



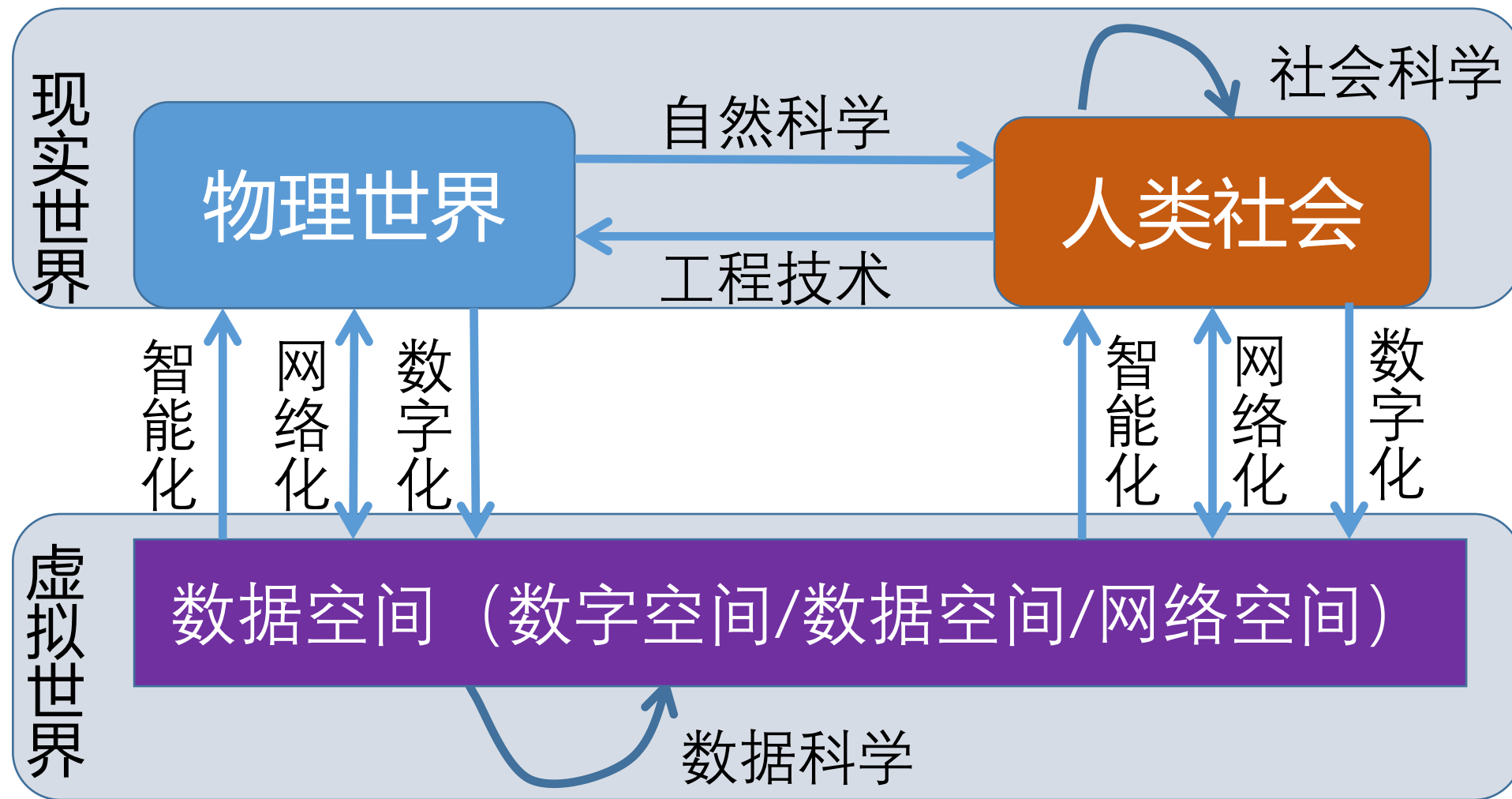
Data Structures and Algorithms

数据结构与算法

张 正

zhengzhang@hit.edu.cn

哈尔滨工业大学 (深圳)



