

折半查找判定树及平均查找长度

折半查找判定树及平均查找长度

从折半查找的过程看，以有序表的中间记录作为比较对象，并以中间记录将表分割为两个子表，对子表继续上述操作。所以，对表中每个记录的查找过程，可用二叉树来描述，二叉树中的每个结点对应有序表中的一个记录，结点中的值为该记录在表中的位置。通常称这个描述折半查找过程的二叉树为折半查找判定树。

长度为 n 的折半查找判定树的构造方法为：

- (1) 当 $n=0$ 时，折半查找判定树为空；
- (2) 当 $n>0$ 时，折半查找判定树的根结点是有有序表中序号为 $\text{mid}=(n+1)/2$ 的记录，根结点的左子树是与有序表 $r[1] \sim r[\text{mid}-1]$ 相对应的折半查找判定树，根结点的右子树是与 $r[\text{mid}+1] \sim r[n]$ 相对应的折半查找判定树。

例如，长度为10的折半查找判定树的具体生成过程为：

- (1) 在长度为10的有序表中进行折半查找，不论查找哪个记录，都必须先和中间记录进行比较，而中间记录的序号为 $(1+10)/2=5$ （注意是整除即向下取整），即判定树的根结点是5，如图7-2(a)所示；
- (2) 考虑判定树的左子树，即将查找区间调整到左半区，此时的查找区间是 $[1, 4]$ ，也就是说，左分支上为根结点的值减1，代表查找区间的高端high，此时，根结点的左孩子是 $(1+4)/2=2$ ，如图7-2(b)所示；
- (3) 考虑判定树的右子树，即将查找区间调整到右半区，此时的查找区间是 $[6, 10]$ ，也就是说，右分支上为根结点的值加1，代表查找区间的低端low，此时，根结点的右孩子是 $(6+10)/2=8$ ，如图7-2(c)所示；
- (4) 重复(2)(3)步，依次确定每个结点的左右孩子，如图7-2(d)所示。



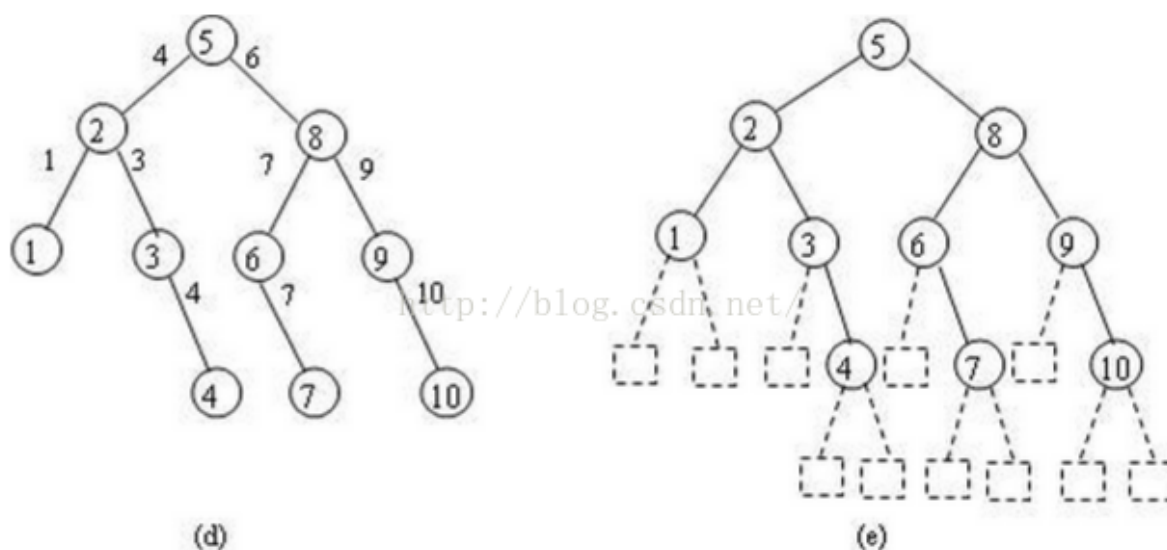


图 7-2 折半查找判定树的生成过程

对于折半查找判定树，需要补充以下两点：

- (1) 折半查找判定树是一棵二叉排序树，即每个结点的值均大于其左子树上所有结点的值，小于其右子树上所有结点的值；
- (2) 折半查找判定树中的结点都是查找成功的情况，将每个结点的空指针指向一个实际上并不存在的结点——称为外结点，所有外结点即是查找不成功的情况，如图7-2(e)所示。如果有序表的长度为 n ，则外结点一定有 $n+1$ 个。在折半查找判定树中，某结点所在的层数即是查找该结点的比较次数，整个判定树代表的有序表的平均查找长度即为查找每个结点的比较次数之和除以有序表的长度。例如，长度为10的有序表的平均查找长度为：

$$ASL = (1 \times 1 + 2 \times 2 + 3 \times 4 + 4 \times 3) / 10 = 29 / 10$$

在折半查找判定树中，查找不成功时的比较次数即是查找相应外结点时与内结点的比较次数。整个判定树代表的有序表在查找失败时的平均查找长度即为查找每个外结点的比较次数之和除以外结点的个数。例如，长度为10的有序表在查找失败时的平均查找长度为：

$$ASL = (3 \times 5 + 4 \times 6) / 11 = 39 / 11$$