

Codage de l'information

Codes et codages

1 Exercices

1.1 Première série

Exercice 1-1 Avec le codage ASCII

Pour cet exercice, utilisez la table ASCII de la page 6

Question 1 Codez votre nom en hexadécimal avec le codage ASCII.

Question 2 Voici le début d'un fichier texte vu avec un éditeur hexadécimal.

```
00000000  4C 61 20 43 69 67 61 6C 65 20 65 74 20 6C 61 20
00000010  46 6F 75 72 6D 69 0A 0A 4C 61 20 43 69 67 61 6C
00000020  65 2C 20 61 79 61 6E 74 20 63 68 61 6E 74 65 0A
00000030  54 6F 75 74 20 6C 27 65 74 65 2C 0A 53 65 20 74
```

Quel est ce texte ?

Question 3

Si on change le 3ème bit du second octet du fichier précédent, que devient le fichier texte ?

Exercice 1-2 Avec le codage ISO-8859-1

On rappelle que le codage ISO-8859-1 (encore appelé ISO-LATIN-1), est un codage de 224 caractères tous codés sur un octet (codes 0 à 127, et 160 à 255. Les 128 premiers caractères sont identiques aux caractères du codage ASCII, les 96 suivants contiennent (presque) tous les autres caractères utilisés en français comme par exemple les caractères accentués.

Voici une vue hexadécimale sur le début de la fable de La Fontaine « La cigale et la fourmi » codée en ISO-LATIN-1.

```
00000000  4C 61 20 43 69 67 61 6C 65 20 65 74 20 6C 61 20  La Cigale et la
00000010  46 6F 75 72 6D 69 0A 0A 4C 61 20 43 69 67 61 6C  Fourmi..La Cigal
00000020  65 2C 20 61 79 61 6E 74 20 63 68 61 6E 74 E9 0A  e, ayant chant..
00000030  54 6F 75 74 20 6C 27 E9 74 E9 2C 0A 53 65 20 74  Tout l'.t.,.Se t
```

Quel est le code du caractère é ?

Exercice 1-3 Avec le codage UTF-8

On rappelle que le codage UTF-8 est un codage de longueurs variables, dont les mots du code sont constitués de 1, 2, 3 ou 4 octets. La table ?? (page ??) donne le format des codes selon le nombre d'octets.

Question 1 Combien de caractères le codage UTF-8 permet-il de coder ?

Voici une vue hexadécimale sur le début de la fable de La Fontaine « La cigale et la fourmi » codée en UTF-8.

```

00000000  4C 61 20 43 69 67 61 6C 65 20 65 74 20 6C 61 20 La Cigale et la
00000010  46 6F 75 72 6D 69 0A 0A 4C 61 20 43 69 67 61 6C Fourmi..La Cigal
00000020  65 2C 20 61 79 61 6E 74 20 63 68 61 6E 74 C3 A9 e, ayant chant..
00000030  0A 54 6F 75 74 20 6C 27 C3 A9 74 C3 A9 2C 0A 53 .Tout l'...t...S

```

Question 2 Quel est le code du caractère é ? Ce code a-t-il un rapport avec le code du même caractère dans le codage ISO-LATIN-1 ?

Exercice 1-4 Avec un codage de longueur variable

Dans cet exercice on considère le codage décrit par la table 1.

A 1010	B 0010011	C 01001	D 01110	E 110	F 0111100	G 0111110
H 0010010	I 1000	J 011111110	K 01111111001	L 0001	M 00101	N 1001
O 0000	P 01000	Q 0111101	R 0101	S 1011	T 0110	U 0011
V 001000	W 01111111000	X 01111110	Y 011111111	Z 0111111101	ESP 111	

TABLE 1 – Codage de longueur variable

Question 1 Codez votre nom avec ce codage de longueur variable.

Question 2 Décodez le message 100010010111100000001010010110100110100000001001.

Question 3 Si on change le 3ème bit du message précédent, que devient ce message ? Même question en remplaçant 3ème par 2ème, puis en remplaçant 3ème par 12ème.

Exercice 1-5

À l'aide d'un éditeur hexadécimal, on visualise le contenu d'un fichier.

```
00000000 26 12 FE 03 50
```

Question 1 Sachant que ce fichier contient une information textuelle codée avec le codage de longueur variable (cf table 1), trouvez l'information codée.

Question 2 Quelle taille (en octets) aurait cette information si elle avait été codée à l'aide du code ASCII ?

Question 3 Quelle serait cette taille si on utilisait un codage de longueur fixe minimale ?