数据：是用以描述客观事物的数值、字符，以及一切可以输入到计算机中并由计算机程序加以处理的符号的集合；



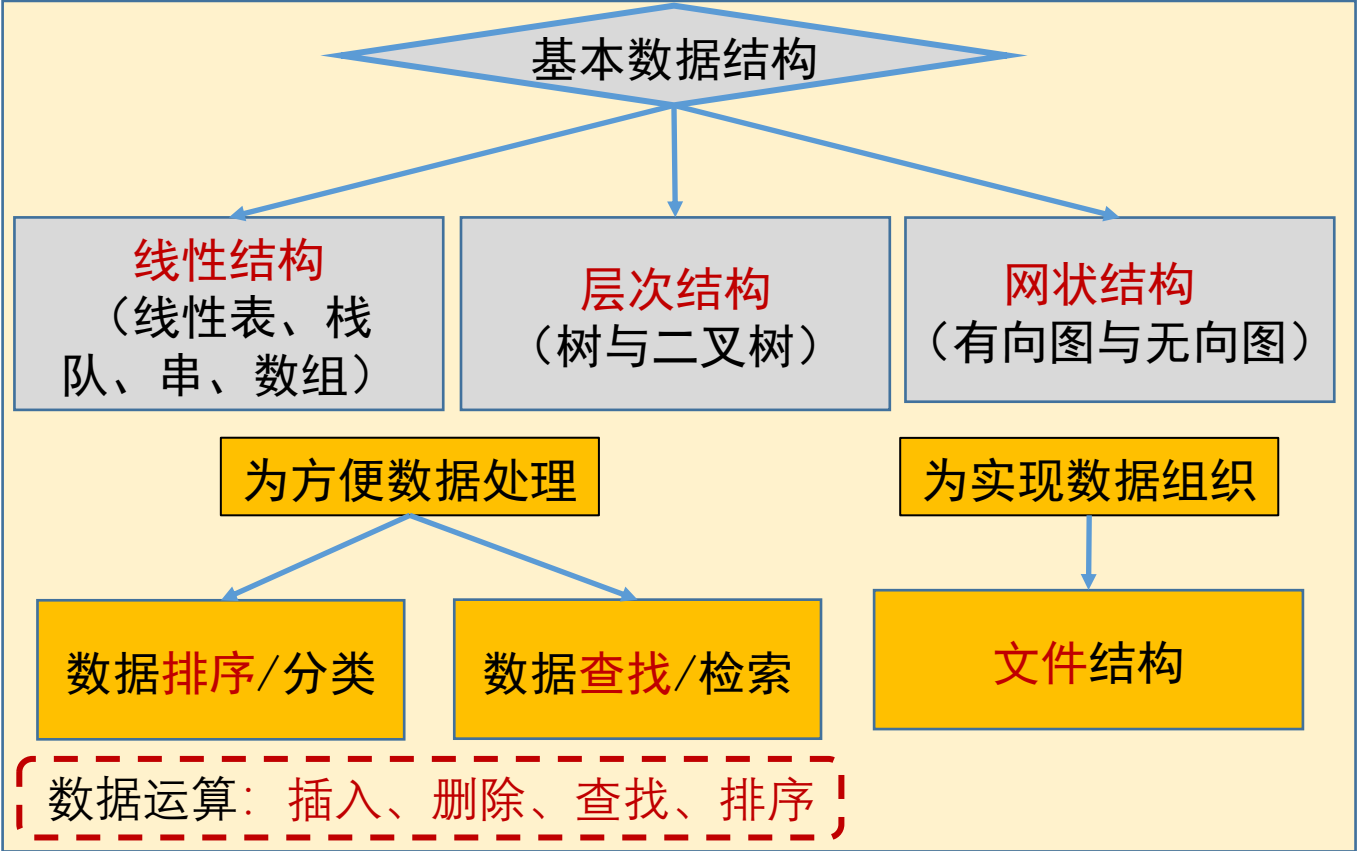
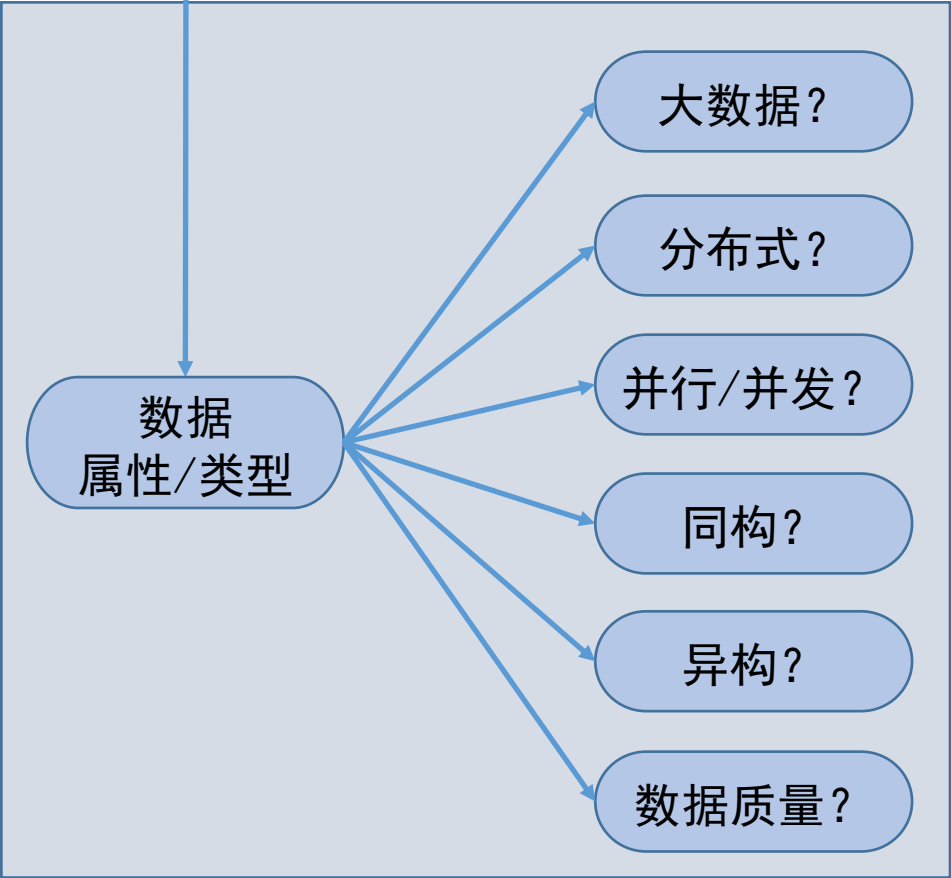
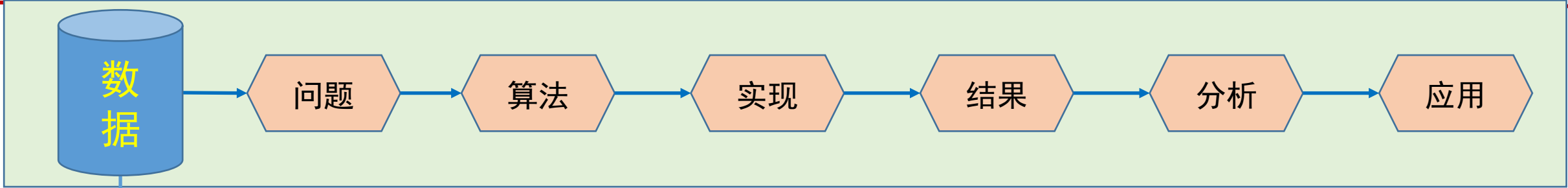
实现数据存储

