

一、选择题

1.

5. 以下关于数据结构的说法中，正确的是（ ）。

- A. 数据的逻辑结构独立于其存储结构
- B. 数据的存储结构独立于其逻辑结构
- C. 数据的逻辑结构唯一决定了其存储结构
- D. 数据结构仅由其逻辑结构和存储结构决定

5. A

数据的逻辑结构是从面向实际问题的角度出发的，只采用抽象表达方式，独立于存储结构，数据的存储方式有多种不同的选择；而数据的存储结构是逻辑结构在计算机上的映射，它不能独立于逻辑结构而存在。数据结构包括三个要素，缺一不可。

2.

5. 设线性表中有 $2n$ 个元素，（ ）在单链表上实现要比在顺序表上实现效率更高。

- A. 删除所有值为 x 的元素
- B. 在最后一个元素的后面插入一个新元素
- C. 顺序输出前 k 个元素
- D. 交换第 i 个元素和第 $2n-i-1$ 个元素的值 ($i=0, \dots, n-1$)

5. A

对于 A，在单链表和顺序表上实现的时间复杂度都为 $O(n)$ ，但后者要移动很多元素，因此在单链表上实现效率更高。对于 B 和 D，顺序表的效率更高。C 无区别。

3.

10. 【2014 统考真题】循环队列放在一维数组 $A[0..M-1]$ 中， $end1$ 指向队头元素， $end2$ 指向队尾元素的后一个位置。假设队列两端均可进行入队和出队操作，队列中最多能容纳 $M-1$ 个元素。初始时空。下列判断队空和队满的条件中，正确的是（ ）。

- | | |
|---------------------------------------|--------------------------------------|
| A. 队空: $end1 == end2$; | 队满: $end1 == (end2 + 1) \bmod M$ |
| B. 队空: $end1 == end2$; | 队满: $end2 == (end1 + 1) \bmod (M-1)$ |
| C. 队空: $end2 == (end1 + 1) \bmod M$; | 队满: $end1 == (end2 + 1) \bmod M$ |
| D. 队空: $end1 == (end2 + 1) \bmod M$; | 队满: $end2 == (end1 + 1) \bmod (M-1)$ |

10. A

$end1$ 指向队头元素，可知出队操作是先将 $A[end1]$ 读数，然后 $end1$ 再加 1。 $end2$ 指向队尾元素的后一个位置，可知入队操作是先存数到 $A[end2]$ ，然后 $end2$ 再加 1。若用 $A[0]$ 存储第一个元素，队列初始时，入队操作是先把数据放到 $A[0]$ 中，然后 $end2$ 自增，即可知 $end2$ 初值为 0；而 $end1$ 指向的是队头元素，队头元素在数组 A 中的下标为 0，所以得知 $end1$ 的初值也为 0，可知队空条件为 $end1 == end2$ ；然后考虑队列满时，因为队列最多能容纳 $M-1$ 个元素，假设队列存储在标为 0 到 $M-2$ 的 $M-1$ 个区域，队头为 $A[0]$ ，队尾为 $A[M-2]$ ，此时队列满，考虑在这种情况下 $end1$ 和 $end2$ 的状态， $end1$ 指向队头元素，可知 $end1=0$ ， $end2$ 指向队尾元素的后一个位置，可知 $end2=M-2+1=M-1$ ，所以可知队满的条件为 $end1 == (end2 + 1) \bmod M$ ，选 A。

注意：考虑这类具体问题时，用一些特殊情况判断往往比直接思考问题能更快地得到答案，并可以画出简单的草图以方便解题。

4.

