



Architectures des Ordinateurs

Chapitre 5

(Représentation des caractères)

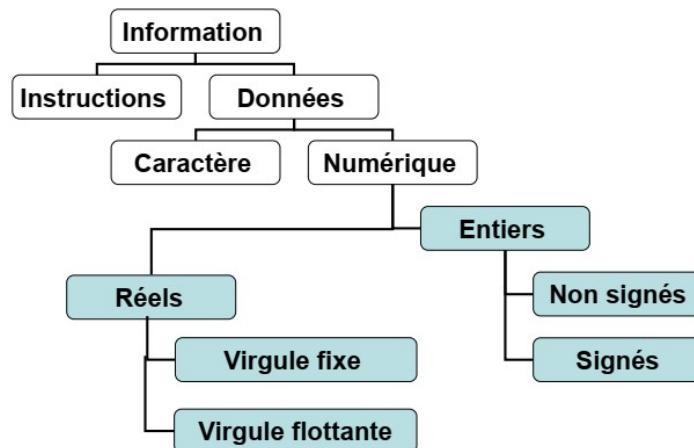
Mokrani Hocine
dr.mokrani@gmail.com

1

Représentation des caractères

2

Rappel des types de base



3

Caractères ASCII

- L'ASCII (American Standard Code for Information Interchange), créé en 1961, associe un nombre à chaque caractère.
- Les caractères sont codés sur 7 bits.
- Il y a 128 caractères différents (de 00 à 7F en hexadécimal).
- Contient les caractères nécessaires pour écrire en anglais.
- Les caractères de numéro 0 à 31 et le 127 ne sont pas affichables.

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

4

Limites de code ASCII

- Absence des caractères des langues étrangères à l'anglais.
 - Les ordinateurs travaillant sur un multiple de huit bits (multiple d'un octet) depuis les années 1970. (Chaque caractère d'un texte en ASCII est souvent stocké dans un octet dont le 8^e bit est 0).
 - Pourquoi l'utilisé alors?

=> Héritage des programmes qui fonctionne encore sur le code ASCII.

=> certains systèmes de messagerie électronique et de SMS fonctionnent avec des mots composés de seulement sept bits.

5

ASCII étendu

Il existe des extensions (sur 8 bits, de 80 à FF). Cependant, ces extensions ne sont pas standardisées.

128	Ҫ	144	Ӯ	160	Ӵ	176	ӷ	193	ӹ	209	Ӯ	225	Ӯ	241	Ӹ
129	ӻ	145	Ӯ	161	ӵ	177	ӷ	194	ӹ	210	Ӯ	226	Ӯ	242	ӹ
130	Ӷ	146	Ӯ	162	Ӷ	178	ӷ	195	ӹ	211	Ӯ	227	Ӷ	243	ӹ
131	Ӹ	147	Ӷ	163	Ӹ	179	ӷ	196	ӹ	212	Ӷ	228	Ӷ	244	ӹ
132	ӹ	148	Ӷ	164	ӹ	180	ӷ	197	ӹ	213	Ӷ	229	Ӷ	245	ӹ
133	ӷ	149	Ӷ	165	ӷ	181	ӷ	198	ӹ	214	Ӷ	230	ӷ	246	ӹ
134	Ӹ	150	ӷ	166	Ӹ	182	ӷ	199	ӹ	215	ӷ	231	Ӹ	247	ӷ
135	ӹ	151	ӷ	167	ӹ	183	ӷ	200	ӹ	216	ӷ	232	ӷ	248	ӹ
136	Ӹ	152	ӹ	168	Ӹ	184	ӹ	201	ӷ	217	ӹ	233	Ӹ	249	ӹ
137	ӹ	153	ӷ	169	ӹ	185	ӷ	202	ӹ	218	ӷ	234	ӹ	250	ӹ
138	Ӹ	154	ӷ	170	Ӹ	186	ӷ	203	ӷ	219	ӷ	235	Ӹ	251	ӷ
139	ӹ	156	Ӹ	171	ӹ	187	ӷ	204	ӷ	220	ӷ	236	ӹ	252	ӹ
140	Ӹ	157	ӹ	172	Ӹ	188	ӷ	205	=	221	ӷ	237	Ӹ	253	ӹ
141	ӹ	158	ӷ	173	ӹ	189	ӷ	206	ӷ	222	ӷ	238	ӹ	254	ӷ
142	Ӹ	159	ӷ	174	ӹ	190	ӷ	207	ӷ	223	ӷ	239	ӹ	255	ӷ
143	ӹ	192	ӷ	175	ӹ	191	ӷ	208	ӷ	224	ӷ	240	ӹ	256	ӷ

