

<p style="text-align: center;">info223 : Science informatique TD 3 : codage par bloc</p>

Pierre Hyvernat
Laboratoire de mathématiques de l'université de Savoie
bâtiment Chablais, bureau 22, poste : 94 22
email : Pierre.Hyvernat@univ-savoie.fr
www : <http://www.lama.univ-savoie.fr/~hyvernat/>

Question 1. L'ADN est formé de 4 bases : l'adénine, la thymine, la cytosine et la guanine. Quel codage taille fixe peut-on utiliser ?

Si les fréquences des 4 bases sont identiques, que se passe-t'il si on décide de coder une des bases sur 1 bit (dans le cadre d'un codage par bloc de taille variable) ?

Donnez le codage complet ainsi que le gain ou la perte de taille si chaque base est présente n fois dans une suite.

Question 2. Si l'on veut stocker des suites des caractères n'utilisant que les voyelles, en proportions identiques, quel codage par bloc peut on utiliser pour optimiser l'espace occupé ? Justifiez que votre solution est optimal en utilisant l'inégalité de Kraft.

Question 3.

- Peut-on construire un code non-ambigu avec des mots de longueur 1, 2, 3, 3, 3 ?
- Peut-on construire un code non-ambigu avec des mots de longueur 1, 3, 3, 3, 3 ?
- Peut-on construire un code non-ambigu avec des mots de longueur 1, 3, 3, 3, 3, 3 ?

Quant cela est possible, construisez un tels code. Quand cela est impossible, justifiez.

Question 4. En utilisant l'algorithme de Sardinas-Patterson, décidez si les codes suivants sont ambigus ou non :

- $\{\underline{010101}, \underline{0101}\}$,
- $\{\underline{1}, \underline{010}, \underline{101}, \underline{001}\}$,
- $\{\underline{1}, \underline{010}, \underline{100}, \underline{000}\}$,
- $\{\underline{00}, \underline{01}, \underline{10}, \underline{11}, \underline{100010100}\}$,
- $\{\underline{1}, \underline{011}, \underline{01110}, \underline{1110}, \underline{1001}\}$,
- $\{\underline{00}, \underline{10}, \underline{11}, \underline{100}, \underline{110}\}$.

Lorsque le code est ambigu, donnez une suite de bit qui a au moins deux décodages possibles.

Question 5. Transformez le code $\{\underline{1}, \underline{010}, \underline{100}, \underline{000}\}$ en code préfixe de même taille. Qu'en pensez-vous ?

Question 6. Supposons qu'on ne dispose que du début d'un mot, peut-on décoder le (début du) mot avec les codes non-ambigus suivants :

- $\{\underline{1}, \underline{0}\}$,
- $\{\underline{01}, \underline{10}\}$,
- $\{\underline{0110}, \underline{0100}\}$,
- $\{\underline{0}, \underline{01}, \underline{011}\}$,
- $\{\underline{1}, \underline{10}, \underline{00}\}$.

Question 7. En reprenant les codes de la question précédente, que se passe-t'il si on ne dispose que de la fin d'un mot ?

Idem s'il manque le début et la fin d'un mot : peut-on encore décoder la partie du milieu ?