	58
3.1	栈	58
3.1.1	抽象数据类型栈的定义	58
3.1.2	栈的表示和实现	60
3.2	栈的应用举例	66
3.2.1	数制转换	67
3.2.2	括号匹配的检验	67
3.2.3	行编辑程序	68
3.2.4	迷宫求解	69
3.2.5	表达式求值	71
3.3	栈与递归的实现	72
3.4	队列	76
3.4.1	抽象数据类型队列的定义	76
3.4.2	链队列——队列的链式表示和实现	78
3.4.3	循环队列——队列的顺序表示和实现	81
3.5	离散事件模拟	87
3.6	队列的应用	90
3.7	上机实验	92
3.7.1	实验内容	92
3.7.2	实验指导	92
3.8	本章小结	96
3.9	练习题	96
第4章	串、数组和广义表	100
4.1	串类型的定义	100
4.2	串的表示和实现	102
4.2.1	定长顺序存储表示	102
4.2.2	堆分配存储表示	107
4.2.3	串的链存储表示	110
4.3	串的模式匹配算法	116
4.3.1	求子串位置的定位函数Index (S, T, pos)	116

4.3.2 模式匹配的一种改进算法	118
4.4 串操作应用举例	119
4.4.1 文本编辑	119
4.4.2 建立词索引表	121
4.5 数组的定义	121
4.6 数组的顺序表示和实现	122
4.7 矩阵的压缩存储	123
4.7.1 特殊矩阵	124
4.7.2 稀疏矩阵	125
4.8 广义表的定义	132
4.9 广义表的存储表示	134
4.10 广义表的基本运算算法	135
4.11 广义表的递归算法	141
4.12 上机实验	144
4.12.1 实验内容	144
4.12.2 实验指导	144
4.13 本章小结	147
4.14 练习题	147
第5章 树与二叉树	150
5.1 树	150
5.1.1 树的定义	151
5.1.2 树的表示	151
5.1.3 树的基本术语	152
5.1.4 树的存储结构	153
5.2 二叉树及其基本性质	155
5.2.1 什么是二叉树	155
5.2.2 二叉树的基本性质	156
5.2.3 满二叉树与完全二叉树	157
5.3 二叉树基本运算算法	158
5.3.1 二叉树的基本运算	158
5.3.2 二叉树基本运算实现算法	158
5.4 二叉树的遍历	162
5.5 确定唯一的二叉树	166
5.6 二叉树的存储结构	167
5.6.1 顺序存储结构	167
5.6.2 链式存储结构	168
5.7 线索二叉树	169
5.8 二叉树、树以及森林之间的转换	173

5.8.1	树和森林转换为二叉树	173
5.8.2	二叉树还原为树或森林	175
5.9	哈夫曼树及其应用	175
5.9.1	什么是哈夫曼树	175
5.9.2	哈夫曼树的构造	176
5.9.3	哈夫曼编码	178
5.10	上机实验	180
5.10.1	实验内容	180
5.10.2	实验指导	180
5.11	本章小结	182
5.12	练习题	182
第6章	图	186
6.1	图的基本概念	186
6.1.1	图的定义	186
6.1.2	图的基本术语	187
6.2	图的存储结构	189
6.2.1	邻接矩阵	190
6.2.2	邻接表	192
6.3	图的遍历	195
6.3.1	广度优先搜索法	196
6.3.2	深度优先搜索法	198
6.4	最小生成树	201
6.4.1	普里姆算法	201
6.4.2	克鲁斯卡尔算法	203
6.5	最短距离问题	207
6.5.1	单源最短距离	208
6.5.2	每对顶点之间的最短距离	210
6.6	拓扑排序	214
6.7	关键路径	217
6.8	上机实验	220
6.8.1	实验内容	220
6.8.2	实验指导	220
6.9	本章小结	222
6.10	练习题	222
第7章	查找	225
7.1	顺序查找	225
7.2	有序表的对分查找	227

