



Sommaire

Préambule	1
1 Introduction	2
2 La représentation des caractères	2
2.1 Le code Ascii	2
2.2 Le code Ascii étendu	3
2.3 Le code Ansi	3
2.4 L'Unicode	4
3 Exercices :	5



Préambule

Traduisez ce message :

†□◆◆ ☉❖ℳ⌘ ℳ□○□□⌘◆ ℳℳ ○ℳ◆◆
 ☉ℳℳ ☉ ○☉⌘◆ ℳ■ ☉❖ℳ⌘☉◆◆ ℳ
 □○□□⌘◆ ●ℳ ◆ℳ■◆ ✎

-	☉
?	✎
,	☉
a	☉
c	ℳ

e	ℳ
g	ℳ
i	⌘
l	●
m	○

n	■
o	□
p	□
r	□
s	◆

u	◆
v	†
v	❖
z	⌘

Vous avez compris ce message, mais en avez-vous compris le sens ?



1 Introduction

Comme vu précédemment, la représentation d'un message et donc de caractères peut prendre des formes bizarres. Dans le cours précédent, nous avons vu la codification des nombres, et surtout comment les ordinateurs manipulaient les données dans un langage qui leur était propre. Pour représenter les autres informations, nous avons besoin d'une symbologie standard qui prenne en compte :

Les lettres de l'alphabet	26
La différenciation majuscule, minuscule	26 de plus
Les chiffres 0 → 9	10
Les autres caractères (+ - * , ? @...)	
Des caractères de commande (retour, dring...)	
....	

2 La représentation des caractères

2.1 *Le code Ascii*

En fait, la première codification a été liée au mode de transmission des données vers l'ordinateur. Les premiers ordinateurs étaient liés à des télétypes qui servaient de point d'entrée des données. Ces premiers systèmes étaient raccordés en liaison série sur les ordinateurs.

Les données étaient codées sur le système de base, à savoir l'octet.

Pour des raisons de fiabilité, cet octet a été séparé en 2 :

- Les 7 bits de poids fort représentent la donnée
- Le bit de poids fort est réservé pour un contrôle de la donnée (bit de parité)
- Ce qui donne 128 possibilités

Ainsi est née la codification ASCII (American Standard Code for Information Interchange)

Chaque caractère est représenté par un code hexadécimal :

A → 41H

B → 42H

a → 61H

...

De plus, la codification ASCII a respecté l'ordre de l'alphabet, ce qui permet de manipuler les données par leur codification et d'en effectuer un tri alphanumérique. Il est possible dans les logiciels d'utiliser le caractère ASCII ou le code correspondant.





		000	001	010	011	100	101	110	111		
		0	1	2	3	4	5	6	7		
0000	0	NUL	DLE	SP	0	@	P	`	p	NUL	Absence de caractère, blanc, espace
0001	1	SOH	DC1		1	A	Q	a	q	SOH	Start of Heading : début en-tête
0010	2	STX	DC2	"	2	B	R	b	r	STX	Start of Text
0011	3	ETX	DC3	#	3	C	S	c	s	ETX	End of Text
0100	4	EOT	DC4	\$	4	D	T	d	t	EOT	End of Transmission
0101	5	ENQ	NAK	%	5	E	U	e	u	ENQ	Enquiry Demande
0110	6	ACK	SYN	&	6	F	V	f	v	ACK	Acknowledge, accusé réception
0111	7	BEL	ETB	'	7	G	W	g	w	BEL	Bell, sonnette
1000	8	BS	CAN	(8	H	X	h	x	BS	Backspace marche arrière 1 caractère
1001	9	HT	EM)	9	I	Y	i	y	HT	Horizontale Tabulation
1010	10	LF	SUB	*	:	J	Z	j	z	LF	Line Fed retour à la nouvelle ligne
1011	11	VT	ESC	+	;	K	[k	{	VT	Vertical Tabulation
1100	12	FF	FS	,	<	L	\	l	!	FF	Form Fed, passage page suivante
1101	13	CR	GS	-	=	M]	m	}	CR	Carriage Return, retour chariot
1110	14	SO	RS	.	>	N	'	n	~	SO	Shift Out caractère suivant non std
1111	15	SI	US	/	?	O	-	o	DEL	SI	Shift In retour au caractère std