Exercice 1-6 Codage bitmap des images

Le codage *bitmap* d'une image consiste à coder la couleur de chaque pixel (point) de l'image par une chaîne binaire de longueur fixée. Ainsi, pour représenter une image ne contenant que du noir et du blanc, il suffit de coder chacune de ces deux couleurs sur un bit. Les formats les plus fréquents codent chaque pixel sur 1 bit (2 couleurs), 4 bits (16 couleurs), 8 bits (256 couleurs) et 24 bits ($\approx 16 \times 10^6$ couleurs).

Question 1

Une image Noir et Blanc de 10 lignes de 16 pixels est codée en bitmap. Voici la représentation hexadécimale de cette image

0000 0F00 1088 0388 0CBE 1088 1188 0E80 0000 7FFE

Dessinez l'image.

Question 2

Calculez la taille (en nombre d'octets) de la plus petite représentation possible en bitmap d'une image comprenant 200×300 pixels et 100 couleurs.

Question 3 *Images compressibles*

La représentation "bitmap" d'une image peut être inutilement gourmande en espace mémoire (Pensez à une image blanche).

Proposez une représentation plus compacte des images.

1.2 Deuxième série

Exercice 1-7 *Nombre de mots*

Pour les deux questions qui suivent l'alphabet contient au moins les deux lettres **a** et **b**.

Question 1 Combien y a-t-il de mots de 5 lettres commençant par **a** et terminant par **b**?

Question 2 Combien y a-t-il de mots de 10 lettres comprenant trois **a** et 2 **b**?

Exercice 1-8 *Peut-on tout compresser ?*

Soit \mathcal{A} un alphabet à q lettres.

Question 1 Combien y a-t-il de mots de longueur strictement inférieure à n ($n \neq 0$) ?

Un fichier est une suite finie d'octets. On peut considérer que c'est un mot sur l'alphabet \mathcal{A} des 256 octets.

Un logiciel de compression transforme un fichier en un autre fichier qu'on espère être plus petit. Bien entendu une compression n'est utile que l'opération inverse, la décompression, existe. On peut modéliser un logiciel de compression comme une fonction z de l'ensemble des mots dans lui-même :

$$z : \mathcal{A}^* \rightarrow \mathcal{A}^*,$$

telle que pour tout mot $\mathbf{u} \in \mathcal{A}^*$, on a

$$|z(\mathbf{u})| < |\mathbf{u}|,$$

autrement dit le fichier obtenu par compression est strictement plus petit que le fichier d'origine.

Question 2 Quelle propriété mathématique la fonction z doit-elle posséder pour assurer l'existence de l'opération de décompression ?

Question 3 En considérant la compression des fichiers de taille n , n étant un entier fixé, déduisez de la propriété trouvée pour z l'inexistence d'une telle fonction, et concluez.

Exercice 1-9 *Mots commutants*

Soit l'alphabet $=\{\mathbf{a}, \mathbf{b}\}$.

Question 1 Trouvez deux mots \mathbf{u} et \mathbf{v} non vides sur l'alphabet \mathcal{A} tels que $\mathbf{u.v} = \mathbf{v.u}$.

Question 2 Montrez que pour tout mot $\mathbf{w} \in \mathcal{A}^*$ et tout entier $p, q \in \mathbb{N}$, les mots $\mathbf{u} = \mathbf{w}^p$ et $\mathbf{v} = \mathbf{w}^q$ commutent.

Question 3 Réciproquement, montrez que la condition précédente est nécessaire pour que deux mots commutent.

Exercice 1-10 *Concaténation de langages*

Soient L et M deux langages sur un alphabet \mathcal{A} . La *concaténation* de L et M est le langage noté $L.M$ défini par

$$L.M = \{\mathbf{u.v} \mid \mathbf{u} \in L, \mathbf{v} \in M\}$$

autrement dit, $L.M$ est l'ensemble de tous les mots obtenus en concaténant un mot de L à un mot de M .

On considère l'alphabet $\mathcal{A} = \{\mathbf{a}, \mathbf{b}\}$.

Question 1 Calculez le langage $L.M$ dans chacun des cas suivants

1. $L = \{\varepsilon, \mathbf{ab}, \mathbf{b}, \mathbf{aba}\}$, $M = \{\mathbf{a}, \mathbf{ab}\}$
2. $L = \{\mathbf{a}, \mathbf{ab}, \mathbf{bba}\}$, $M = \{\mathbf{a}, \mathbf{ba}, \mathbf{aab}\}$
3. $L = \emptyset$, M quelconque.

Question 2 À quelle condition nécessaire et suffisante, a-t-on $L \subseteq L.M$? et $M \subseteq L.M$?

Question 3 Lorsque L et M sont deux langages finis, quel est le nombre maximal de mots de $L.M$? Dans quel cas ce maximum est-il atteint?