6

Caractères ISO-8859-1

- souvent appelée **Latin-1** ou **Europe occidentale**.
- reprend le codage des caractères imprimables d'US-ASCII.
- cette norme était utilisée par de nombreux systèmes d'exploitation, dont Unix et Windows.

ISO-8859-1																
	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1x	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2x	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	-
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
8x	PAD	HOP	BPH	NBH	IND	NEL	SSA	ESA	HTS	HTJ	VTS	PLD	PLU	RI	SS2	SS3
9x	DCS	PU1	PU2	STS	CCH	MW	SPA	EPA	SOS	SGCI	SCI	CSI	ST	OSC	PM	APC
Ax	NBSP	i	¢	£	¤	¥	ı	§	„	®	™	«	»	¬	®	-
Bx	°	±	²	³	‘	μ	¶	·	¸	¹	º	»	¼	½	¾	½
Cx	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Î	Ï	Ӯ	ӻ	ӻ
Dx	Đ	Ñ	Ó	Ó	Ó	Ó	Ó	Ø	Ù	Ú	Ú	Ü	Ý	Þ	Þ	Þ
Ex	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ï	í	ӯ	ӻ
Fx	ö	ñ	ò	ó	ô	ó	ö	÷	ø	ù	ú	ü	ÿ	þ	ÿ	ÿ

7

Quelques limites de ISO-8859-1

- Elle ignore quelques caractères européens:
 - Le symbole de l'euro € (qui n'existe pas encore lorsque ce jeu a été normalisé)
 - certaines des lettres normalement nécessaires à certaines langues théoriquement couvertes:
 - les lettres « œ », « œ » et « Ÿ » en français.
 - les lettres « š » et « Š » en tchèque.
- La nécessité d'utiliser plusieurs jeux de caractères codés pour couvrir plusieurs alphabets. (utilisation des meta-données).

8

Caractère sur 16 bits (Unicode)

- Unicode propose de coder les caractères sur 16 bits (65536 valeurs possibles).
- Il permet de coder notamment les alphabets latins, cyrilliques, les caractères chinois, indiens et arabes.

Char	Code														
Ā	100	Đ	110	Ę	118	Ķ	136	Ń	143	Ó	d3	Ś	15a	Ű	170
ā	101	đ	111	ę	119	ķ	137	ń	144	ó	f3	ś	15b	ű	171
Ă	102	Ď	10e	Ě	11a	Ľ	139;	Ň	145	Œ	152	Š	160	Ӯ	172
ă	103	ď	10f	ě	11b	ľ	13a	ň	146	œ	153	š	161	ӻ	173
Ӑ	104	Ӗ	112	Ӗ	122	Ӆ	13b	Ӯ	147	Ӱ	155	Ӯ	162	Ӯ	178
ӑ	105	ӗ	113	ӗ	123	ԓ	13c	ڻ	148	ڦ	156	ڻ	163	ڻ	179
Ҫ	106	Ҽ	115	Ҽ	124	Ը	13d	Ծ	14c	Ծ	157	Ծ	164	Ծ	17a
ҫ	107	Ҽ	116	Ҽ	12a	Ը	13e	Ծ	14d	Ծ	158	Ծ	165	Ծ	17b
Ҫ	10c	Ҽ	117	Ҽ	12b	Ը	13f	Ծ	150	Ծ	159	Ծ	166	Ծ	17c
ҫ	10d	ҽ	12e	ҽ	12f	Ը	141	Ծ	151	Ծ	15e	Ծ	167	Ծ	17d
				î	12f	î	142			§	15f			ž	17e

- Limites: il existe différentes représentations binaires dont l'interprétation peut être très différente si la représentation effective est mal déterminée.

9

Conclusion

Nous savons maintenant,
Comment en peut représenter les éléments en binaire.

Ce qu'il reste à savoir?

Représentation des réels.

Représentation du son, vidéo ...

10