

## 哈夫曼编码

计算机传输数据时，必须将信息的内容编码成0或1的信息流，比如说可以讲一个字母或者数字转换成ASCII码，成为8位的0/1串，但是这么编码生成出的0/1信息流还是比较长。可以将一些出现频数较高的字母缩短编码长度，而频度较低的字母加长编码长度，以达到缩短总长度的目的。

假设信息只由A,B,C,D,E这几个字母组成，其出现的次数分别是A: 5, B: 10, C: 13, D: 14, E: 20。请设计一种0/1编码，使编码后的总长度最小。

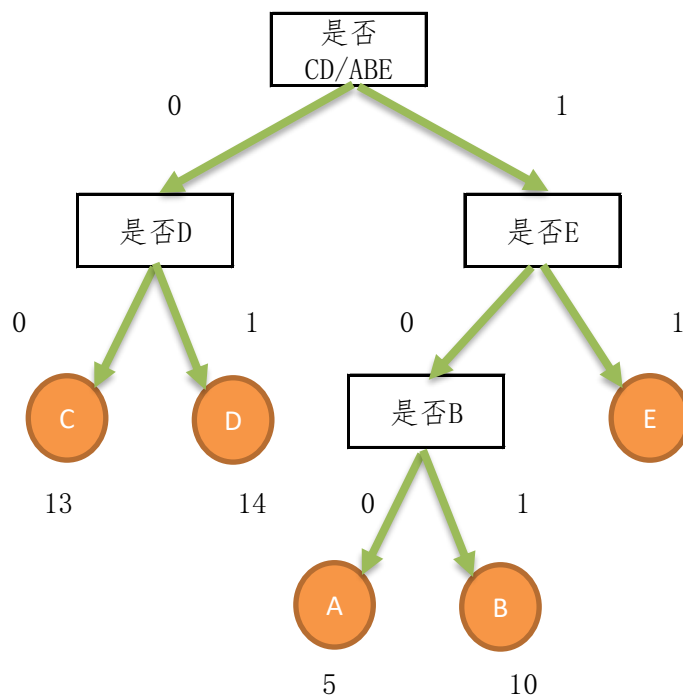
### 分析

既然要将每个字母变成一组0/1编码，出现次数多的字母要短，那可以从一开始给这些使用频数最多的字母编码，如果一位不够就两位……可以得到这样的编码方案。

E: 0 D: 1 C: 00 B: 01 A: 10

然而，这么做是不行的。如果收到了001，对应的原文可能是EED、CD或者EB，这就造成了歧义。为了不造成歧义，我们要求一个字母的0/1编码不是其他编码的前缀。（比如01就是011的前缀）

那么该如何正确的构建呢？我们使用如下方法。



要查询某个字母的编码就相当于从头开始将字母分类，往左边分就是0，往右边分就是1，直到不可分为止。因此，最终的编码如下：

C: 00 D: 01 A: 100 B: 101 E: 11

因此“BAD”可以编码成10110001，而编码“001101”可以解码成“CED”。这样的编码方式就被称为哈夫曼编码。

逻辑航线培优教育，信息学奥赛培训专家。

扫码添加作者获取更多内容。