7.3 分块查找	229
7.4 二叉排序树查找	231
7.4.1 二叉排序树的基本概念	231
7.4.2 二叉排序树的插入	232
7.4.3 二叉排序树的删除	233
7.4.4 二叉排序树的查找	235
7.5 二叉平衡树	235
7.6 哈希表查找	240
7.6.1 哈希表查找的基本概念	240
7.6.2 哈希函数的构造方法	241
7.6.3 哈希冲突的解决方法	243
7.7 多层索引树查找	258
7.7.1 B-树	258
7.7.2 B+树	262
7.8 上机实验	265
7.8.1 实验内容	265
7.8.2 实验指导	265
7.9 本章小结	267
7.10 练习题	267
第8章 排序	270
8.1 互换类排序	270
8.1.1 冒泡排序	271
8.1.2 快速排序	273
8.2 插入类排序	274
8.2.1 简单插入排序	274
8.2.2 希尔排序	276
8.3 选择类排序	278
8.3.1 简单选择排序	278
8.3.2 堆排序	279
8.3.3 树形选择排序	284
8.4 其他排序方法简介	285
8.4.1 归并排序	285
8.4.2 基数排序	288
8.4.3 二叉树排序	291
8.5 各种排序方法的比较和选择	294
8.5.1 各种排序方法的比较	294
8.5.2 各种内排序方法的选择	294
8.6 上机实验	295

8.6.1 实验内容.....	295
8.6.2 实验指导.....	295
8.7 本章小结.....	298
8.8 练习题.....	299
主要参考文献.....	302

第 1 章

绪论

数据的组织形式和表示方式直接关系到计算机对数据的处理效率, 因此, 为了更好地进行程序设计, 有效地利用计算机, 就需要对计算机程序加工处理的对象进行系统和深入地研究。研究各种数据的特性以及数据之间存在的关系, 进而根据实际应用的要求, 合理地组织和存储数据, 设计出相应的算法, 这就是“数据结构”要讨论的问题。



本章主要内容

- 数据结构的逻辑结构、存储结构和数据运算三方面的相互关系
- 抽象数据类型的表示与实现, 总结了C/C++语言中常用的数据类型
- 算法定义及特性
- 使用C语言进行算法描述
- 算法的时间复杂度分析

1.1 什么是数据结构

自 1946 年第一台计算机问世以来, 计算机产业的飞速发展已远远超出了人们的预料, 在某些生产线上, 几秒钟就能生产出一块微型计算机芯片, 产量猛增, 价格低廉, 这就使得计算机的应用范围迅速扩展。如今, 计算机已深入到人类社会的各个领域。计算机的应用已不再局限于科学计算, 而更多地用于控制、管理及数据处理等非数值计算的处理工作。与此相应, 计算机加工处理的对象由纯粹的数值发展到字符、表格和图像等多种具有一定结构的数据, 这就给程序设计带来了一些新的问题。为了编写出一个“好”的程序, 必须分析要处理对象的特性以及各处理对象之间的关系。

数据是对客观事物的符号表示, 在计算机科学中是指所有能输入到计算机中并被计算机程序处理的符号的总称, 它是计算机程序加工的“原料”。例如, 一个班的全部学生记录、 $a \sim z$ 的字母集合、 $1 \sim 1\,000$ 的所有素数等都称为数据。

数据元素 (也称为节点) 是数据的基本单位, 在程序中通常作为一个整体进行考虑和处理。有时一个数据元素可以由若干个数据项组成。数据项是具有独立含义的最小标识单位。例如, 在整数集合中, “10” 就可称为一个数据元素。又如, 在一个数据库 (关系数据库) 中, 一个记录可称为一个数据元素, 而这个元素中的某个字段就是一个数据项。

数据结构是相互之间存在一种或多种特定关系的数据元素的集合。这些数据元素不是孤立存在的，而是有着某种关系，这种关系称为结构。

1.1.1 数据的逻辑结构

数据元素之间的逻辑关系称为数据的逻辑结构。根据数据元素之间逻辑关系的不同特性，分为下列 4 类基本结构：

- 集合。结构中的数据元素同属于一个集合（集合类型由于元素之间的关系过于松散，数据结构课程中较少讨论）。
- 线性结构。线性结构中的数据元素存在一对一的关系。
- 树形结构。树形结构中的数据元素存在一对多的关系。
- 图形结构。图形结构中的数据元素存在多对多的关系。