1.3 Troisième série

Exercice 1-11 *Code ou pas code?*

Parmi les langages suivants sur l'alphabet $\mathcal{A} = \{0, 1\}$ indiquez ceux qui sont des codes.

1. $L = \emptyset$
2. $L = \{\varepsilon\}$
3. $L = \{\mathbf{u}\}$, où \mathbf{u} est un mot non vide sur l'alphabet \mathcal{A}
4. $L = \{0011, 1001, 0110\}$
5. $L = \{00, 010, 111, 01101, 1100\}$
6. $L = \{010101, 0101\}$
7. $L = \{0, 01, 110, 1101, 1111\}$
8. $L = \{0, 01, 100\}$

Exercice 1-12

Question 1 Peut-on trouver un sous-ensemble strict de $Y = \{0, 01, 110, 1101, 1111\}$ qui ne soit pas un code?

Question 2 Peut-on trouver un code contenant strictement Y ?

Question 3 Reprendre les deux questions avec $Y = \{0, 01, 100\}$.

Exercice 1-13

Quels sont les codes sur un alphabet à une seule lettre?

Exercice 1-14

Soient u et v deux mots sur un alphabet \mathcal{A} . À quelle condition le langage constitué des deux mots u et v est-il un code?

Exercice 1-15

Pour quelles valeurs des entiers naturels m , n et p , le langage

$$L = \{a^m, baba, ba^n, ab, bab^p\}$$

est-il un code ?

Exercice 1-16

Soit $C \subset \mathcal{A}^*$ un code fini contenant k mots, et $n \in \mathbb{N}$ un entier.

Question 1

Quel est le cardinal de C^n ?

Question 2

Quel est le cardinal de $C^{\leq n} = \bigcup_{0 \leq i \leq n} C^i$?

1.4 Quatrième série

Exercice 1-17

Les langages binaires ci-dessous sont-ils des codes ? Si non, existe-t-il un code ayant la même distribution de longueurs de mots ?

1. $L_1 = \{0, 01, 101, 110\}$
2. $L_2 = \{0, 01, 101, 110, 111\}$

Exercice 1-18 Codage binaire de l'alphabet latin

\mathcal{S} désigne l'alphabet latin (26 lettres) et $\mathcal{A} = \{0, 1\}$.

On veut coder chaque lettre de \mathcal{S} par un mot de \mathcal{A}^*

Question 1 On choisit un codage de longueur fixe. Quelle est la longueur minimale des mots du code ?

Question 2 On choisit l'application définie par $\mathbf{c}(a) = 0$, $\mathbf{c}(b) = 00$, $\mathbf{c}(c) = 000$, ... Cette application est-elle un codage ?

Existe-t-il un codage préfixe dont les mots ont même longueur que ceux de \mathbf{c} ?

Question 3 Existe-t-il un codage préfixe possédant

- 8 mots de code de longueur 4,
- 8 mots de code de longueur 5,
- 8 mots de code de longueur 6,
- et 2 mots de code de longueur 7.

Si oui trouvez-en un.

Question 4 Peut-on remplacer les mots de longueur 7 par des mots plus courts tout en gardant le caractère préfixe du codage et sans changer les autres mots ?

Question 5 Existe-t-il un codage préfixe à 8 mots de longueur 4 et 18 de longueur 5 ?

Exercice 1-19

On désire coder, à l'aide d'un alphabet $\mathcal{A} = \{0, 1, 2\}$, un alphabet \mathcal{S} de dix lettres avec un mot-code d'une lettre, les neuf autres ayant trois lettres au maximum. On désire de plus que ce codage soit préfixe.

Question 1 Cela est-il possible ?

Question 2 Combien de mots-code de trois lettres faut-il au minimum ?

Exercice 1-20

Parmi les mots de code utilisé par un codage binaire préfixe, on trouve 0, 10 et 1100. Combien de mots de code de longueur 6 ce codage peut-il avoir ?

Exercice 1-21

Question 1 Vérifier que l'on peut raccourcir la longueur d'un mot d'un code binaire si et seulement si l'inégalité de Kraft est stricte.

Question 2 Qu'en est-il pour des codes sur un alphabet à plus de deux lettres ?

Exercice 1-22 Codes maximaux

Un code C est dit *maximal* si pour tout mot $u \in \mathcal{A}^*$, $C \cup \{u\}$ n'est pas un code. Montrez qu'un code préfixe fini est maximal si et seulement si $K(C) = 1$.

2 Le codage ASCII

Code initié en 1963 puis complété en 1967, le code ASCII-1967 est devenu une norme ISO-646 en 1983. Il permet de coder 128 caractères, à l'origine sur 7 bits, puis sur huit bits. Cf table 2.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2	ESP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	'	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

TABLE 2 – Table du codage ASCII (ISO-646)