12. 【2014 统考真题】假设栈初始为空，将中缀表达式 $a/b+(c*d-e*f)/g$ 转换为等价的后缀表达式的过程中，当扫描到 f 时，栈中的元素依次是 ()。

- A. $+(*-$ B. $+(*-*$
C. $/+(*-*$ D. $/+--*$

12. B

将中缀表达式转换为后缀表达式的算法思想如下：

从左向右开始扫描中缀表达式；

遇到数字时，加入后缀表达式；

遇到运算符时：

a. 若为 '('，入栈；

b. 若为 ')'，则依次把栈中的运算符加入后缀表达式，直到出现 '('，从栈中删除 '('；

c. 若为除括号外的其他运算符，当其优先级高于除 '(' 外的栈顶运算符时，直接入栈。否则从栈顶开始，依次弹出比当前处理的运算符优先级高和优先级相等的运算符，直到一个比它优先级低的或遇到了一个左括号为止。

当扫描的中缀表达式结束时，栈中的所有运算符依次出栈加入后缀表达式。

待处理序列	栈	后缀表达式	当前扫描元素	动作
$a/b+(c*d-e*f)/g$			a	a 加入后缀表达式
$/b+(c*d-e*f)/g$		a	/	/ 入栈
$b+(c*d-e*f)/g$	/	a	b	b 加入后缀表达式
$+(c*d-e*f)/g$	/	ab	+	+ 优先级低于栈顶的 /，弹出 /
$+(c*d-e*f)/g$		ab/	+	+ 入栈
$(c*d-e*f)/g$	+	ab/	((入栈
$c*d-e*f)/g$	+(ab/	c	c 加入后缀表达式
$*d-e*f)/g$	+(ab/c	*	栈顶为 (，* 入栈
$d-e*f)/g$	+(*	ab/c	d	d 加入后缀表达式
$-e*f)/g$	+(*	ab/cd	-	- 优先级低于栈顶的 *，弹出 *
$-e*f)/g$	+(ab/cd*	-	栈顶为 (，- 入栈
$e*f)/g$	+(-	ab/cd*	e	e 加入后缀表达式
$*f)/g$	+(-	ab/cd*e	*	* 优先级高于栈顶的 -，* 入栈
$f)/g$	+(-*	ab/cd*e	f	f 加入后缀表达式
$) /g$	+(-*	ab/cd*ef)	把栈中 (之前的符号加入表达式
$/g$	+	ab/cd*ef*	/	/ 优先级高于栈顶的 +，/ 入栈
g	+/	ab/cd*ef*	g	g 加入后缀表达式
	+/	ab/cd*ef*-g		扫描完毕，运算符依次退栈加入表达式
		ab/cd*ef*-g/+		完成

由此可知，当扫描到 f 时，栈中的元素依次是 $+(-*$ ，选 B。

在此，以上面给出的中缀表达式为例，给出中缀表达式转换为前缀或后缀表达式的手工做法。

步骤 1：按照运算符的优先级对所有的运算单位加括号。

式子变成 $((a/b)+(((c*d)-(e*f))/g))$ 。

步骤 2：转换为前缀或后缀表达式。

前缀：把运算符符号移动到对应的括号前面，式子变成 $+(/(ab)/(-(*(cd)*(ef))g))$ 。

把括号去掉： $+/ab/-*cd*efg$ 前缀式子出现。

后缀：把运算符符号移动到对应的括号后面，式子变成 $((ab)/(((cd)*(ef)*-g)/+)$ 。

把括号去掉： $ab/cd*ef*-g/+$ 后缀式子出现。

当题目要求直接求前缀或后缀表达式时，这种方法会比上一种方法快捷得多。

8. 【2015 统考真题】已知字符串 S 为 'abaabaabacacaabaabcc', 模式串 t 为 'abaabc'。采用 KMP 算法进行匹配, 第一次出现“失配”($s[i] \neq t[j]$) 时, $i=j=5$, 则下次开始匹配时, i 和 j 的值分别是 ()。

8. C

编号	0	1	2	3	4	5
t	a	b	a	a	b	c
next	-1	0	0	1	1	2

6.

A. 在中序线索树中, 若某结点有右孩子, 则其后继结点是它的右子树的最左下结点
B. 在中序线索树中, 若某结点有左孩子, 则其前驱结点是它的左子树的最右下结点
C. 线索二叉树是利用二叉树的 $n+1$ 个空指针来存放结点的前驱和后继信息的
D. 每个结点通过线索都可以直接找到它的前驱和后继

27. D

7.