在大多数情况下，数据的逻辑结构可以用二元组来表示：

$$S = (D, R)$$

其中， D 是数据节点的有限集合； R 是 D 上的关系的有限集合。其中每个关系都是从 D 到 D 的关系。在表示每个关系时，用尖括号表示有向关系，如 $\langle a, b \rangle$ 表示存在节点 a 到节点 b 之间的关系；用圆括号表示无向关系，如 (a, b) 表示既存在节点 a 到节点 b 之间的关系，又存在节点 b 到节点 a 之间的关系。设 r 是一个 D 到 D 的关系， $r \in R$ ，若 $d, d' \in K$ ，且 $\langle d, d' \rangle \in r$ ，则称 d' 是 d 的后继节点， d 是 d' 的前驱节点，这时 d 和 d' 是相邻的节点（都是相对 r 而言的）；如果不存在一个 d' ，使 $\langle d, d' \rangle \in r$ ，则称 d 为 r 的终端节点；如果不存在一个 d' ，使 $\langle d', d \rangle \in r$ ，则称 d 为 r 的开始节点；如果 d 既不是终端节点也不是开始节点，则称 d 是内部节点。

例如，一个城市表（见表 1-1）就可称之为一个数据结构，它由很多记录（这里的数据元素就是记录）组成，每个元素又包括多个字段（数据项）。那么这个表的逻辑结构是怎么样的呢？分析数据结构都是从数据元素之间的关系开始分析，对于这个表中任意一个记录，它只有一个前驱节点和一个后继节点，整个表只有一个开始节点和一个终端节点。当知道了这些关系之后，就能明白这个表的逻辑结构即为线性结构。其逻辑结构表示如下：

City = (D, R)
 $D = \{ \text{北京, 上海, 武汉, 西安, 南京} \}$
 $R = \{ r \}$
 $r = \{ \langle \text{北京, 上海} \rangle, \langle \text{上海, 武汉} \rangle, \langle \text{武汉, 西安} \rangle, \langle \text{西安, 南京} \rangle \}$

表 1-1 城市表

城 市	区 号	说 明	城 市	区 号	说 明
北京	010	首都	西安	029	陕西省省会
上海	021	直辖市	南京	025	江苏省省会
武汉	027	湖北省省会			

数据的逻辑结构可以用相应的关系图来表示，称之为逻辑结构图。

【例 1-1】 设数据逻辑结构如下：

$B1 = (D, R)$

$D = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$R = \{r\}$

$r = \{<1, 2>, <1, 3>, <3, 4>, <3, 5>, <4, 6>, <4, 7>, <5, 8>, <7, 9>\}$ /* 尖括号表示有向关系 */

试画出对应的逻辑结构图，并给出哪些是开始节点，哪些是终端节点，说明是何种数据结构。

解：B1 对应的逻辑结构如图 1-1 所示。其中，1 是开始节点；2, 6, 8, 9 是终端节点。它是一种树形结构。

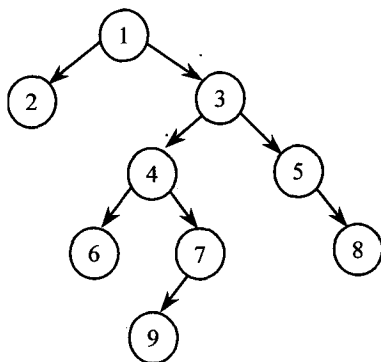


图 1-1 逻辑结构图

1.1.2 数据的存储结构

数据在计算机中的存储表示称为数据的存储结构，也称为物理结构。

把数据存储到计算机中时，通常要求既存储各节点的数值，又存储节点之间的逻辑关系。在实际应用中，数据的存储方法是灵活多样的，可根据问题规模（通常是指节点数目的多少）和运算种类等因素适当选择。本书将介绍 4 种基本的存储结构，即顺序存储结构、链式存储结构、索引存储结构和散列存储结构。下面简要介绍这些存储结构的特点。

1. 顺序存储结构

顺序存储结构是把逻辑上相邻的元素存储在一组连续的存储单元中，其元素之间的逻辑关系由存储单元地址间的关系隐含表示。

对于前面的逻辑结构 City，假定每个元素占用 30 个存储单元，数据从第 100 个存储单元开始由低地址向高地址方向存放，对应的顺序存储结构如表 1-2 所示。

表 1-2 顺序存储结构

地 址	城 市 名	区 号	说 明
100	北京	010	首都
130	上海	021	直辖市
160	武汉	027	湖北省省会
190	西安	029	陕西省省会
210	南京	025	江苏省省会

顺序存储结构的主要优点是节省存储空间。因为分配给数据的存储单元全部用于存放节点的数据值，节点之间的逻辑关系没有占用额外的存储空间。用这种方法来存储线性结构的数据元素时，可实现对各数据元素的随机访问。这是因为线性结构中每个数据元素都对应一个序号（开始节点的序号为 1，其后继节点的序号为 2，……），可以根据节点的序