外部存储

确定数据结构

设计实现算法

内部存储+处理



说明：建立在相同类型数据元素集合的基础上

数据结构与算法

数据结构：数据对象中数据元素的彼此“关系”，主要是指“位序”

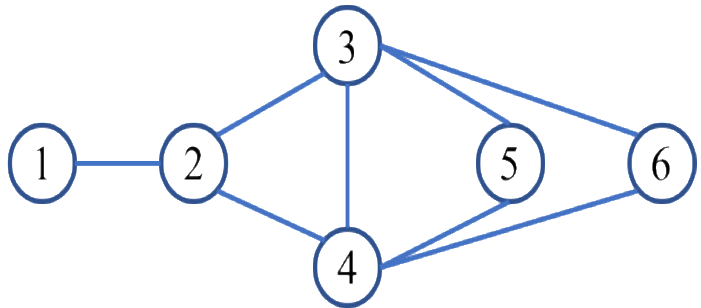
三大结构：

1、线性表：线性结构

2、树及二叉树：层次结构

3、图：网状结构

} 非线性结构



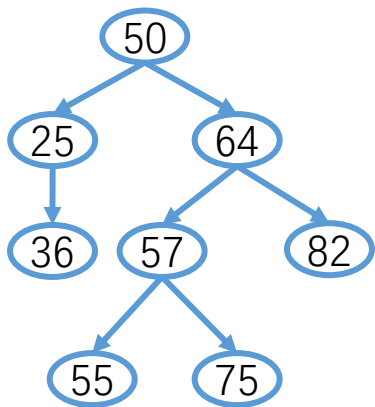
线性表：

除第一个数据元素没有前驱、最后一个数据元素没有后继元素外，其他任何一个数据元素只有唯一的前驱元素和唯一的后继元素。

$(a_1, a_2, a_3, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_n)$

树及二叉树：

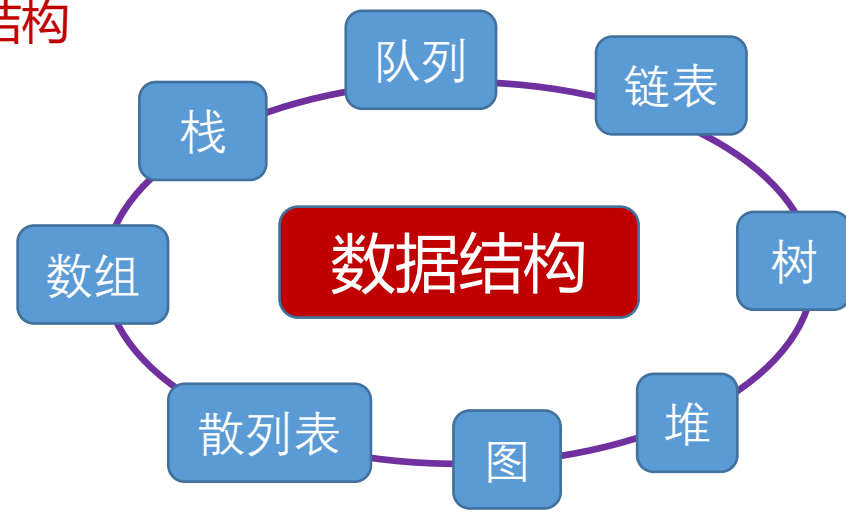
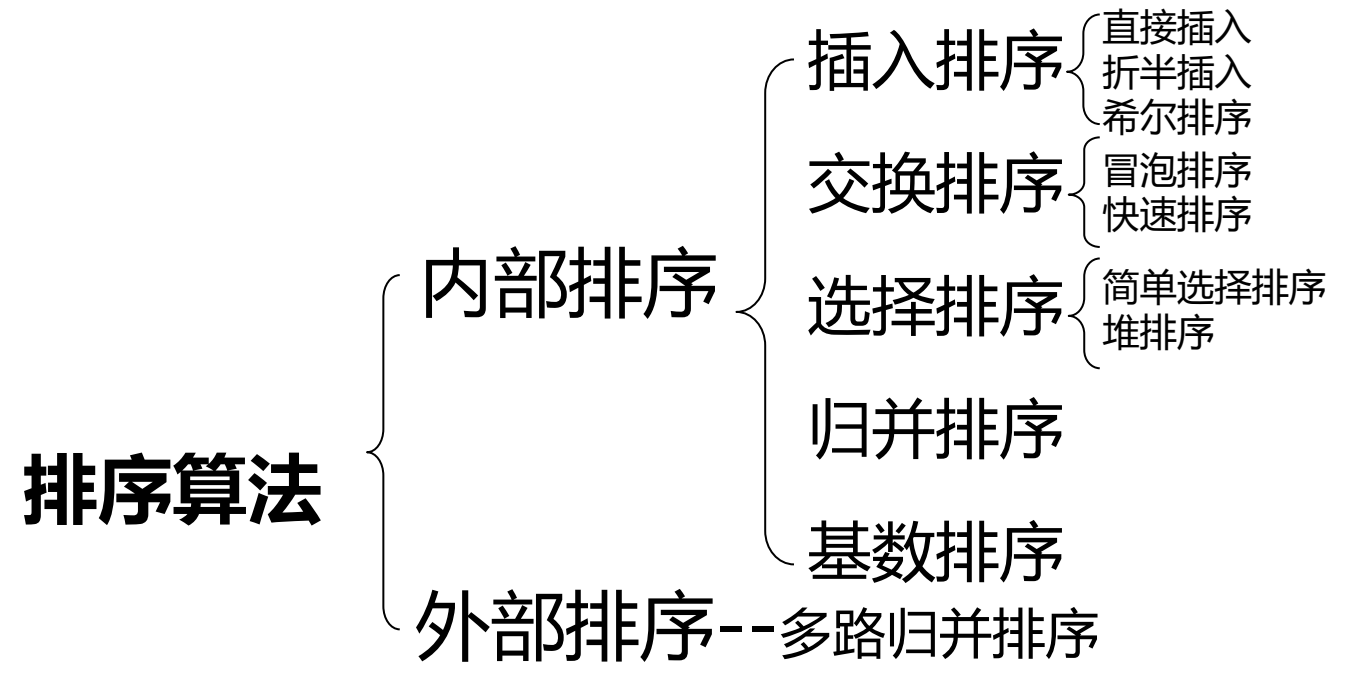
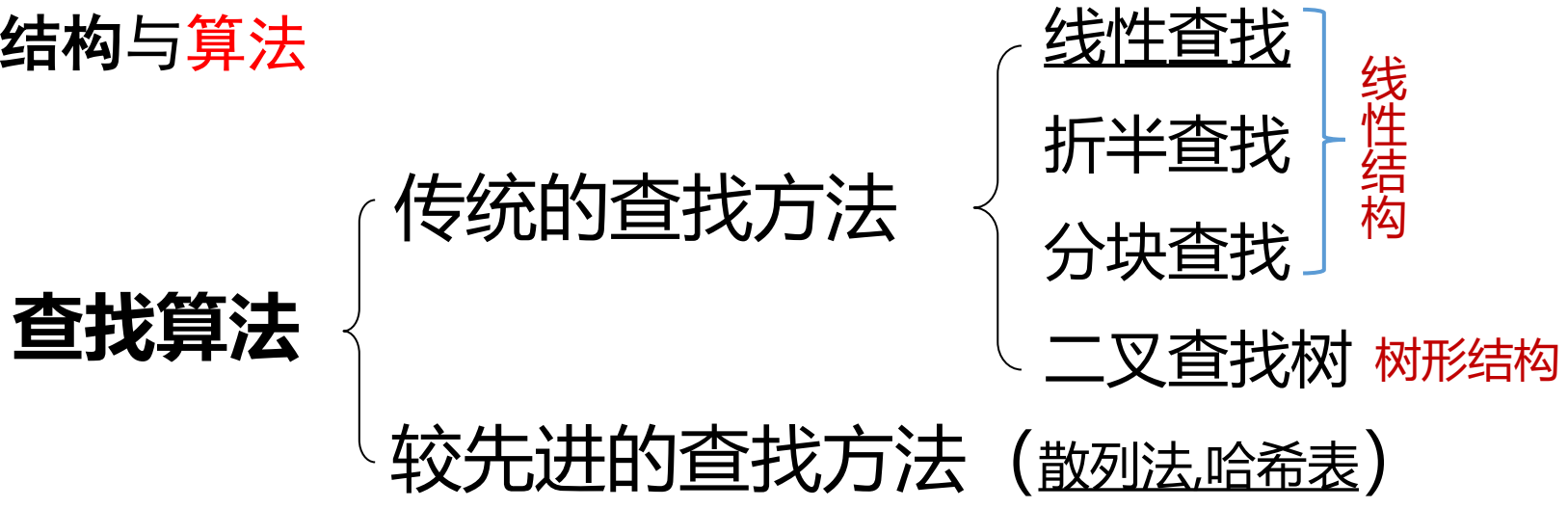
除第一个数据元素没有前驱之外（特殊），其他任何一个数据元素只有唯一的前驱元素，但可以有0个或多个后继元素。