A. 求关键路径的方法 B. 求最短路径的 Dijkstra 算法

12. C

考试是反过来问的

8. 没找到一样的 但考试那题选 A (B 树的每个结点有 m 个子树)

2. 下列关于 m 阶 B 树的说法中, 错误的是 ()。

- A. 根结点至多有 m 棵子树
- B. 所有叶结点都在同一层次上
- C. 非叶结点至少有 $m/2$ (m 为偶数) 或 $(m+1)/2$ (m 为奇数) 棵子树
- D. 根结点中的数据是有序的

3. 以下关于 m 阶 B 树的说法中, 正确的是 ()。

- I. 每个结点至少有两棵非空子树
- II. 树中每个结点至多有 $m-1$ 个关键字
- III. 所有叶结点在同一层
- IV. 插入一个元素引起 B 树结点分裂后, 树长高一层

- A. I、II
- B. II、III
- C. III、IV
- D. I、II、IV

4. 【2009 统考真题】下列叙述中, 不符合 m 阶 B 树定义要求的是 ()。

- A. 根结点至多有 m 棵子树
- B. 所有叶结点都在同一层上
- C. 各结点内关键字均升序或降序排列
- D. 叶结点之间通过指针链接

2. C

除根结点外的所有非终端结点至少有 $\lceil m/2 \rceil$ 棵子树。对于根结点, 最多有 m 棵子树, 若其不是叶结点, 则至少有 2 棵子树。

3. B

每个非根的内部结点必须至少有 $\lceil m/2 \rceil$ 棵子树, 而根结点至少要有两棵子树, 所以 I 不正确。II、III 显然正确。对于 IV, 插入一个元素引起 B 树结点分裂后, 只要从根结点到该元素插入位置的路径上至少有一个结点未滿, B 树就不会长高, 所以 IV 不正确。

4. D

m 阶 B 树不要求将各叶结点之间用指针链接。选项 D 描述的实际上是 B+树。

9.

9. 数据序列 $F = \{2, 1, 4, 9, 8, 10, 6, 20\}$ 只能是下列排序算法中的 () 两趟排序后的结果。

- A. 快速排序
- B. 冒泡排序
- C. 选择排序
- D. 插入排序

9. A

若为插入排序, 则前三个元素应该是有序的, 显然不对。而冒泡排序和选择排序经过两趟排序后应该有两个元素处于最终位置 (最左/右端), 无论是按从小到大还是从大到小排序, 数据序列中都没有两个满足这样的条件的元素, 因此只可能选 A。

另解: 先写出排好序的序列, 并和题中的序列做对比。

题中序列: 2 1 4 9 8 10 6 20

已排好序序列: 1 2 4 6 8 9 10 20

在已排好序的序列中, 与题中序列相同元素的有 4、8 和 20, 最左和最右两个元素与题中的序列不同, 故不可能是冒泡排序、选择排序或插入排序。

10.

5. 设二叉排序树中关键字由 1 到 1000 的整数构成，现要查找关键字为 363 的结点，下述关键字序列中，不可能是在二叉排序树上查找的序列是（ ）。

- A. 2, 252, 401, 398, 330, 344, 397, 363
- B. 924, 220, 911, 244, 898, 258, 362, 363
- C. 925, 202, 911, 240, 912, 245, 363
- D. 2, 399, 387, 219, 266, 382, 381, 278, 363

5. C

在二叉排序树上查找时，先与根结点值进行比较，若相同，则查找结束，否则根据比较结果，沿着左子树或右子树向下继续查找。根据二叉排序树的定义，有左子树结点值 \leq 根结点值 \leq 右子树结点值。C 序列中，比较 911 关键字后，应转向其左子树比较 240，左子树中不应出现比 911 更大的数值，但 240 竟有一个右孩子结点值为 912，所以不可能是正确的序列。

二、填空题

1.

【2009 统考真题】设栈 S 和队列 Q 的初始状态均为空，元素 *abcdefg* 依次进入栈 S。若每个元素出栈后立即进入队列 Q，且 7 个元素出队的顺序是 *bdcfeag*，则栈 S 的容量至少是（ ）。

答案 3

2. 层次遍历二叉树需要使用 队列

3. 先序序列 1, 2, 3, 4 对应的二叉树有多少种 14

4. n 个结点二叉树用三叉链表存储，有多少个空指针域 $n+2$

5. 糖宝答案是 20 我记不得题目了

6. 有 2015 个结点的完全二叉树的叶结点数 1008

7. 好像是问 n 个结点二路归并的趟数 $\text{ceil}(\log_2 n)$

8.

