

## Codis de Huffman

P62266\_ca

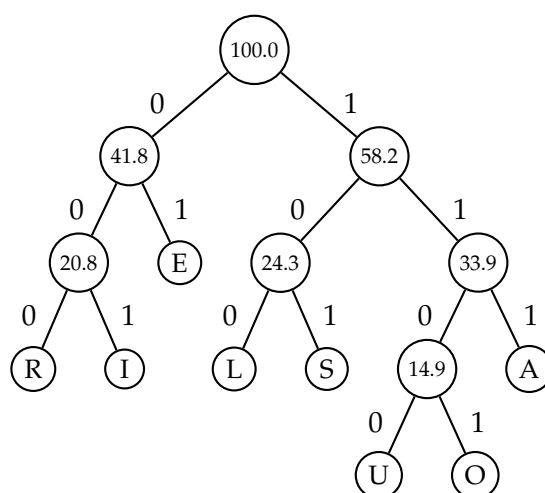
Suposeu que en l'escriptura d'un determinat llenguatge s'utilitzen  $n$  símbols diferents. Una manera simple de codificar-hi un text consisteix a assignar  $\lceil \log_2 n \rceil$  bits a cada símbol. Per exemple, considereu les cinc vocals i les tres consonants més freqüents en llengua catalana. Segons la taula següent, una codificació "normal" de la paraula RES seria 101000010:

Lletra	Codificació "normal"	Freqüència relativa	Codificació de Huffman
E	000	21	01
A	001	19	111
S	010	12.7	101
L	011	11.6	100
I	100	10.6	001
R	101	10.2	000
O	110	8.6	1101
U	111	6.3	1100

Suposeu ara que coneixem les probabilitats (o freqüències relatives) de cadascun dels  $n$  símbols (per exemple, segons la taula, de cada 100 símbols entre els vuit escollits, 21 són es, 19 són as, etcètera). Es pot aconseguir una codificació més eficient en mitjana, si a cada lletra se li assignen menys bits com més freqüent és. Segons la taula, la codificació de Huffman de la paraula RES seria 00001101, amb només vuit bits en lloc de nou.

La construcció d'una codificació de Huffman és relativament senzilla: repetidament, s'escullen els dos símbols menys freqüents, se'ls assigna arbitràriament un bit (0 o 1) a cadascun per diferenciar-los, i a partir d'aquell moment se'ls considera un únic símbol. L'algorisme acaba quan només queda un símbol.

Per exemple, aquest és l'arbre corresponent a l'algorisme de Huffman per als vuit caràcters de la taula anterior. La quantitat dins de cada node és la freqüència relativa de tots els símbols que en penjen.



En aquest exemple, la longitud mitjana de la codificació de Huffman d'un símbol és només

$$0.21 * 2 + 0.19 * 3 + 0.127 * 3 + 0.116 * 3 + 0.106 * 3 + 0.102 * 3 + 0.086 * 4 + 0.063 * 4 \simeq 2.9390,$$

més petita que la longitud mitjana d'una codificació "normal", la qual és evidentment 3. Per a distribucions de probabilitats més esbiaxades, s'aconsegueixen estalvis molt més significatius.

Feu un programa que llegeixi les freqüències relatives d'unes quantes lletres, i en calculi la longitud mitjana de la seva codificació de Huffman.

### **Entrada**

L'entrada consisteix en el nombre de símbols  $n \geq 2$ , seguit de les freqüències relatives dels  $n$  símbols. Aquestes freqüències són totes no negatives, i sumen 100.

### **Sortida**

Escriviu amb quatre decimals el nombre esperat de bits per lletra.

#### **Exemple d'entrada 1**

```
8
19 21 10.6 8.6 6.3 12.7 11.6 10.2
```

#### **Exemple de sortida 1**

```
nombre esperat de bits per lletra: 2.9390
```

#### **Exemple d'entrada 2**

```
4
5 2 3 90
```

#### **Exemple de sortida 2**

```
nombre esperat de bits per lletra: 1.1500
```

### **Informació del problema**

Autor : Salvador Roura

Generació : 2024-05-02 20:44:56

© *Jutge.org*, 2006–2024.

<https://jutge.org>