图：

任何一个数据元素可以有0个或多个前驱元素，也可以有0个或多个后继元素。

数据结构与算法





唐纳德·克努特 (Donald Ervin Knuth)

- 1938年12月7日出生于美国威斯康星州密尔沃基市;
- 29岁提出“算法”(Algorithm)和“数据结构”(Data Structure);
- 1973年首创双向链表;
- 著名算法Knuth-Bendix和Knuth-Morris-Pratt的发明者;
- 排版系统TEX和METAFONT发明人。

1956年: 进入俄亥俄州克利夫兰的凯斯理工学院(现并入凯斯西储大学)物理专业;

1958年: 改学数学, 并从此与计算机结缘;

1960年: Knuth 被破例同时授予学士和硕士学位, 进入加州理工学院数学系;

1963年: 取得博士学位, 1966年副教授;

1968年: 斯坦福大学计算机科学系教授。开始撰写著名的《计算机程序设计艺术》系列图书;

1971年: 美国计算机协会格蕾丝·赫柏奖;

1973年: 当选为美国科学艺术学院院长;

1974年: 图灵奖历史上最年轻的获奖者;

1975年: 当选为美国国家科学院院士。

《计算机程序设计艺术》

1968年第一卷《基本算法》出版;

1969年第二卷《半数字化算法》出版;

1973年第三卷《排序与搜索》出版;

目前第四卷《组合算法》已出版。



《计算机程序设计艺术》一书与牛顿的《自然哲学的数学原理》等书一起, 被评为“世界历史上最伟大的十种科学著作”之一。

看了这部书后, 再谈起编程序都会变得谦虚谨慎。

比尔·盖茨曾说: “如果你能读懂整套书的话, 请给我发一份你的简历”。

数据结构与算法课程大纲

一、课程目标

《数据结构》是计算机大类的专业基础课，既是对前序课程的深入和扩展，也是为将来更加深入地学习其他专业课程打下基础。通过本课程的学习，学生将掌握数据结构和算法基础知识，能够提高程序设计的质量；根据问题的性质，能够选择合理的数据结构，完成算法的设计，并对时间复杂性进行必要的控制，培养解决实际问题的能力。