对关键码序列 {23, 17, 72, 60, 25, 8, 68, 71, 52} 进行堆排序，输出两个最小关键码后的剩余堆是（ ）。

筛选法初始建堆为 {8, 17, 23, 52, 25, 72, 68, 71, 60}，输出 8 后重建的堆为 {17, 25, 23, 52, 60, 72,

68, 71}，输出 17 后重建的堆为 {23, 25, 68, 52, 60, 72, 71}。

9. n 个顶点的无向图各顶点最大度 $n-1$

10. n 个元素，折半查找，失败结点数 $n+1$

三、简答题

1. n 个结点， e 条边。邻接矩阵存储，问有多少个非零元素。

若用邻接表存储，有奇数个表结点，该图一定是有向图还是无向图，为什么。

2. 森林与二叉树的转换

四、应用题

1. 如果一棵非空 k ($k \geq 2$) 叉树 T 中每个非叶结点都有 k 个孩子，则称 T 为正则 k 叉树。请回答下列问题并给出推导过程。

(1) 若 T 有 m 个非叶结点，则 T 中的叶结点有多少个？

(2) 若 T 的高度为 h (单结点的树 $h=1$)，则 T 的结点数最多为多少个？

最少为多少个？

解答：

(1) 根据定义，正则 k 叉树中仅含有两类结点：叶结点（个数记为 n_0 ）和度为 k 的分支结点（个数记为 n_1 ）。树 T 中的结点总数 $n = n_0 + n_1 = n_0 + m$ 。树中所含的边数 $e = n - 1$ ，这些边均为 m 个度为 k 的结点发出的，即 $e = m \times k$ 。整理得： $n_0 + m = m \times k + 1$ ，故 $n_0 = (k-1) \times m + 1$ 。（3分）

(2) 高度为 h 的正则 k 叉树 T 中，含最多结点的树形为：除第 h 层外，第 1 到第 $h-1$ 层的结点都是度为 k 的分支结点；而第 h 层均为叶结点，即树是“满”树。此时第 j ($1 \leq j \leq h$) 层结点数为 k^{j-1} ，结点总数 M_1 为：

$$M_1 = \sum_{j=1}^h k^{j-1} = \frac{k^h - 1}{k - 1} \quad (3\text{分})$$

含最少结点的正则 k 叉树的树形为：第 1 层只有根结点，第 2 到第 $h-1$ 层仅含 1 个分支结点和 $k-1$ 个叶结点，第 h 层有 k 个叶结点。即除根外第 2 到第 h 层中每层的结点数均为 k ，故 T 中所含结点总数 M_2 为：

$$M_2 = 1 + (h-1) \times k \quad (2\text{分})$$

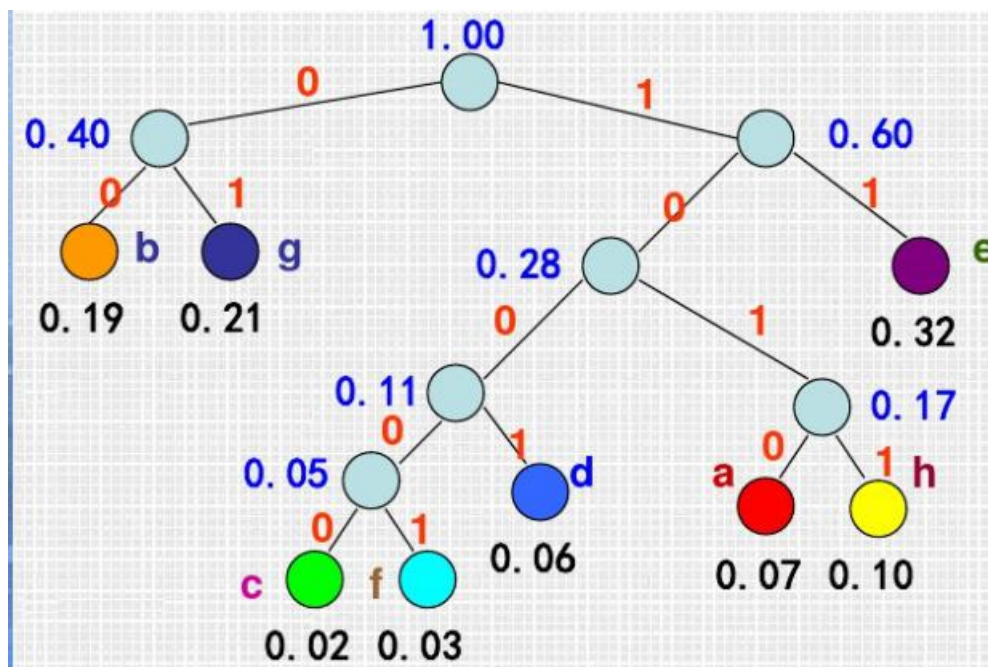
2.

