

Multimédia et interaction humain-machine 2005-6

ED texte

Question 1

Ecrire un programme perl qui calcule les fréquences d'apparitions de caractères d'un fichier texte et les affiche sous forme décroissante :

```
FREQUENCY OF ' ' IS 115049
FREQUENCY OF 'e' IS 76701
FREQUENCY OF 'a' IS 45149
FREQUENCY OF 's' IS 43120
FREQUENCY OF 't' IS 39070
FREQUENCY OF 'i' IS 38524
FREQUENCY OF 'n' IS 35235
FREQUENCY OF 'r' IS 34459
FREQUENCY OF 'l' IS 33785
FREQUENCY OF 'u' IS 33042
FREQUENCY OF 'o' IS 27526
FREQUENCY OF 'd' IS 19275
FREQUENCY OF 'c' IS 14857
FREQUENCY OF 'm' IS 14696
FREQUENCY OF 'p' IS 13857
FREQUENCY OF ',' IS 12417
```

Question 2

Calculer les codes d'Huffman pour le texte DBACEFADACBFDEACACDEBFACBD
Quel est le taux de compression obtenu ?

Question 3

décrypter le texte suivant :

```
qvfpbhef qr yn zrgubqr
cbhe ovra pbaqhver fn envfba,
rg purepure yn irevgr qnaf yrf fpvrprf
```

corrigé - question 1

```
#!/usr/bin/perl
@x = <>;
foreach $l (@x) { $l =~ s/(.)/$chars{$l}++;$l/eg; }
print "FREQUENCY OF '$_' IS $chars{$_}\n"
    foreach (sort {$chars{$b} <=> $chars{$a}} keys %chars);
```

corrigé - question 2

Table des fréquences correspondant au texte :

Caractère	Code	Fréquence
A	65	6
B	66	4
C	67	5
D	68	5
E	69	3
F	70	3

On commence par chercher les car. de plus petite fréquence.

Attention ! Il sera impératif de respecter la convention suivante pour déterminer sans ambiguïté les deux plus petites fréquences : une fréquence f_1 sera considérée inférieure à une fréquence f_2 dans les deux cas suivants :

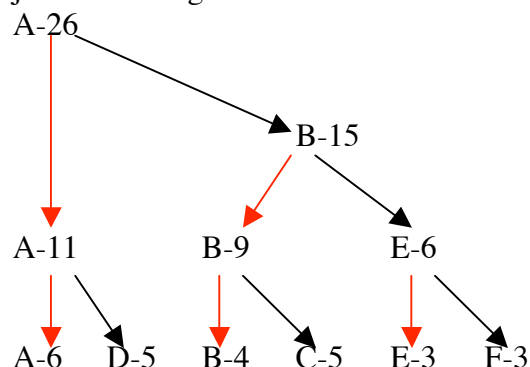
$f_1 < f_2$

ou bien

$f_1 = f_2$ et le code associé à f_1 est inférieur à celui associé à f_2

A chaque étape de construction de l'arbre, le nouveau noeud créé porte comme information la somme des fréquences de ses deux fils ainsi que le plus petit des codes de ses deux fils.

je note en rouge les arcs de bit 0



A	00
B	100
C	101
D	01
E	110
F	111

corrigé question 3

frequency of ' ' is 13
frequency of 'r' is 13
frequency of 'f' is 8
frequency of 'e' is 7
frequency of 'p' is 6
frequency of 'v' is 6
frequency of 'b' is 5
frequency of 'q' is 5
frequency of 'a' is 5
frequency of 'n' is 5
frequency of 'y' is 3
frequency of 'h' is 3
frequency of 'g' is 3
frequency of 'u' is 3
frequency of ',' is 1
frequency of 'c' is 1
frequency of 'o' is 1
frequency of 'i' is 1
frequency of 'z' is 1

si on fait $r \Rightarrow E$

qvfpbhef qE yn zEgubqE
cbhe ovEa pbaqhveE fn envfba,
Eg puEepuEe yn iEevgE qnaf yEf fpvEapEf

on essaie $f \Rightarrow A$

qvApbheA qE yn zEgubqE
cbhe ovEa pbaqhveE An envAba,
Eg puEepuEe yn iEevgE qnaA yEA ApvEapEA
pas bon : des mots terminent par A

on re-essaie avec $f \Rightarrow S$

qvSpbheS qE yn zEgubqE
cbhe ovEa pbaqhveE Sn envSba,
Eg puEepuEe yn iEevgE qnaS yES SpvEapES
pas mal : en particulier yES qui peut être LES, MES, DES, CES
et qE qui peut être LE ou DE

etc...

avec beaucoup de courage, on arrive à :

discours de la methode
pour bien conduire sa raison,
et chercher la verite dans les sciences

la méthode employée est le rot13, logiciel bien connu des groupes de News