NUL Absence de caractère, blanc, espace
 SOH Start of Heading : début en-tête
 STX Start of Text
 ETX End of Text
 EOT End of Transmission
 ENQ Enquiry Demande
 ACK Acknowledge, accusé réception
 BEL Bell, sonnette
 BS Backspace marche arrière 1 caractère
 HT Horizontale Tabulation
 LF Line Fed retour à la nouvelle ligne
 VT Vertical Tabulation
 FF Form Fed, passage page suivante
 CR Carriage Return, retour chariot
 SO Shift Out caractère suivant non std
 SI Shift In retour au caractère std
 DLE DataLink Escape chgmt de signific.
 NAK Negative Acknowledgment
 SYN Synchronous, caractère de synchro.
 ETB End Of Transmission Block
 CAN Cancel annulation de la donnée précédente
 SUB Substitute remplacement
 ESC Escape caractère de ctrl d'extension
 FS File Separator
 GS Groupe Separator
 RS Record Separator
 US United Separator
 SP Space Espace
 DEL Delete, suppression
 DC1 à DC4 : caractères de commandes

TABLE DES CODES ASCII

2.2 Le code Ascii étendu

Certains constructeurs, dont IBM suivis par tous les fabricants de compatibles, ont enrichi cette table ASCII en utilisant le 8^{ème} bit, ce qui double le nombre de caractères représentables (256). Les caractères supplémentaires sont essentiellement :

- Les caractères accentués utilisés dans la langue française.
- Des jeux de caractères utilisés dans d'autres langues ($\approx, \sqrt{\quad}, \dots$).
- Quelques symboles mathématiques (\exists, \int).
- Des caractères semi-graphiques qui permettent de réaliser des petits dessins géométriques (cadres, soulignés, etc.).

Le code ASCII 8 bits existe en 2 variantes : le jeu de caractères **IBM PC** (c'est le jeu de caractères standard du DOS) et le jeu de caractères **ISO-ANSI** (jeu de caractères international utilisé dans Windows). Ces deux jeux diffèrent notamment au niveau des caractères nationaux accentués et des caractères semi-graphiques.

2.3 Le code Ansi

Les logiciels sous Windows utilisent la norme ANSI, qui reprend en grande partie le code ASCII, et propose des extensions différentes selon le « code de page » retenu. Ainsi, le code page 850 est très employé en France, alors que le code page 864 définit un jeu de caractères « arabe ».

L'utilisation du code ANSI se fait de la même manière que pour un code ASCII (intersections colonne-ligne).



Page de codes 850

	0	16	32	48	64	80	96	112	128	144	160	176	192	208	224	240
0-0	▶		0	@	P	'	p	Ç	É	á	⋮	Ł	ó	Ó	-	
1-1	☺	◀	!	l	A	Q	a	q	ü	æ	i	⋮	Ł	Đ	β	±
2-2	☹	↓	ˆ	2	B	R	b	r	é	Æ	ó	⋮	Ł	É	Ó	=
3-3	♥	!!	#	3	C	S	c	s	á	ó	ú	⋮	Ł	É	Ó	¼
4-4	♦	§	5	4	D	T	d	t	á	ó	ñ	⋮	Ł	É	ó	¶
5-5	♣	§	%	5	E	U	e	u	á	ó	ñ	⋮	Ł	É	ó	§
6-6	♠	—	&	6	F	V	f	v	á	ó	ñ	⋮	Ł	É	ó	÷
7-7	•	↑	ˆ	7	G	W	g	w	ç	ú	º	⋮	Ł	É	ó	.
8-8	■	†	(8	H	X	h	x	é	ý	ı	⋮	Ł	É	ó	°
9-9	○	↓)	9	I	Y	i	y	ē	ö	⊙	⋮	Ł	É	ó	..
10-A	☉	→	.	:	J	Z	j	z	è	ü	⋮	Ł	É	ó	.	
11-B	♂	←	+	:	K	[k	{	ı	ø	¼	⋮	Ł	É	ó	ı
12-C	♀	↳	.	<	L	\	l		i	é	¼	⋮	Ł	É	ó	ı
13-D	♫	↔	-	=	M]m	})	i	ø	i	é	⋮	Ł	É	ó	ı
14-E	♪	▲	.	>	N	^	n	ˆ	Ä	x	«	⋮	Ł	É	ó	ı
15-F	♫	▼	/	?	O	_	o	Δ	À	f	»	⋮	Ł	É	ó	ı

