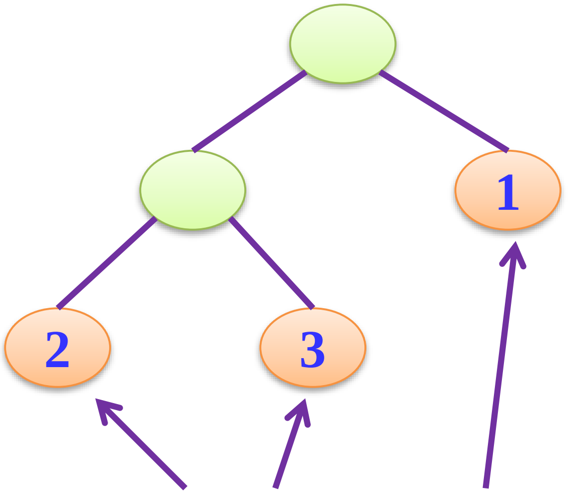
7.9 哈夫曼树

7.9.1 哈夫曼树的定义

设二叉树具有n个带权值的叶节点，那么从根节点到各个叶节点的路径长度与相应节点权值的乘积的和，叫做二叉树的**带权路径长度**。

$$WPL = \sum_{i=1}^{n} w_i l_i$$

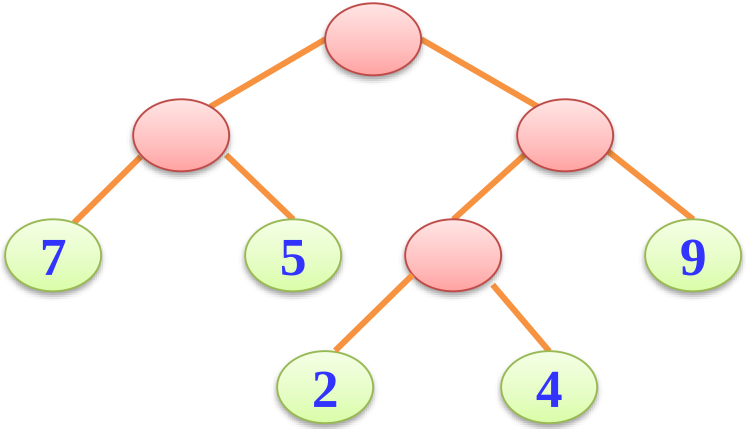
WPL的计算：

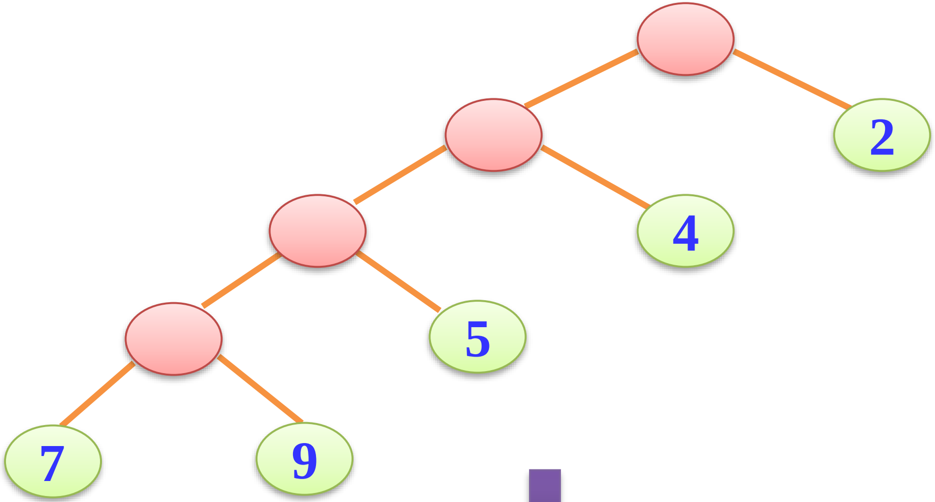


$$WPL = (2+3)\times 2 + 1\times 1 = 11$$

具有最小带权路径长度的二叉树称为**哈夫曼树（也称最优树）**。

相同的叶节点构造出不同的二叉树





WPL(T1) =

7×2+5×2+2×3+4×3+9×2 =60

WPL(T2) =

7×4+9×4+5×3+4×2+2×1=89

7.9.2 构造哈夫曼树

构造哈夫曼树的原则：

- 权值越大的叶节点越靠近根节点。
- 权值越小的叶节点越远离根节点。

构造哈夫曼树的过程：

（1）给定的n个权值$\{W_1, W_2, \ldots, W_n\}$构造$n$棵只有一个叶节点的二叉树，从而得到一个二叉树的集合$F=\{T_1, T_2, \ldots, T_n\}$。

