

**info223 : Science informatique**  
**TD 3 : codage par bloc**

Pierre Hyvernât  
Laboratoire de mathématiques de l'université de Savoie  
bâtiment Chablais, bureau 22, poste : 94 22  
email : [Pierre.Hyvernât@univ-savoie.fr](mailto:Pierre.Hyvernât@univ-savoie.fr)  
www : <http://www.lama.univ-savoie.fr/~hyvernât/>

*Question 1.* L'ADN est formé de 4 bases : l'adénine, la thymine, la cytosine et la guanine. Plutôt que le codage ASCII sur un octet (adénine : A, thymine : T, cytosine : C et guanine : G), proposez un codage de taille fixe qui gagne de la place.

On peut utiliser un codage de taille variable en codant une des bases sur un seul bit. Sur combien de bits faut-il coder les autres bases ?

Si les fréquences des 4 bases sont identiques (25% chacune) quelle est la taille de la suite codée avec le codage de taille variable? A t'on un gain ou une perte de taille par rapport au codage de taille fixe ?

*Question 2.* Si l'on veut stocker des suites de A, E, I, O, U et Y (les voyelles), qui apparaissent avec des fréquences identiques (16.66% chacune). Quel codage par bloc peut on utiliser pour optimiser l'espace occupé ?

Justifiez que votre solution est optimale 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 :

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

Lorsque le code est ambigu, donnez une suite de bits 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 :

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

*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 ?