课程具体目标：

课程目标1：能够根据算法的实际操作需求，选择、设计合理的数据结构，并能实现相关算法；

课程目标2：能够对算法进行时间（空间）复杂性的分析；

课程目标3：能够通过具体案例，学习并掌握基本数据结构（线性、树、图）的设计与使用；

课程目标4：通过实验验收与实验报告的撰写，培养学生的书面表达能力和口头表达能力。

考研大纲

2021年全国硕士研究生入学联考（408）专业课考试大纲

考试科目名称：计算机学科专业基础

考试科目代码：[408]

综合试卷满分150分；考试时间180分钟。包括数据结构45分，计算机组成原理45分，操作系统35分，计算机网络25分。

试卷题型结构：单项选择题，80分(40小题，每小题2分)；综合题，70分。

重点考查考生掌握相关基础知识、基本理论和分析问题解决问题的能力。

统考（408）数据机构考查内容

考查目标：

- 1、掌握数据结构的基本概念、基本原理和基本方法。
- 2、掌握数据的逻辑结构、存储结构及基本操作的实现，能够对算法进行基本的时间复杂度与空间复杂度的分析。
- 3、能够运用数据结构的基本原理和方法进行问题的分析与求解，具备采用C或C++语言设计与实现算法的能力。

统考（408）**数据结构**考查内容

一、线性表

- （一）线性表的基本概念
- （二）线性表的实现
- （三）线性表的应用

二、栈、队列和数组

- （一）栈和队列的基本概念
- （二）栈和队列的顺序存储结构
- （三）栈和队列的链式存储结构
- （四）多维数组的存储
- （五）特殊矩阵的压缩存储
- （六）栈、队列和数组的应用

三、树与二叉树

- （一）树的基本概念
- （二）二叉树
- （三）树、森林
- （四）树与二叉树的应用

四、图

- （一）图的基本概念
- （二）图的存储及基本操作
- （三）图的遍历
- （四）图的基本应用

数据结构

五、查找

- （一）查找的基本概念
- （二）顺序查找法
- （三）分块查找法
- （四）折半查找法
- （五）B树及其基本操作，B+树的基本概念
- （六）散列（Hash）表
- （七）字符串模式匹配
- （八）查找算法的分析及应用

六、排序

- （一）排序的基本概念
- （二）插入排序
- （三）起泡排序
- （四）简单选择排序
- （五）希尔排序
- （六）快速排序
- （七）堆排序
- （八）二路归并排序
- （九）基数排序
- （十）外部排序

算法

HIT考研大纲

2021 年硕士研究生入学考试大纲

考试科目名称：计算机基础

考试科目代码：[854]

本考试科目考试时间 180 分钟，满分 150 分。包括计算机系统、计算机网络、数据结构与算法共三部分，每部分各 70、40、40 分。

HIT考研大纲

考试要求

1. 要求考生全面系统地掌握数据结构与算法的基本概念、数据的逻辑结构和存储结构及操作算法，并能灵活运用；能够利用数据结构和算法的基本知识，为应用问题设计有效的数据结构和算法；能够分析算法的复杂性。

2. 要求能够用 **C/C++/Java** 等程序设计语言描述数据结构和算法。

注：考试内容范围主要以参考书目 1 为标准，带*号部分不在考试范围之内。

参考书目

1. 廖明宏，郭福顺，张岩，李秀坤，数据结构与算法（第 4 版），高等教育出版社，2007. 11
2. 严蔚敏，吴伟民，数据结构（C 语言版），清华大学出版社，2002. 09

HIT考研大纲 - 考试内容

1) 数据结构与算法的概念

- a: 数据结构与算法及其相关的基本概念
- b: 算法及其复杂性分析

2) 线性表

- a: 线性结构及其操作算法
- b: 线性表的应用及算法

3) 树与二叉树

- a: 二叉树的定义、性质、表示、遍历算法
- b: 树的表示、操作算法
- c: 森林与二叉树关系
- d: 树与二叉树的应用及算法

4) 图及其相关算法

- a: 图的相关概念
- b: 图的存储结构与搜索算法
- c: 图的应用及算法

5) 查找与排序

- a: 查找与排序的相关概念
- b: 典型算法的描述及复杂性分析
- c: 查找与排序算法的应用

题型结构

- a: 选择题 (0—20 分)
- b: 简答题 (0—20 分)
- c: 算法设计题 (0—20 分)

注：题型分数在以上范围内浮动，总分为 40 分

课程整体安排

1. 课堂授课:

- ◆ 每位同学请准备好教材，上课之前，请务必先自学教材！这是前提！（授人以鱼不如授人以渔，大学要学会自学的能力）
- ◆ 课堂教学：通过PPT、场景分析、习题讲解相结合的方式，在课堂讲授本课程各章节的知识单元

往年课程情况简介

2. 课后作业:每一次专题结束后，会以电子版的形式将题目及要求发在QQ群 813406975 中

3. 实验：完成实验项目，将理论与实践相结合，全面实践数据结构的各个结构以及结构之间的关系，并掌握如何利用数据结构算法解决实际问题



群名称:2021春HITSZ数据结构4班
群 号:813406975

二、课程内容

章节	主要内容	授课时数	实验时数
第一章 绪论	①数据结构基本概念和研究方法 ②算法与算法分析	4	0
第二章 线性表	①线性表基本概念 ②线性表存储结构及操作 ③栈与队列 ④字符串 ⑤数组	8	3
第三章 树	①树及二叉树基本概念 ②树及二叉树存储结构及操作 ③线索二叉树 ④树的应用	8	4
第四章 图	①图的基本概念 ②图的存储结构及操作 ③图的连通性（最小生成树） ④有向无环图（拓扑排序、关键路径） ⑤最短路径	8	3
第五章 查找	①静态查找 ②动态查找 ③哈希表	6	3
第六章 排序	①插入排序 ②选择排序 ③快速排序 ④归并排序 ⑤基数排序	4	3
第七章 文件与外排	①顺序文件 ②索引文件 ③直接存取文件 ④多关键字文件	2	