Page de codes 864

	0	16	32	48	64	80	96	112	128	144	160	176	192	208	224	240
0-0	▶		0	@	P	'	p	°	β	.	ı	ı	ı	ı	ı	ı
1-1	☺	◀	!	l	A	Q	a	q	.	∞	.	ı	ı	ı	ı	ı
2-2	♫	↓	ˆ	2	B	R	b	r	.	∞	.	ı	ı	ı	ı	ı
3-3	♫	!!	#	3	C	S	c	s	√	±	ı	ı	ı	ı	ı	ı
4-4	♫	§	5	4	D	T	d	t	¼	ı	ı	ı	ı	ı	ı	ı
5-5	♫	§	%	5	E	U	e	u	¼	ı	ı	ı	ı	ı	ı	ı
6-6		—	&	6	F	V	f	v		≈	ı	ı	ı	ı	ı	ı
7-7	♫	↑	ˆ	7	G	W	g	w	+	ı	ı	ı	ı	ı	ı	ı
8-8	♫	†	(8	H	X	h	x	ı	ı	ı	ı	ı	ı	ı	ı
9-9	♫	↓)	9	I	Y	i	y	ı	ı	ı	ı	ı	ı	ı	ı
10-A	♫	→	.	:	J	Z	j	z	ı	ı	ı	ı	ı	ı	ı	ı
11-B	♫	←	+	:	K	[k	{	ı	ı	ı	ı	ı	ı	ı	ı
12-C	♫	↳	.	<	L	\	l		ı	ı	ı	ı	ı	ı	ı	ı
13-D	♫	↔	-	=	M]m	})	ı	ı	ı	ı	ı	ı	ı	ı	ı
14-E	♫	▲	.	>	N	^	n	ˆ	ı	ı	ı	ı	ı	ı	ı	ı
15-F	♫	▼	/	?	O	_	o	Δ	ı	ı	ı	ı	ı	ı	ı	ı

2.4 L'Unicode

Compte-tenu de l'extension mondiale de l'informatique et de la diversité de plus en plus importante des caractères à stocker, les organismes de normalisation ISO travaillent depuis 1988 à la création d'un code universel (UNIversal CODE). Ce code cherche à supprimer les pages de code et à reprendre l'ensemble des caractères utilisés de par le monde. Il reste basé sur la codification ASCII.

Il se représente sous 3 formes :

- UTF8 étant beaucoup utilisé sur Internet et présentant un codage de taille variable.
- UTF16 de taille fixe sur 2 octets → Théoriquement 65 535 caractères différents.
- UTF32 sur 4 octets

En UTF8, tous les caractères de base ASCII 7 bits sont codés sur 1 octet. Au-delà, les caractères peuvent prendre de 2 à 4 octets.

Exemple :

é	C3 A9	à	C3 A0	ù	C3 B9
μ	C2 B5	§	C2 A7	Ê	C3 8A

Pour plus d'infos : <http://www.unicode.org/charts/>

Pour utiliser un convertisseur : <http://www.macchiato.com/unicode/convert.html>



3 Exercices :

1. Codez en ASCII :

Nous sommes en BTS :

2. Transcodez cette séquence :

56 69 76 65 20 6C 65 20 42 54 53 20 65 74 20 6C 65 73 20 63 6F 75 72 73 20 64 27 41 4D
53 49

3. Repérez et décodez les séquences de texte. De quel type de fichier s'agit-il ?

0043EFF0	6F 20 32 20 43 72 65 61 74 69 6F 6E 20 64 27 75
0043F000	6E 65 20 64 69 73 71 75 65 74 74 65 20 64 65 20
0043F010	42 6F 6F 74 20 0A 20 20 20 20 20 20 55 74 69 6C
0043F020	69 73 61 74 69 6F 6E 20 64 65 20 57 69 6E 68 65
0043F030	78 0A 20 20 20 20 20 20 56 6F 69 72 20 6C 65 73
0043F040	20 64 69 66 66 E9 72 65 6E 63 65 73 0A 20 20 20
0043F050	20 20 20 72 E9 64 69 67 65 72 20 75 6E 20 63 6F
0043F060	6D 70 74 65 20 72 65 6E 64 75 07 00 00 00 31 00

4. Repérez et décodez les séquences de texte. De quel type de fichier s'agit-il ?

0057CA00	47 49 46 38 39 61 3D 00 45 00 E6 00 00 00 00 00
0057CA10	FF FF FF FD FD FD FA FA FA F8 F8 F8 F5 F5 F5 F3
0057CA20	F3 F3 F0 F0 F0 EE EE EE EB EB EB E9 E9 E9 E6 E6
0057CA30	E6 E4 E4 E4 E1 E1 E1 DF DF DF DC DC DC DA DA DA

5. Repérez et décodez les séquences de texte. De quel type de fichier s'agit-il ?

00448A00	4D 5A 50 00 02 00 00 00 04 00 0F 00 FF FF 00 00
00448A10	B8 00 00 00 00 00 00 00 40 00 1A 00 00 00 00 00
00448A20	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00448A30	00 00 00 00 00 00 00 00 00 00 00 00 00 01 00 00
00448A40	BA 10 00 0E 1F B4 09 CD 21 B8 01 4C CD 21 90 90
00448A50	54 68 69 73 20 70 72 6F 67 72 61 6D 20 6D 75 73
00448A60	74 20 62 65 20 72 75 6E 20 75 6E 64 65 72 20 57
00448A70	69 6E 33 32 0D 0A 24 37 00 00 00 00 00 00 00 00

