



www.enjoylinux.cn

第三十八课 霍夫曼树



版权声明：本课件及其印刷物、视频的版权归成都国嵌信息技术有限公司所有，并保留所有权力：任何单位或个人未经成都国嵌信息技术有限公司书面授权，不得使用该课件及其印刷物、视频从事商业、教学活动。已经取得书面授权的，应在授权范围内使用，并注明“来源：国嵌”。违反上述声明者，我们将追究其法律责任。

历史上的远距通信问题



老师：最初的远距通信用于传递文本信息，主要是电报。

小A：那如何将一段文字内容为“BADCADFEED”通过网络传递给别人呢？

小B：利用二进制对这些字母进行编码，然后传输这个编码就行啊。

小D：是的，电报不就“滴”和“答”嘛？！

小A：那怎么编码呢？

小C：难道你没听说过ASCII码？

小A：那时有没有ASCII码哦？

小C：。。。

嵌入式Linux技术咨询QQ号：550491596
嵌入式Linux学习交流QQ群：65212116



最初的解决方案

- ❖ 对于文本“BADCADFEED”的传输而言，因为重复出现的只有“ABCDEF”这6个字符，因此可以用下面的方式编码：

A	B	C	D	E	F
000	001	010	011	100	101

BADCADFEED →

001000011010000011101100100011

接收方可以根据每3个bit进行一次
字符解码的方式还原文本信息。

嵌入式Linux技术咨询QQ号: 550491596
嵌入式Linux学习交流QQ群: 65212116





www.enjoylinux.cn

存在的问题

- ❖ 这样的编码方式需要30个bit位才能表示10个字符
- ❖ 那么当传输一篇500个字符的情报时，需要15000个bit位

在战争年代，这种编码方式对于情报的发送和接受是很低效且容易出错的。

如何提高收发效率？



嵌入式Linux技术咨询QQ号: 550491596
嵌入式Linux学习交流QQ群: 65212116



另一种编码方式

- ❖ 要提高效率，必然要从编码方式的改进入手，要避免每个字符都占用相同的bit位

A	B	C	D	E	F
01	1001	101	00	11	1000

BADCADFEED



1001010010101001000111100

嵌入式Linux技术咨询QQ号: 550491596
嵌入式Linux学习交流QQ群: 65212116





www.enjoylinux.cn

效率提高了吗？

- ❖ 改进的编码方式只需要25个bit位就能表示10个字符
- ❖ 随着传输字符的增加，这种优势会更明显
- ❖ . . .

效率上得到**17%**的提高!!!

问题：如何解码？



嵌入式Linux技术咨询QQ号：550491596
嵌入式Linux学习交流QQ群：65212116



精妙之处



www.enjoylinux.cn

A	B	C	D	E	F
01	1001	101	00	11	1000

BADCADFEED



1001010010101001000111100

任一字符的编码都不是另一个字符编码的前缀！

嵌入式Linux技术咨询QQ号: 550491596

嵌入式Linux学习交流QQ群: 65212116



精妙之处



问题:

这个编码方式是怎么得到的?



嵌入式Linux技术咨询QQ号: 550491596
嵌入式Linux学习交流QQ群: 65212116



霍夫曼树



www.enjoylinux.cn

1. 给定 n 个数值 $\{ v_1, v_2, \dots, v_n \}$
2. 根据这 n 个数值构造二叉树集合 F

$$F = \{ T_1, T_2, \dots, T_n \}$$

T_i 的数据域为 v_i , 左右子树为空

3. 在 F 中选取两棵根结点的值最小的树作为左右子树构造一棵新的二叉树, 这棵二叉树的根结点中的值为左右子树根结点中的值之和
4. 在 F 中删除这两棵子树, 并将构造的新二叉树加入 F 中
5. 重复3和4, 直到 F 中只剩下一个树为止。这棵树即霍夫曼树。

