

三、下面的数据表给出了在一篇有 19710 个词的英文行文中出现最普遍的 15 个词的出现频度。

(1) 假设一个英文字符等价于 $\log_2 6 = 4.7010$ bits, 那么这些词按 bits 计的平均长度是多少? (5 分)

(2) 假定一篇正文仅由上述的数据表中的词组成, 那么它们的最佳编码是什么? 平均长度是多少? (5 分)

词	the	of	a	to	and	in	that	he
出现频度	1192	677	541	518	462	450	242	195

词	is	at	on	for	his	are	be
出现频度	190	181	174	157	138	124	123

四、设有 12 个数据 {25, 40, 33, 47, 12, 66, 72, 87, 94, 22, 5, 58}, 它们存储在散列表中, 利用双散列解决冲突, 要求为插入新数据的平均查找次数不超过 3 次。

(1) 该散列表的大小 m 应设计多大? (4 分)

(2) 试为该散列表设计相应的散列函数(用除留余数法)和计算寻找下一个“空位”时向前跨步步长的再散列函数。(6 分)

(3) 顺次将各个数据散列到表中。(6 分)

(4) 计算查找成功的平均查找次数。(2 分)

五、二项式 $(a + b)^n$ 展开式的系数为

$$C(n, 0) = 1, \quad C(n, n) = 1, \quad \text{对于 } n \geq 0$$

$$C(n, k) = C(n-1, k) + C(n-1, k-1), \quad \text{对于 } 0 < k < n$$

形成著名的杨辉三角形, 如图所示。

(1) 试写一个递归算法, 根据以上公式生成 $C(n, k)$ 。(6 分)

(2) 试画出计算 $C(6, 4)$ 的递归树。(4 分)

(3) 试写一个非递归算法, 既不用数组也不用栈, 对于任意的 $0 \leq k \leq n$ 计算 $C(n, k)$ 。(6 分)

数据结构和程序设计

				1						$n=0$			
				1	1					$n=1$			
				1	2	1				$n=2$			
				1	3	3	1			$n=3$			
				1	4	6	4	1		$n=4$			
				1	5	10	10	5	1	$n=5$			
				1	6	15	20	15	6	1	$n=6$		
				1	7	21	35	35	21	7	1	$n=7$	
				1	8	28	56	70	56	28	8	1	$n=8$

六、一个双端队列 *deque* 是限定在两端 *end1*、*end2* 都可进行插入和删除的线性表。队空条件是 $end1 = end2$ 。若用顺序方式来组织双端队列，试根据下列要求，定义双端队列的结构，并给出在指定端 i ($i = 1, 2$) 的插入 *enq* 和删除 *deq* 操作的实现。(12分)

- (1) 当队满时，最多只能有一个元素空间可以是空的。
- (2) 在做两端的插入和删除时，队列中其它元素一律不动。

七、设 L_1 和 L_2 是两个有序顺序表，分别包含 n_1 和 n_2 个整数。

- (1) 试利用折半查找的思想描述如何查找两个表中的中值。(4分)
(所有整数)
- (2) 试用类 Pascal 或类 C 语言编写一个算法，实现你的想法。(6分)

八、一棵高度为 h 的满 k 叉树有如下性质：根结点所在层次为 0；第 h 层上的结点都是叶结点；其余各层上每个结点都有 k 棵非空子树，如果按层次自顶向下，同一层自左向右，顺序从 1 开始对全部结点进行编号，试问：

- ☛ 各层的结点个数是多少？(3分)
- ☛ 编号为 i 的结点的双亲结点(若存在)的编号是多少？(3分)
- ☛ 编号为 i 的结点的第 m 个孩子结点(若存在)的编号是多少？(3分)
- ☛ 编号为 i 的结点有右兄弟的条件是什么？其右兄弟结点的编号是多少？(3分)