6. Repérez et décodez les séquences de texte. Que remarquez-vous de particulier ?

Trouvez une explication

1A 01 41 00 20 00 72 00 65 00 63 00 6F 00 70 00
69 00 65 00 72 00 20 00 43 00 20 00 70 00 72 00
6F 00 67 00 72 00 61 00 6D 00 20 00 66 00 69 00
6C 00 65 00 73 00 C4 01 B0 38 00 00 00 00 01 00
60 00 50 00 00 00 00 00 05 00 00 00 00 00 05 00



Correction des exercices

1. Codez en ASCII :

```
4E 6F 75 73 20 73 6F 6D 6D 65 73 20 65
6E 20 42 54 53
```

Nous sommes en BTS :

2. Transcodez cette séquence :

56 69 76 65 20 6C 65 20 42 54 53 20 65 74 20 6C 65 73 20 63 6F 75 72 73 20 64 27 41 4D 53 49 → Vive le BTS et les cours d'AMSI

3. Repérez et décodez les séquences de texte. De quel type de fichier s'agit-il ?

0043EFFF0	6F 20 32 20 43 72 65 61 74 69 6F 6E 20 64 27 75	o 2 Creation d'u
0043F000	6E 65 20 64 69 73 71 75 65 74 74 65 20 64 65 20	ne disquette de
0043F010	42 6F 6F 74 20 0A 20 20 20 20 20 55 74 69 6C	Boot . Util
0043F020	69 73 61 74 69 6F 6E 20 64 65 20 57 69 6E 68 65	isation de Winhe
0043F030	78 0A 20 20 20 20 20 20 56 6F 69 72 20 6C 65 73	x. Voir les
0043F040	20 64 69 66 66 E9 72 65 6E 63 65 73 0A 20 20 20	différences.
0043F050	20 20 20 72 E9 64 69 67 65 72 20 75 6E 20 63 6F	rédiger un co
0043F060	6D 70 74 65 20 72 65 6E 64 75 07 00 00 00 31 00	mpte rendu...1.

Fichier Excel : Mon cahier de texte

4. Repérez et décodez les séquences de texte. De quel type de fichier s'agit-il ?

0057CA00	47 49 46 38 39 61 3D 00 45 00 E6 00 00 00 00 00	GIF89a= .E.æ.....
0057CA10	FF FF FF FD FD FD FA FA FA F8 F8 F8 F5 F5 F3	yyyyyyuuuuææððóó
0057CA20	F3 F3 F0 F0 F0 EE EE EE EB EB EB E9 E9 E6 E6	óóððííííèèééææ
0057CA30	E6 E4 E4 E4 E1 E1 DF DF DF DC DC DC DA DA DA	ææææááááBáBúUúUú

Fichier Image extension Gif

5. Repérez et décodez les séquences de texte. De quel type de fichier s'agit-il ?

00448A00	4D 5A 50 00 02 00 00 00 04 00 0F 00 FF FF 00 00	MZP.....ÿÿ..
00448A10	B8 00 00 00 00 00 00 00 40 00 1A 00 00 00 00 00@.....
00448A20	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00448A30	00 00 00 00 00 00 00 00 00 00 00 00 00 01 00 00
00448A40	BA 10 00 0E 1F B4 09 CD 21 B8 01 4C CD 21 90 90!.,Lf!
00448A50	54 68 69 73 20 70 72 6F 67 72 61 6D 20 6D 75 73	This program mus
00448A60	74 20 62 65 20 72 75 6E 20 75 6E 64 65 72 20 57	t be run under W
00448A70	69 6E 33 32 0D 0A 24 37 00 00 00 00 00 00 00 00	in32..\$7.....
00448A80	00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00	

Un logiciel Exécutable (Winhex)

6. Repérez et décodez les séquences de texte. Que remarquez-vous de particulier ?
Trouvez une explication

3DE5666E0	1A 01 41 00 20 00 72 00 65 00 63 00 6F 00 70 00	..A. r.e.c.o.p.
3DE5666F0	69 00 65 00 72 00 20 00 43 00 20 00 70 00 72 00	i.e.r. .C. .p.r.
3DE566700	6F 00 67 00 72 00 61 00 6D 00 20 00 66 00 69 00	o.g.r.a.m. .f.i.
3DE566710	6C 00 65 00 73 00 C4 01 B0 38 00 00 00 00 01 00	l.e.s.Å.*8.....
3DE566720	60 00 50 00 00 00 00 00 05 00 00 00 00 00 05 00	`P.....

Il s'agit d'un répertoire en NTFS

Les caractères sont codés sur 2 octets → Codage en UTF16