

## Codage de l'information

### Devoir surveillé n° 1

Samedi 13 novembre 2010 - Durée 2h - Documents non autorisés. Calculatrices autorisées.

Veuillez indiquer le numéro de votre groupe de TD sur la copie.

Ce sujet contient trois exercices indépendants.

#### Exercice 1-1 *Codage de nombres*

Voici un mot de 16 bits (écrit en hexadécimal).

0xA674.

**Question 1** Donnez l'écriture binaire de ce mot.

**Question 2** Si on considère ce mot comme la représentation binaire d'un entier positif ou nul, quel est cet entier ?

**Question 3** Si on considère ce mot comme la représentation binaire d'un entier signé codé sur 16 bits en complément à deux, quel est cet entier ?

**Question 4** Quel est le mot binaire  $u$  tel que

$$u \oplus 0xA674 = 0xFFFF ?$$

Répondez en hexadécimal.

**Question 5** Donnez une expression ne dépendant que d'un entier positif  $n$  compris entre 0 et  $2^{16} - 1$ , et ne faisant intervenir que les opérations logiques sur les entiers telle que si l'entier  $n$  s'écrit 0xWXYZ en hexadécimal alors la valeur de l'expression s'écrit 0xZWXY en hexadécimal. Par exemple avec l'entier 0xA674, l'expression vaut 0x4A67.

#### Exercice 1-2 *Codage de textes*

Un fichier contient le texte<sup>1</sup> suivant :

*L'amour des mathématiques conduit les esprits sérieux à s'attaquer à des problèmes reconnus difficiles, dont la résolution nécessite abnégation, abstraction et aptitudes hors du commun.*

Une vue hexadécimale de ce fichier est

00000000	4C 27 61 6D	6F 75 72 20	64 65 73 20	6D 61 74 68	L'amour des math
00000010	C3 A9 6D 61	74 69 71 75	65 73 20 63	6F 6E 64 75	..matiques condu
00000020	69 74 20 6C	65 73 20 65	73 70 72 69	74 73 20 73	it les esprits s
00000030	C3 A9 72 69	65 75 78 20	C3 A0 0A 73	27 61 74 74	..rieux ...s'att
00000040	61 71 75 65	72 20 C3 A0	20 64 65 73	20 70 72 6F	aqer .. des pro
00000050	62 6C C3 A8	6D 65 73 20	72 65 63 6F	6E 6E 75 73	bl..mes reconnus
00000060	20 64 69 66	66 69 63 69	6C 65 73 2C	20 64 6F 6E	difficiles, don
00000070	74 20 6C 61	0A 72 C3 A9	73 6F 6C 75	74 69 6F 6E	t la.r..solution

1. Premier paragraphe de l'article de Jean-Paul Delahaye dans le numéro d'octobre 2010 de la revue Pour la Science.

```

00000080  20 6E C3 A9  63 65 73 73  69 74 65 20  61 62 6E C3  n..cessite abn.
00000090  A9 67 61 74  69 6F 6E 2C  20 61 62 73  74 72 61 63  .gation, abstrac
000000A0  74 69 6F 6E  20 65 74 20  61 70 74 69  74 75 64 65  tion et aptitude
000000B0  73 20 0A 68  6F 72 73 20  64 75 20 63  6F 6D 6D 75  s .hors du commu
000000C0  6E 2E 0A                                     n..

```

**Question 1** Des deux codages UTF-8 et ISO-8859-1 quel est celui utilisé dans le fichier ?

**Question 2** En sachant que dans le codage utilisé dans ce fichier, les lettres (non accentuées) sont codées par des nombres consécutifs, et en sachant que la lettre A majuscule est codée par l'entier 65, donnez en hexadécimal le codage du mot AMOUR.

**Question 3** Quelle serait la taille du fichier (en octets) si le même texte avait été codé dans l'autre codage ?

**Question 4** Si on code le fichier initial dans le codage base 64, quelle est la taille du fichier obtenu ?

**Exercice 1-3** *Étude d'un code*

On considère le langage binaire

$$L = \{01, 00, 10, 011, 111\}.$$

**Question 1** En utilisant l'algorithme de Sardinas et Patterson, vérifiez que  $L$  est un code.

**Question 2** Est-il possible de remplacer dans le code  $L$  l'un des mots par un mot plus court tout en conservant au langage ainsi obtenu la propriété d'être un code ?

On utilise le code  $L$  pour coder des mots sur l'alphabet  $\mathcal{S} = \{C, G, A, T, U\}$ , avec le codage défini par la table

$s$	C	G	A	T	U
$c(s)$	10	011	01	111	00

Les mots utilisés pour ce codage sont tels que le codage d'un mot  $\mathbf{u} \in \mathcal{S}^*$  n'a pas une longueur multiple de 8, et ne peut donc pas être découpé en un nombre entier d'octets. Si on veut stocker le résultat du codage d'un mot dans un fichier, il est donc nécessaire d'utiliser une *complétion*.

La complétion utilisé pour l'écriture dans le fichier consiste à ajouter quelques bits dont le nombre dépend du bit manquant.

1. S'il manque  $1 \leq r \leq 7$  bits au mot codé pour obtenir un nombre entier d'octets, on ajoute  $r$  bits dont le premier est 1 et les  $r - 1$  suivant 0. Autrement dit on ajoute le mot  $\mathbf{10}^{r-1}$ .
2. S'il ne manque aucun bit au mot codé, par homogénéité avec les cas précédents, on ajoute un octet complet dont le bit de poids fort est 1, et les sept autres bits 0. Autrement dit on ajoute le mot  $\mathbf{10}^7$ .

Dans tous les cas, cette complétion est telle que le mot codé s'arrête juste avant le dernier bit égal à 1 dans le fichier.

**Question 3** Un fichier codant un mot de  $\mathcal{S}^*$  contient deux octets : 0x6B suivi de 0x80. Quel est le mot de  $\mathcal{S}^*$  codé dans ce fichier ?

**Question 4** Le dernier bit du fichier précédent est accidentellement changé en son opposé. Peut-on encore décoder le fichier ainsi transformé ? Si oui, quel est le mot de  $\mathcal{S}^*$  qui est codé ?