例：设通信用的电文由字符集 {a, b, c, d, e, f, g, h} 中的字母构成，这8个字母在电文中出现的概率分别为 { 0.07, 0.19, 0.02, 0.06, 0.32, 0.03, 0.21, 0.10 }，试为这8个字母设计哈夫曼编码。

1) 画出哈夫曼树

2) 最长编码与对应字母，最短编码与对应字母

3) 传输“ABCDEFGH”，原先传输一个字母需要 8bit,求使用哈夫曼编码后的压缩率



答案：1)

2)

3) $28/64=0.4375$

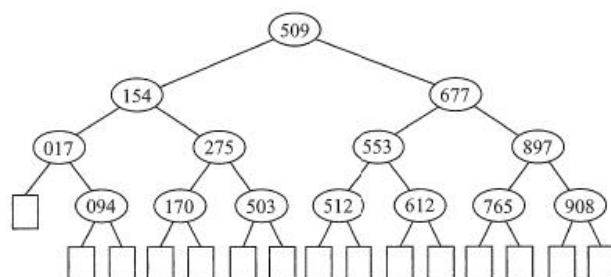
3.

设有序顺序表中的元素依次为 017, 094, 154, 170, 275, 503, 509, 512, 553, 612, 677, 765, 897, 908。

- 1) 试画出对其进行折半查找的判定树。
- 2) 若查找 275 或 684 的元素，将依次与表中的哪些元素比较？
- 3) 计算查找成功的平均查找长度和查找不成功的平均查找长度。

2. 解答：

1) 判定树如下图所示。



- 2) 若查找 275，依次与表中元素 509, 154, 275 进行比较，共比较 3 次。若查找 684，依次与表中元素 509, 677, 897, 765 进行比较，共比较 4 次。
- 3) 在查找成功时，会找到图中的某个圆形结点，其平均查找长度为

$$ASL_{\text{成功}} = \frac{1}{14} \sum_{i=1}^{14} C_i = \frac{1}{14} (1 + 2 \times 2 + 3 \times 4 + 4 \times 7) = \frac{45}{14}$$

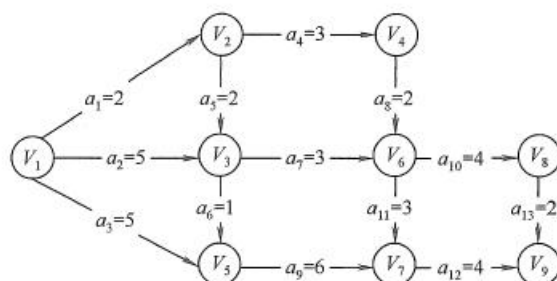
在查找失败时，会找到图中的某个方形结点，但这个结点是虚构的，最后一位的比较元素为其父结点（圆形结点），故其平均查找长度为

$$ASL_{\text{不成功}} = \frac{1}{15} \sum_{i=0}^{14} C'_i = \frac{1}{15} (3 \times 1 + 4 \times 14) = \frac{59}{15}$$

4.