嵌入式Linux技术咨询QQ号: 550491596
嵌入式Linux学习交流QQ群: 65212116

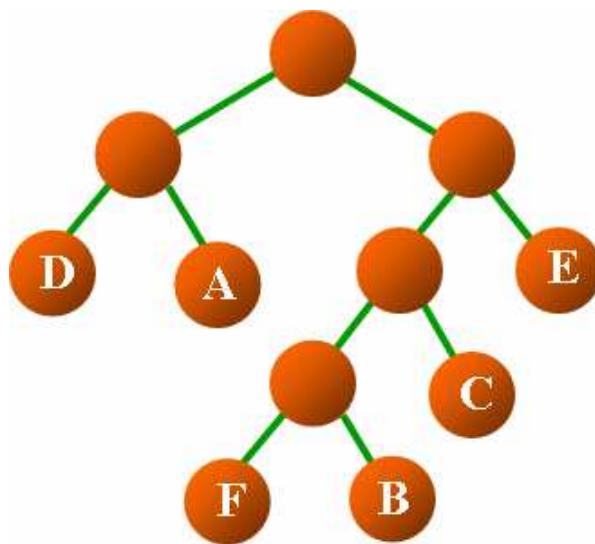


霍夫曼树的应用

❖ 假设经过统计ABCDEF在需要传输的报文中出现的概率如下

A	B	C	D	E	F
27%	8%	15%	15%	30%	5%

❖ 根据这些数值构造的霍夫曼树如下



嵌入式Linux技术咨询QQ号: 550491596
嵌入式Linux学习交流QQ群: 65212116



手把手教你写代码



构建霍夫曼树

嵌入式Linux技术咨询QQ号: 550491596
嵌入式Linux学习交流QQ群: 65212116



霍夫曼树的应用



www.enjoylinux.cn

❖ 该霍夫曼树叶结点在树中的位置编码如下

A: 01

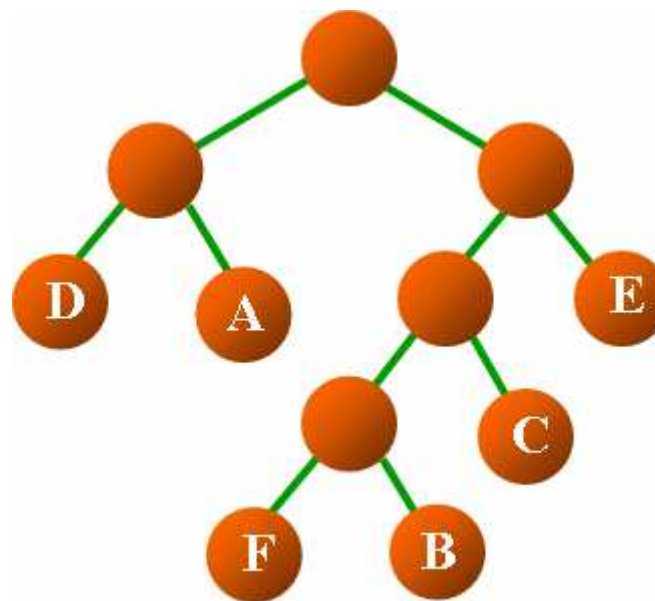
B: 1001

C: 101

D: 00

E: 11

F: 1000



嵌入式Linux技术咨询QQ号: 550491596

嵌入式Linux学习交流QQ群: 65212116

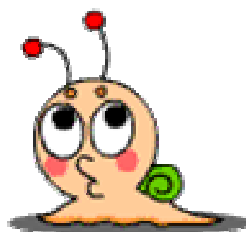


小结



www.enjoylinux.cn

- ❖ 霍夫曼树是一种特殊的二叉树
- ❖ 霍夫曼树应用于信息编码和数据压缩领域
- ❖ 霍夫曼树是现代压缩算法的基础



嵌入式Linux技术咨询QQ号: 550491596
嵌入式Linux学习交流QQ群: 65212116