教材

严蔚敏 吴伟民

清华大学计算机系列教材

第1章 绪论	1
1.1 什么是数据结构	1
1.2 基本概念和术语	4
1.3 抽象数据类型的表示与实现	9
1.4 算法和算法分析	13
1.4.1 算法	13
1.4.2 算法设计的要求	13
1.4.3 算法效率的度量	14
1.4.4 算法的存储空间需求	17
第2章 线性表	18
2.1 线性表的类型定义	18
2.2 线性表的顺序表示和实现	21
2.3 线性表的链式表示和实现	27
2.3.1 线性链表	27
2.3.2 循环链表	27
2.3.3 双向链表	27
2.4 一元多项式的表示及相加	39
第3章 栈和队列	44
3.1 栈	44
3.1.1 抽象数据类型栈的定义	44
3.1.2 栈的表示和实现	45
3.2 栈的应用举例	48
3.2.1 数制转换	48
3.2.2 括号匹配的检验	49
3.2.3 行编辑程序	49
3.2.4 迷宫求解	50
3.2.5 表达式求值	52
** 3.3 栈与递归的实现	54
3.4 队列	58
3.4.1 抽象数据类型队列的定义	58
3.4.2 链队列——队列的链式表示和实现	60
3.4.3 循环队列——队列的顺序表示和实现	63
** 3.5 离散事件模拟	65
第4章 串	70
4.1 串类型的定义	70
4.2 串的实现和表示	72

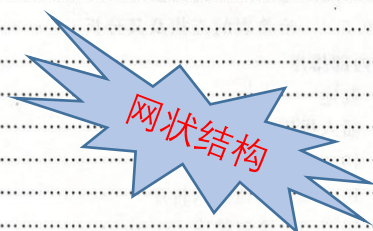
绪论

4.2.1 定长顺序存储表示	73
4.2.2 堆分配存储表示	75
4.2.3 串的块链存储表示	78
** 4.3 串的模式匹配算法	79
4.3.1 求子串位置的定位函数 Index(S,T,pos)	79
4.3.2 模式匹配的一种改进算法	80
4.4 串操作应用举例	84
4.4.1 文本编辑	84
** 4.4.2 建立词索引表	86
第5章 数组和广义表	90
5.1 数组的定义	90
5.2 数组的顺序表示和实现	91
5.3 矩阵的压缩存储	95
5.3.1 特殊矩阵	95
5.3.2 稀疏矩阵	96
5.4 广义表的定义	106
5.5 广义表的存储结构	109
** 5.6 m 元多项式的表示	110
** 5.7 广义表的递归算法	112
5.7.1 求广义表的深度	113
5.7.2 复制广义表	115
5.7.3 建立广义表的存储结构	115
第6章 树和二叉树	118
6.1 树的定义和基本术语	118
6.2 二叉树	121
6.2.1 二叉树的定义	121
6.2.2 二叉树的性质	123
6.2.3 二叉树的存储结构	126
6.3 遍历二叉树和线索二叉树	128
6.3.1 遍历二叉树	128
6.3.2 线索二叉树	132
6.4 树和森林	135
6.4.1 树的存储结构	135
6.4.2 森林与二叉树的转换	137
6.4.3 树和森林的遍历	138
** 6.5 树与等价问题	139
6.6 赫夫曼树及其应用	144
6.6.1 最优二叉树(赫夫曼树)	144
6.6.2 赫夫曼编码	146
** 6.7 回溯法与树的遍历	149
** 6.8 树的计数	152