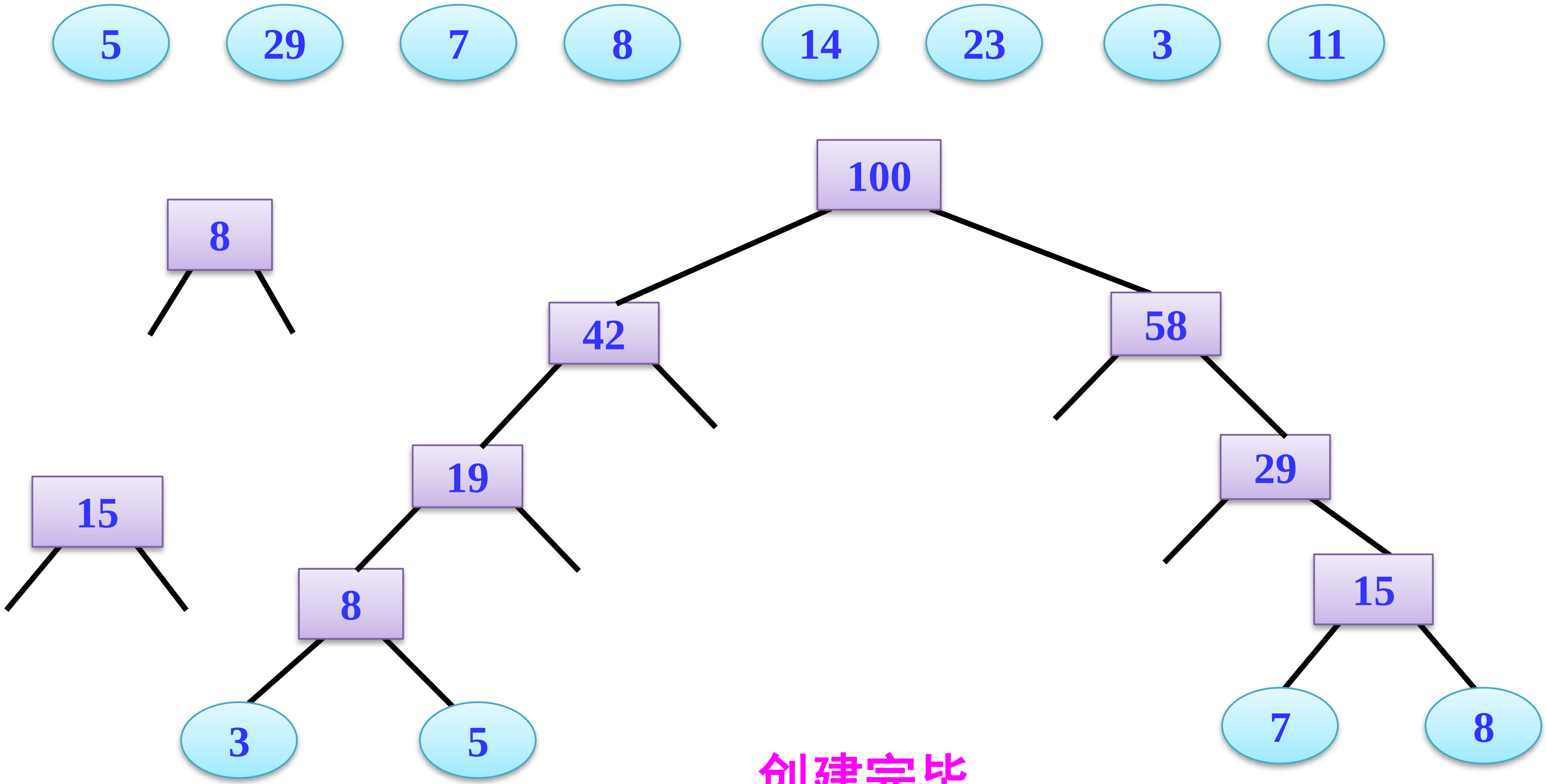
（2）在F中选取根节点的权值最小和次小的两棵二叉树作为左、右子树构造一棵新的二叉树，这棵新的二叉树根节点的权值为其左、右子树根节点权值之和。

