

問題二 (30%)

- 霍夫曼編碼：是一種用於檔案壓縮的方法。
- 請同學實作檔案壓縮的功能，將輸入的文字資料編碼後所需bit數(總長度)，印在螢幕上。
- 輸入資料為多行不等長的字串，使用的字元包括空格以及：
- !, - . : ; ? 0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K
L M N O P Q R S T U V W X Y Z a b c d e f g h i j
k l m n o p q r s t u v w x y z

編碼方式(*Huffman*)

字元	頻率
H	1
u	1
f	2
m	1
a	1
n	3

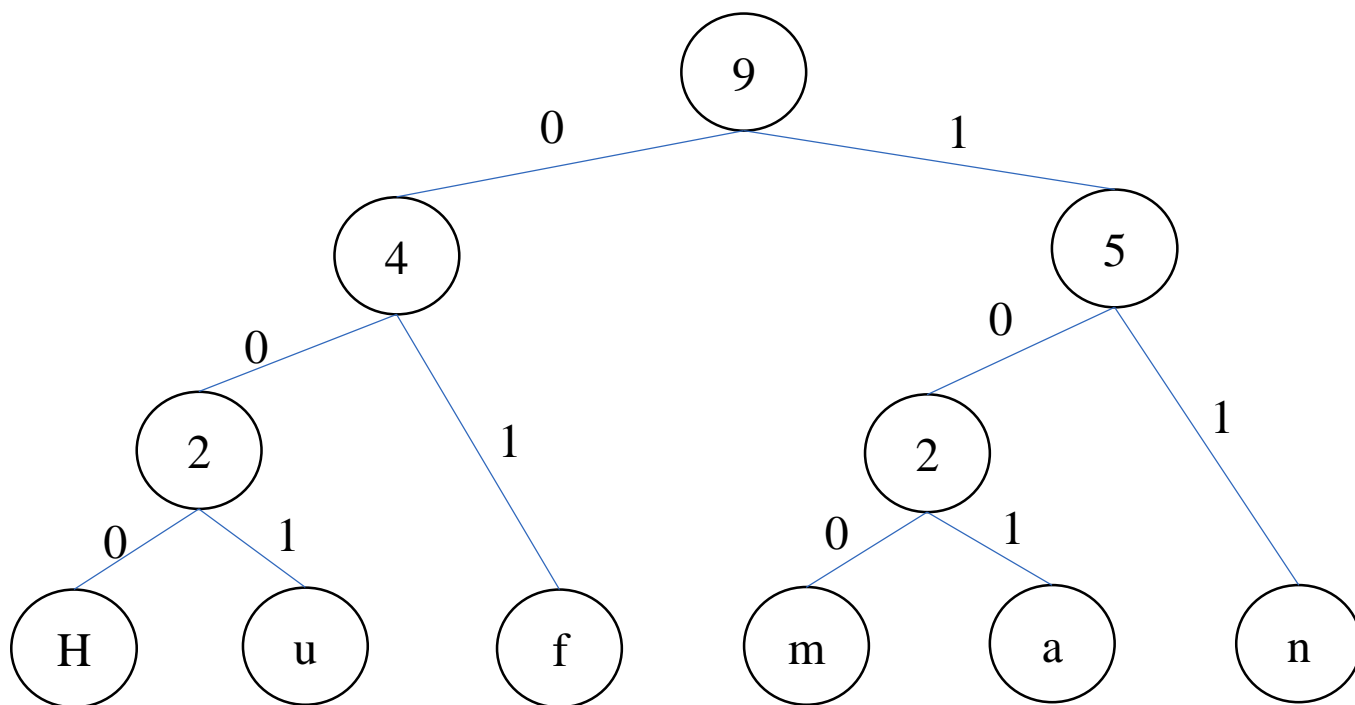
- 範例輸入:Huffmannn

1. 先統計字元出現的頻率/次數。
2. 將字串中的所有字元放在最底層的葉節點上。
3. 目前尚無父節點的所有節點中，尋找值(出現次數)最小的兩個進行合併，合併後的父節點值等於兩個子節點的值相加。
4. 合併直到剩下一個根節點為止。
5. 從根節點開始向下編碼，左分支為0、右分支為1，每個葉節點的字元會產生一個霍夫曼碼。

編碼方式(*Huffman*)

- 範例輸入:Huffmannn

字元	頻率	編碼
H	1	000
u	1	001
f	2	01
m	1	100
a	1	101
n	3	11



編碼方式(*Huffman*)

字元	頻率	編碼
H	1	000
u	1	001
f	2	01
m	1	100
a	1	101
n	3	11

- 範例輸入:Huffmannn
- 一個Char的大小為1 byte = 8 bits
- 原本”Huffmannn”這個字串要用 $9 * 8 = 72$ bits儲存
- 編碼後總長度 = $1 * 3 + 1 * 3 + 2 * 2 + 1 * 3 + 1 * 3 + 3 * 2$
 $= 3 + 3 + 4 + 3 + 3 + 6$
 $= 22$ bits

輸入/輸出

- 由input_2.txt讀取測資。
- input_2.txt的第一行包含一個整數 m ，表示有多少行的字串，($1 \leq m \leq 100$ ，字串長1~1000)。
- 可能有多筆測資計算，若 m 讀取到0則表示測資結束。
- 最後在螢幕上印出輸入字串經過編碼後的總長度。

範例

(input)

2

Hello!

oH He lolo

0

(output)

40

字元	出現頻率	編碼
l	4	10
o	4	11
H	3	01
e	2	001
_ (space)	2	0001
!	1	0000

$$\begin{aligned}\text{總長度} &= 4*2 + 4*2 + 3*2 + 2*3 + 2*4 + 1*4 \\ &= 8 + 8 + 6 + 6 + 8 + 4 \\ &= 40\end{aligned}$$