线性结构

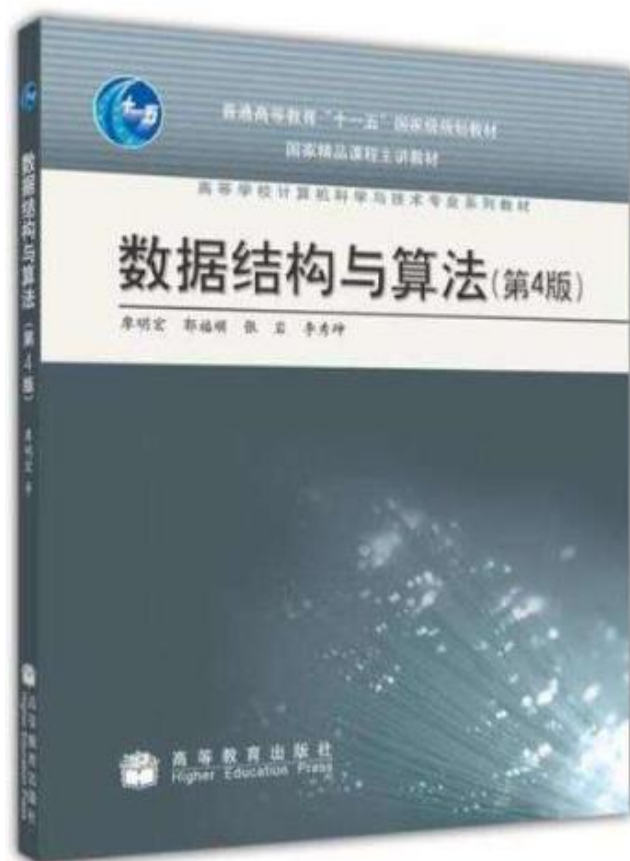
层次结构

第7章 图	156
7.1 图的定义和术语	156
7.2 图的存储结构	160
7.2.1 数组表示法	161
7.2.2 邻接表	163
7.2.3 十字链表	164
7.2.4 邻接多重表	166
7.3 图的遍历	167
7.3.1 深度优先搜索	167
7.3.2 广度优先搜索	169
7.4 图的连通性问题	170
7.4.1 无向图的连通分量和生成树	170
** 7.4.2 有向图的强连通分量	172
7.4.3 最小生成树	173
** 7.4.4 关节点和重连通分量	176
7.5 有向无环图及其应用	179
7.5.1 拓扑排序	180
7.5.2 关键路径	183
7.6 最短路径	186
7.6.1 从某个源点到其余各顶点的最短路径	187
7.6.2 每一对顶点之间的最短路径	190
第8章 动态存储管理	193
8.1 概述	193
8.2 可利用空间表及分配方法	195
8.3 边界标识法	198
8.3.1 可利用空间表的结构	198
8.3.2 分配算法	199
8.3.3 回收算法	201
8.4 伙伴系统	203
8.4.1 可利用空间表的结构	203
8.4.2 分配算法	204
8.4.3 回收算法	205
** 8.5 无用单元收集	206
** 8.6 存储紧缩	212
第9章 查找	214
9.1 静态查找表	216
9.1.1 顺序表的查找	216
9.1.2 有序表的查找	218
** 9.1.3 静态树表的查找	222
9.1.4 索引顺序表的查找	225
9.2 动态查找表	226
9.2.1 二叉排序树和平衡二叉树	227
9.2.2 B-树和B ⁺ 树	238



章 绪论

9.2.3 键树	247
9.3 哈希表	251
9.3.1 什么是哈希表	251
9.3.2 哈希函数的构造方法	253
9.3.3 处理冲突的方法	256
9.3.4 哈希表的查找及其分析	259
第10章 内部排序	263
10.1 概述	263
10.2 插入排序	265
10.2.1 直接插入排序	265
10.2.2 其他插入排序	266
10.2.3 希尔排序	271
10.3 快速排序	272
10.4 选择排序	277
10.4.1 简单选择排序	277
10.4.2 树形选择排序	278
10.4.3 堆排序	279
10.5 归并排序	283
10.6 基数排序	284
10.6.1 多关键字的排序	284
10.6.2 链式基数排序	286
10.7 各种内部排序方法的比较讨论	288
第11章 外部排序	293
11.1 外存信息的存取	293
11.2 外部排序的方法	295
** 11.3 多路平衡归并的实现	297
** 11.4 置换-选择排序	299
** 11.5 最佳归并树	304
第12章 文件	306
12.1 有关文件的基本概念	306
12.2 顺序文件	308
12.3 索引文件	311
12.4 ISAM文件和VSAM文件	313
12.4.1 ISAM文件	313
12.4.2 VSAM文件	316
12.5 直接存取文件(散列文件)	317
12.6 多关键字文件	319
12.6.1 多重表文件	319
12.6.2 倒排文件	319
附录A 名词索引	322
附录B 函数索引	329
参考书目	334



参 考 教 材

数据结构与算法 (第4版)

编著 廖明宏 郭福顺 张岩 李秀坤

高等教育出版社

参考资料

- ✓ Ellis Horowitz, Sartaj Sahni, Susan Anderson-Freed 著, 李建中, 张岩, 李治军 译. Fundamentals of Data Structures in C. 机械工业出版社. 2006.
- ✓ Sartaj Sahni 著, 汪诗林, 孙晓东. Data Structures, Algorithms, and Applications in C++. 机械工业出版社. 2006.
- ✓ 李秀坤, 张岩, 李治军, 娄久. 数据结构与算法实验教程. 高等教育出版社. 2009.

教学安排：讲课学时40，实验学时16

考核要求：期末考试占70%

实验成绩占20%

平时作业占10%



	2020级计算机专业13-16班	
上课时间	3-10, 12, 13周 周三 1-2节	3-12周 周五 1-2节
上课地点	T5304教室	T5204教室
考试时间	实验、考试的时间地点待定	

注意事项

- 1、记住上课时间，按时上课，不旷课
- 2、课前预习，培养自学能力
- 3、按时完成并提交作业（线下 或 E-mail）
- 4、尽可能保持QQ在线，所有课程通知、资料在QQ群发布
- 5、课堂点名



群名称:2021春HITSZ数据结构4班
群 号:813406975



课后答疑

授课老师：张正

地址：信息楼1519

邮件：zhengzhang@hit.edu.cn

课程助教：

黄裕涛（13班），QQ：278580633

李逢君（14班），QQ：543877815

李浩宇（15班），QQ：412689719

何素妍（16班），QQ：1872954023

赵书光，QQ：1329777148

实验课指导老师：汪花梅老师

地址：T2801

QQ：314396326



群名称:2021春HITSZ数据结构4班
群 号:813406975



创新事业群UC事业部-高级搜索推荐算法专家/资深专家-国内信息流

发布时间:	2019-08-16	工作地点:	北京	工作年限:	五年以上
所属部门:	阿里集团	学 历:	本科	招聘人数:	若干

岗位描述:

- 1. 负责全网文本排序、文本理解（问题识别、智能问答）、图片垂直搜索方向的大规模分布式机器学习算法\深度学习算法的开发;
- 2. 带领团队承担公司机器学习/深度学习算法平台的建设，提升整体算法团队的深度和效率;

岗位要求:

- 1. 在业内该方向上有丰富的实战经验和一定的影响力，包括但不限于：深度学习、LTR、自然语言处理、图模型、推荐算法等;
- 2. 算法理论、工程功底扎实，优秀的编程能力：分布式计算系统，精通C/C++语言和数据结构，编码基本功扎实;
- 3. 具有优秀的分析和解决实际问题的能力和态度;
- 4. 对搜索/推荐业务有着不错的理解，有一定的创新能力，能够有效整合资源解决业务问题;

