

### 5.7.3 哈夫曼编码

利用哈夫曼树可以构造一种不等长的二进制编码，并且构造所得的哈夫曼编码是一种最优前缀编码，即所传电文的总长度最短。

**前缀编码：**任何一个字符的编码都不是同一字符集中另一个字符的编码的前缀。

Letter code  
A: 0  
T: 10  
C: 110  
S: 111

| Letter | Freq | Code | Bits |
|--------|------|------|------|
| C      | 6    | 00   | 2    |
| K      | 7    | 01   | 2    |
| E      | 9    | 10   | 2    |
| A      | 5    | 110  | 3    |
| V      | 2    | 111  | 3    |

| Letter | code   |
|--------|--------|
| C:     | 1110   |
| D:     | 101    |
| E:     | 0      |
| M:     | 11111  |
| K:     | 111101 |
| L:     | 110    |
| U:     | 100    |
| Z:     | 111100 |

128

### 哈夫曼编码(2)

----步骤

给定待编码字符集，哈夫曼编码步骤如下：

1. 将各个字符的频度作为权值，构造哈夫曼树；
2. 给每个内部结点左右两条边赋0和1的权值；
3. 将每个叶结点从根开始路径上的权值按顺序排列，即得该叶结点对应字符的编码码字。

129

### 哈夫曼编码(3)---example1

Huffman编码(不等长)

| Letter | Freq | Code | Bits |
|--------|------|------|------|
| C      | 6    | 00   | 2    |
| K      | 7    | 01   | 2    |
| E      | 9    | 11   | 2    |
| A      | 5    | 101  | 3    |
| V      | 2    | 100  | 3    |

字符平均编码长度=  $65/29=2.24$

若有字符串KEAC，其huffman编码结果：011101100  
等长编码结果：001010100000

等长编码

| Letter | Code | Bits |
|--------|------|------|
| C      | 000  | 3    |
| K      | 001  | 3    |
| E      | 010  | 3    |
| A      | 100  | 3    |
| V      | 101  | 3    |

字符平均编码长度= 3

压缩比CR=1.33

130

### 哈夫曼编码(4) ---example2

| Letter | Freq | Code | Bits |
|--------|------|------|------|
| C      | 32   |      |      |
| D      | 42   |      |      |
| E      | 120  |      |      |
| M      | 24   |      |      |
| K      | 7    |      |      |
| L      | 42   |      |      |
| U      | 37   |      |      |
| Z      | 2    |      |      |

该树的加权路径长度 = 785

Huffman编码：字符平均编码长度=  $785/306 = 2.565$

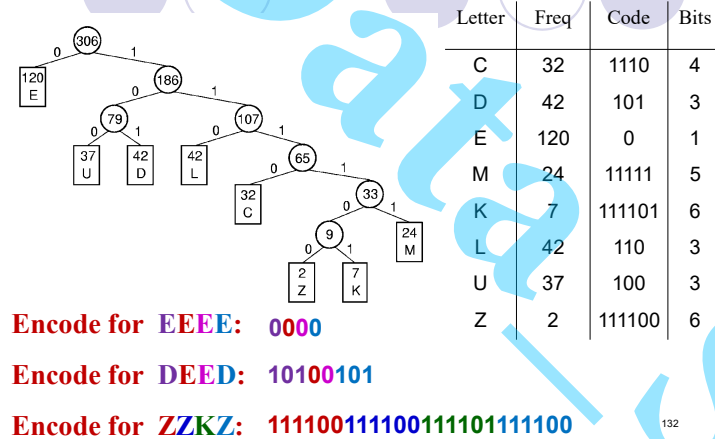
等长编码：字符平均编码长度 =  $\log(8)=3$

CR =  $3/2.565=1.17$

131

131

### 哈夫曼编码(5) ---example2



132

### 哈夫曼编码树

若各叶子权值相同，哈夫曼编码码字的bits是否相同呢

字符集合: {A, B, C, D, E}

权值W: {1, 1, 1, 1, 1}

字符集合: {A, B, C, D, E, F, H, I}

权值W: {1, 1, 1, 1, 1, 1, 1, 1}

若各叶节点权值相同且叶结点个数为2的整数次方，则对应的Huffman Tree是一个叶子节点均在最底层完全满二叉树，此时编码结果等同于等长编码

否则呢? Huffman编码比等长编码有优势吗?

133

### 本章我们学到了

- 树的基本概念和定义
- 二叉树的定义和性质
- 二叉树的遍历，前序，中序，后序，广度优先
- 二叉树结点的执行：基于指针、基于数组
- 二叉搜索树(BST)：定义，基本操作
- 堆(heap)/优先队列：定义，基本操作
- huffman编码树：定义，构建，编码

134

### 本章作业三

● 5.28

135