

Codage de l'information

Devoir surveillé n° 1

Vendredi 24 octobre 2014 - Durée 2h - Documents non autorisés. Calculatrices autorisées.

Veuillez indiquer le numéro de votre groupe de TD sur la copie qu'il est inutile de rendre anonyme, ainsi que votre NIP (figurant sur votre carte d'étudiant).

Ce sujet contient quatre exercices indépendants. Prenez 10mn pour lire l'intégralité du sujet avant de commencer.

Exercice 1-1 *Représentation des caractères et des nombres*

Voici une vue hexadécimale d'un fichier codé en UTF-8 :

00000000 43c8 9f62 69d0 9248 0a

Question 1 De combien d'octets est composé ce fichier ?

Question 2 De combien de caractères est composé ce fichier (on ne demande pas de donner le code de chacun des caractères) ?

Prenons les 10 bits de poids faible des deux premiers octets du fichier précédent (c'est-à-dire les deux bits de poids faible du premier octet suivis du deuxième octet).

Question 3 Quel est l'entier correspondant ?

Question 4 En supposant que ces 10 bits sont la représentation d'un entier en complémentaire à deux, quelle est la valeur stockée ?

Question 5 En supposant que ces 10 bits sont la mantisse d'un nombre réel en base 2, quelle est la valeur de cette mantisse ?

Exercice 1-2 *Langages et codes*

Soit $L_1 = \{\varepsilon, 0, 10\}$ et $L_2 = \{110, 111\}$.

Question 1 Quel est le résultat de $L_1 \cdot L_2$? L'ensemble résultant est-il un code ? Pourquoi ?

Question 2 Mêmes questions pour $L_1 = \{0, 1\}$ et $L_2 = \{1, 10, 011\}$.

Question 3 Mêmes questions pour $L_1 = \{0, 01\}$ et $L_2 = \{1, 111\}$.

Exercice 1-3 *Construction de codes*

On désire définir des codes composés de sept mots définis sur un alphabet $\mathcal{A} = \{0, 1, 2\}$.

Question 1 En suivant ces contraintes, donnez un code à virgule, avec la virgule en préfixe (à gauche).

Question 2 En suivant ces contraintes, donnez un code préfixe avec exactement quatre mots de longueur 2.

Exercice 1-4 *Représentation de l'ADN*

L'ADN, support de l'information génétique, peut être représenté avec quatre lettres A, C, G et T, appelées nucléotides.

Question 1 Dans le codage ASCII, le nucléotide A est codé par 65 (en décimal). Donnez les représentations hexadécimales et binaires (sur huit bits) de 65. Mêmes questions pour les nucléotides C, G et T (dont les codes dans le codage ASCII sont respectivement 67, 71 et 84).

Question 2 Parfois les données génétiques sont stockées en minuscule ce qui nécessite de les convertir en majuscules afin d'uniformiser les données. Les nucléotides a, c, g et t ont les codes 97, 99, 103 et 116 dans le codage ASCII. Donnez la représentation binaire (sur huit bits) de ces quatre codes.

Question 3 Avec une même formule, utilisant les opérateurs logiques, comment convertir un code n correspondant à un nucléotide en minuscule, vers le code correspondant à un nucléotide en majuscule.

Question 4 On souhaite convertir directement les codes ASCII des quatre nucléotides (A, C, G et T) en un entier sur 2 bits selon le codage \mathbf{c} suivant :

$$\mathbf{c}(65) = \overline{00}_2, \mathbf{c}(67) = \overline{01}_2, \mathbf{c}(71) = \overline{11}_2, \mathbf{c}(84) = \overline{10}_2.$$

Donnez une seule formule utilisant uniquement des opérateurs logiques, permettant de calculer \mathbf{c} .

Question 5 Dans l'ADN les A sont toujours associés à un T et les C toujours à un G (et inversement). On dit que A et T, ainsi que C et G sont complémentaires. Donnez une formule permettant de convertir un nucléotide en son nucléotide complémentaire (tous deux codés avec le codage \mathbf{c}). Voici à quoi correspond le passage d'un entier représentant un nucléotide à l'entier représentant le nucléotide complémentaire :

$$\begin{aligned}\overline{00}_2 &\leftrightarrow \overline{10}_2 \\ \overline{01}_2 &\leftrightarrow \overline{11}_2\end{aligned}$$

On considère la fonction f dont l'algorithme est décrit dans l'algorithme 1.1.

Algorithme 1.1 Algorithme de calcul de la fonction f

Entrée : n un entier naturel

Sortie : $f(n)$.

$r \leftarrow 0$

tant que $n > 0$ **faire**

si $n \wedge 1 > 0$ **alors**

$r \leftarrow r + 1$

fin si

$n \leftarrow n \gg 2$

fin tant que

renvoyer r

On suppose donnée une fonction $convDNA$ convertissant une chaîne de caractères d'ADN en un entier dans lequel chaque nucléotide a été codé avec le codage \mathbf{c} .

Question 6 Quel est le résultat de la fonction $convDNA("ACGACTGAAG")$?

Question 7 Quel est le résultat de la fonction $f(convDNA("ACGACTGAAG"))$? Donner la valeur de n (en binaire) à chaque étape au début de la boucle.

Question 8 À quoi sert la fonction f sur la chaîne d'ADN finale ?