腾讯云 SDK 终端研发工程师

职位描述

负责腾讯云终端 SDK 的功能开发和效果优化相关：

- (1) 包括“IM SDK”，用于实现点对点和群组内的文本、图片和文件等即时消息的通信。
- (2) 包括“RTC SDK”，用于视频通话，在线会议，互动教育等领域的实时音视频通信。
- (3) 包括“直播SDK”，用于实现直播推拉流、互动连麦等电商直播和娱乐直播的终端能力。
- (4) 包括“短视频SDK”，用于实现视频录制、特效编辑、视频发布、视频点播等一体化的在线流媒体能力。
- (5) 团队技术氛围浓厚，培训体系完善，产品线多样，有经验丰富的“老司机”手把手帮你入门音视频和网络技术，诚盼客户端开发经验丰富且对以上任一方向感兴趣的技术达人加入我们。

本岗位的要求包括：

- (1) 四年以上的客户端开发经验，至少两年以上的 C++ 语言开发经验，熟练使用 C++11 并能够用其构建大型客户端软件项目。
- (2) 有扎实的数据结构和算法基础，以及客户端调试与性能优化经验。
- (3) 充分理解操作系统在内存管理、线程调度和进程管理等方面的基本知识，并能够运用这些知识优化当前程序的运行效率和稳定性。
- (4) 拥有多线程开发经验，能够写出稳定可靠的多线程代码。
- (5) 有责任心有激情，有良好的团队合作能力和沟通能力。有实用性为导向的 OOP 意识和设计能力。
- (6) 具备良好的工程素养，支持 codereview 文化，能够虚心接受他人对代码的改进建议和问题点评。

计算机视觉工程师 (ARVR、视频、搜索推荐、智能驾驶方向、图像识别、地图、OCR等方向)

华为消费者BG (终端) | 华为技术有限公司

工作地点：上海、北京、深圳、武汉、西安、南京、杭州

招聘对象：海外院校：2019年1月-2021年12月期间毕业的博士生；国内院校：2020年1月-2021年12月期间毕业的博士生；

岗位要求：

掌握如下一项或多项技能：

- 1、图像处理与模式识别、计算机视觉、图形学、计算机、自动控制、通信、信息、电子、应用数学、统计学等相关毕业；
- 2、深入理解图像分析/处理专业知识，熟悉常用图形图像处理算法；
- 3、熟悉SLAM/VIO算法，掌握特征点提取、连续图像帧深度信息提取算法；熟悉视觉传感器和IMU传感器的标定，熟悉基于EKF的组合导航算法，对于利用摄像头进行空间标识、标志点提取等有深入研究；有VINS、PTAM、LSD-SLAM、ORB-SLAM、三维重建或目标跟踪/检测等项目开发经验优先；
- 4、专业研究方向为显示、渲染、GPU相关方向，对显示渲染方案框架理解深入，深入理解安卓和Linux或Windows系统渲染技术和Pipeline，熟悉相关渲染软件栈，熟悉OpenGL、OpenGL ES、vulkan、DirectX中一种的实现原理，对GPU的功耗效率优化设计有较深入理解；
- 5、熟练掌握C/C++/Matlab或者Python，熟悉常用的数据结构和算法，掌握多进程/多线程处理，熟悉面向对象思想和方法，了解STL；

职位:阿里巴巴 前端专家P7 和 资深前端工程师都P6

招聘方式: 内推到阿里巴巴内部招聘系统

职位描述

- 1、负责面向销售、商家、对内运营体系等的各种企业级泛中台产品应用的前端架构设计和研发工作;
- 2、参与本地生活大前端技术体系建设, 包括组件开发、工程效能体系化建设、多端架构设计、Node全栈、稳定性监控保障体系等;
- 3、深入理解业务和产品, 不断提升产品的用户体验, 技术驱动业务的发展;
- 4、关注相关前沿技术研究, 通过新技术服务团队和业务;

职位要求

- 1、前端HTML/CSS/JavaScript等基础技术扎实, 数据结构、网络等基础知识掌握牢固, 熟悉浏览器原理机制;
- 2、精通至少一种主流前端框架, 如React/Vue/Angular等, 能深入理解其设计原理, 并对其他类似框架有所了解, 有库或框架设计者优先;
- 3、理解前端工程化, 熟练使用相关工具, 如Npm/Yarn/Webpack/Rollup等, 有对前端工程化有自己的实践者优先;
- 4、熟悉Node相关知识, 有Koa/Egg/Express等的实践经验, 若有其他后端语言如Java/PHP/Python/Golang等的开发经验优先;5、熟悉iOS/Android等客户端相关知识, 具备ReactNative/Weex/Flutter能力和经验者优先;

简单感觉一下。。。

2020年6月腾讯数据结构与算法面试题

讲一讲，如何构造一致性哈希算法？

如何知道二叉树的深度？

算法题：二叉树层序遍历，并要求每层打印出一个换行符。

怎么求一个二叉树的深度？手撕代码？

请你说一下，B+树和B-树？

描述二叉树的遍历方式，写了个非递归的前序遍历。

写一个函数，找到一个文件夹下所有文件，包括子文件夹？

现在有一个单向链表，谈一谈，如何判断链表中是否出现了环？

写一个算法，可以将一个二维数组顺时针旋转90度，说一下思路。

一个数组，除一个元素外其它都是两两相等，求那个元素？

求一个数组中连续子向量的最大和

介绍一下，排序都有哪几种方法？时间复杂度，空间复杂度？

介绍一下，归并排序的原理是什么？

介绍一下，堆排序的原理是什么？

谈一谈，如何得到一个数据流中的中位数？

介绍一下，堆与栈的不同是什么？

2020年腾讯C++面试题

归并排序求逆序？

红黑树了解吗？b+树了解吗？

红黑树以及其查找复杂度？

如何减少哈希冲突？

利用栈解决一个问题，写出算法？

C++智能指针知道吗？

私有继承的作用是什么？

if-else 和 switch在汇编层面相同吗？

2020年腾讯C++面试题

排序算法有哪些？

冒泡和快排的区别？

知道B树和红黑树吗？实现原理是啥？应用？

C++多态实现机制？

单链表的反转机制？