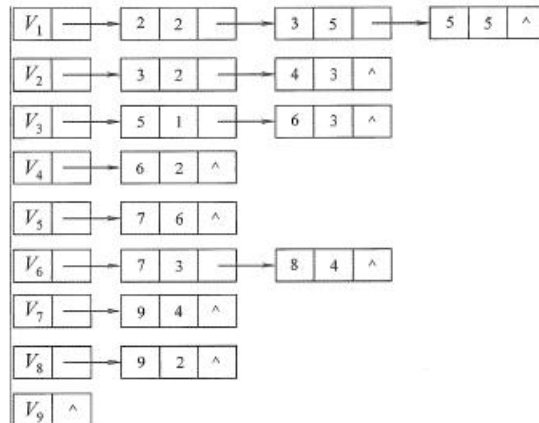
7. 下图所示为一个用 AOE 网表示的工程。

- 1) 画出此图的邻接表表示。
- 2) 完成此工程至少需要多少时间？
- 3) 指出关键路径。
- 4) 哪些活动加速可以缩短完成工程所需的时间？



7. 解答:

1) 该图的邻接表表示如下图所示。



求关键路径的算法如下:

① 输入 e 条弧 $\langle j, k \rangle$, 建立 AOE 网的存储结构。

② 从源点 v_1 出发, 令 $v_e(1) = 0$, 求 $v_e(j)$, $2 \leq j \leq n$ 。

③ 从汇点 v_n 出发, 令 $v_l(n) = v_e(n)$, 求 $v_l(i)$, $1 \leq i \leq n-1$ 。

④ 根据各顶点的 v_e 和 v_l 值, 求每条弧 s (活动) 的最早开始时间 $e(s)$ 和最晚开始时间 $l(s)$,

其中 $e(s) = l(s)$ 为关键活动。

2) 根据以上算法可以得到至少需要时间 16。

3) 关键路径为 $(V_1, V_3, V_5, V_7, V_9)$ 。

4) 活动 a_2, a_6, a_9, a_{12} 加速, 可以缩短工程所需的时间。

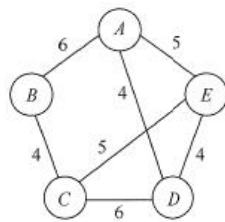
5. Prim 求最小生成树 (大概就下面这样吧 图不一样)

5. 【2017 统考真题】使用 Prim 算法求带权连通图的最小 (代价) 生成树 (MST)。请回答下列问题:

1) 对下列图 G , 从顶点 A 开始求 G 的 MST, 依次给出按算法选出的边。

2) 图 G 的 MST 是唯一的吗?

3) 对任意的带权连通图, 满足什么条件时, 其 MST 是唯一的?



1) Prim 算法属于贪心策略。算法从一个任意的顶点开始，一直长到覆盖图中的所有顶点为止。算法的每一步在连接树集合 S 的顶点和其他顶点的边中，选择一条使得树的总权重增加最小的边加入集合 S 。当算法终止时， S 就是最小生成树。

① S 中顶点为 A ，候选边为 $(A, D), (A, B), (A, E)$ ，选择 (A, D) 加入 S 。

② S 中顶点为 A, D ，候选边为 $(A, B), (A, E), (D, E), (C, D)$ ，选择 (D, E) ，加入 S 。

③ S 中顶点为 A, D, E ，候选边为 $(A, B), (C, D), (C, E)$ ，选择 (C, E) 加入 S 。

④ S 中顶点为 A, D, E, C ，候选边为 $(A, B), (B, C)$ ，选择 (B, C) 加入 S 。

⑤ S 就是最小生成树。

依次选出的边为

$(A, D), (D, E), (C, E), (B, C)$

2) 图 G 的 MST 是唯一的。第一小题的最小生成树包括了图中权值最小的 4 条边，其他边都比这 4 条边大，所以此图的 MST 唯一。

3) 当带权连通图的任意一个环中所包含的边的权值均不相同，其 MST 是唯一的。此题不要求回答充分必要条件，所以回答一个限制边权值的充分条件即可。

五、算法题

1. 新冠传播，每个人与密切接触者形成边，构成有向图，寻找 0 号感染者到其他任意感染者之间的最短距离。

Dijkstra 算法

2. 数组，偶数放偶数位，奇数放奇数位

双指针，快排思想

By: 20 级 kuwernv(xzy)、糖矿 (wzk)、CH₃CHOHCH₃ (ybc)

基本上王道考研上都有，很多题也是课本习题册上的，懂得都懂