（3）在集合F中删除作为左、右子树的两棵二叉树，并将新建立的二叉树加入到集合F中。

（4）重复（2）、（3）两步，当F中只剩下一棵二叉树时，这棵二叉树便是所要建立的哈夫曼树。

建立哈夫曼树示例的演示

W={ 0.05, 0.29, 0.07, 0.08, 0.14, 0.23, 0.03, 0.11}



创建完毕

哈夫曼树的特点：$n_1=0$

所以：$n = n_0+n_1+n_2$

$= n_0+n_2$

$= 2n_0-1$

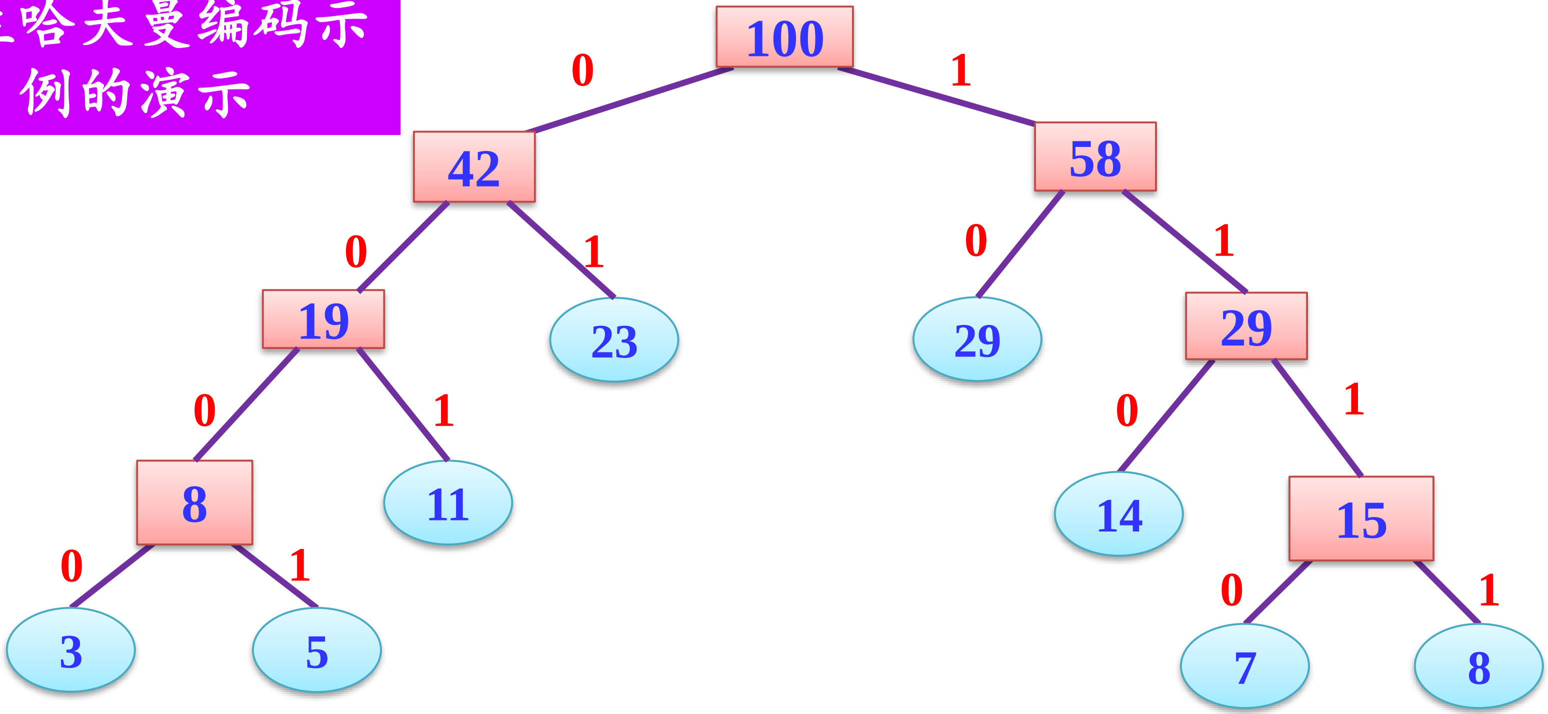
7.9.3 哈夫曼编码

规定哈夫曼树中的左分支为0，右分支为1，则从根节点到每个叶节点所经过的分支对应的0和1组成的序列便为该节点对应字符的编码。这样的编码称为哈夫曼编码。

哈夫曼编码属0、1二进制编码

哈夫曼编码特点：权值越大的字符编码越短，反之越长。

产生哈夫曼编码示例的演示



5 :　0　0　0　1

3:0000	5:0001	11:001	7:1000
8:1111	23:01	29:10	14:110

在一组字符的哈夫曼编码中，不可能出现一个字符的哈夫曼编码是另一个字符哈夫曼编码的前缀。

例如，有4个字符的编码如下：

100，001，0，1

这是哈夫曼编码吗？

×

↓

哈夫曼编码也称为前缀编码。

【例7-13】5个字符有如下4种编码方案，不是前缀编码的是________。

A．01,0000,0001,001,1

B．011,000,001,010,1

C．000,001,010,011,100

D．0,100,110,1110,1100

说明：本题为2014年全国考研题

【例7-14】 对n（$n \geq 2$）个权值均不同的字符构成哈夫曼树，关于该树的叙述中，错误的是______。

A. 该树一定是一棵完全二叉树

B. 该树中一定没有度为1的节点

C. 树中两个权值最小的节点一定是兄弟节点

D. 树中任一非叶子节点的权值一定不小于下一层任一节点的权值

说明：本题为2010年全国考研题

思考题：

哈夫曼编码用什么用途？

——本章完——