

Le codage des caractères

L2

ASCII

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- American Standard Code for Information Interchange (1961)
- 128 caractères (00000000 → 11111111)
- 0 → 31 y 127 ne sont pas visualisés

ASCII

ASCII (2)

lojibaké

nicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040	 	Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	\$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	(	(	72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051	)	)	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[	[	123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174	|	
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135	]	]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177		DEL

ASCII étendu (128 → 255)

ASCII (2)

lojibaké

nicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- ISO 8859-1 (Latin-1)
- ISO 8859-5 (Cyrillic)

ASCII étendu (128 → 255)

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- ISO 8859-1 (Latin-1)
- ISO 8859-5 (Cyrillic)

iso-8859-1										
+	0	1	2	3	4	5	6	7	8	9
160		í	φ	£	¤	¥	¦	§	"	ø
170	≡	«	¬	-	®	-	°	±	²	³
180	´	µ	¶	·	¸	¹	º	»	¼	½
190	¾	¿	À	Á	Â	Ã	Ä	Å	Æ	Ç
200	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ
210	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û
220	Ü	Ý	Þ	ß	à	á	â	ã	ä	å
230	æ	ç	è	é	ê	ë	ì	í	î	ï
240	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù
250	ú	û	ü	ý	þ	ÿ				

iso-8859-5										
+	0	1	2	3	4	5	6	7	8	9
160		Ё	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ
170	Њ	Ћ	Ќ	-	Ў	Ў	А	Б	В	Г
180	Д	Е	Ж	З	И	Й	К	Л	М	Н
190	О	П	Р	С	Т	У	Ф	Х	Ц	Ч
200	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	а	б
210	в	г	д	е	ж	з	и	й	к	л
220	м	н	о	п	р	с	т	у	ф	х
230	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
240	Њ	ё	ђ	ѓ	є	ѕ	і	ї	ј	љ
250	њ	ћ	ќ	ѕ	ў	џ				

ASCII étendu (128 → 255)

ASCII (2)

lojibaké

nicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- ISO 8859-1 (Latin-1)
- ISO 8859-5 (Cyrillic)
- Ambassade de
France à Moscou :
119121,
Москва,
Улица Бурденко, 20

ASCII étendu (128 → 255)

ASCII (2)

lojibaké

nicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- ISO 8859-1 (Latin-1)
- ISO 8859-5 (Cyrillic)
- Ambassade de
France à Moscou :
119121,
Москва,
Улица Бурденко, 20

iso-8859-5										
+	0	1	2	3	4	5	6	7	8	9
160		Ё	Ђ	Ѓ	Є	Š	І	Ї	Ј	Љ
170	Њ	Ћ	Ќ	–	Ў	Џ	А	Б	В	Г
180	Д	Е	Ж	З	И	Й	К	Л	М	Н
190	О	П	Р	С	Т	У	Ф	Х	Ц	Ч
200	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	а	б
210	в	г	д	е	ж	з	и	й	к	л
220	м	н	о	п	р	с	т	у	ф	х
230	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
240	Ń	ë	ђ	ѓ	є	š	ı	ï	ј	љ
250	њ	ћ	ќ	–	ў	џ				

ASCII étendu (128 → 255)

ASCII (2)

lojibaké

nicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- ISO 8859-1 (Latin-1)
- ISO 8859-5 (Cyrillic)
- Ambassade de
France à Moscou :
119121,
Москва,
Улица Бурденко, 20

iso-8859-1										
+	0	1	2	3	4	5	6	7	8	9
160		í	φ	£	κ	¥	ì	§	"	ø
170	≡	«	¬	-	®	-	°	±	²	³
180	˘	µ	¶	·	¸	¹	º	»	¼	½
190	¾	¿	À	Á	Â	Ã	Ä	Å	Æ	Ç
200	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ
210	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û
220	Ü	Ý	Þ	ß	à	á	â	ã	ä	å
230	æ	ç	è	é	ê	ë	ì	í	î	ï
240	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù
250	ú	û	ü	ý	þ	ÿ				

ASCII étendu (128 → 255)

ASCII (2)

lojibaké

nicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- ISO 8859-1 (Latin-1)
- ISO 8859-5 (Cyrillic)
- Ambassade de
France à Moscou :
119121,
Москва,
Улица Бурденко, 20

Mojibaké

SCII (2)

lojibaké

nicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

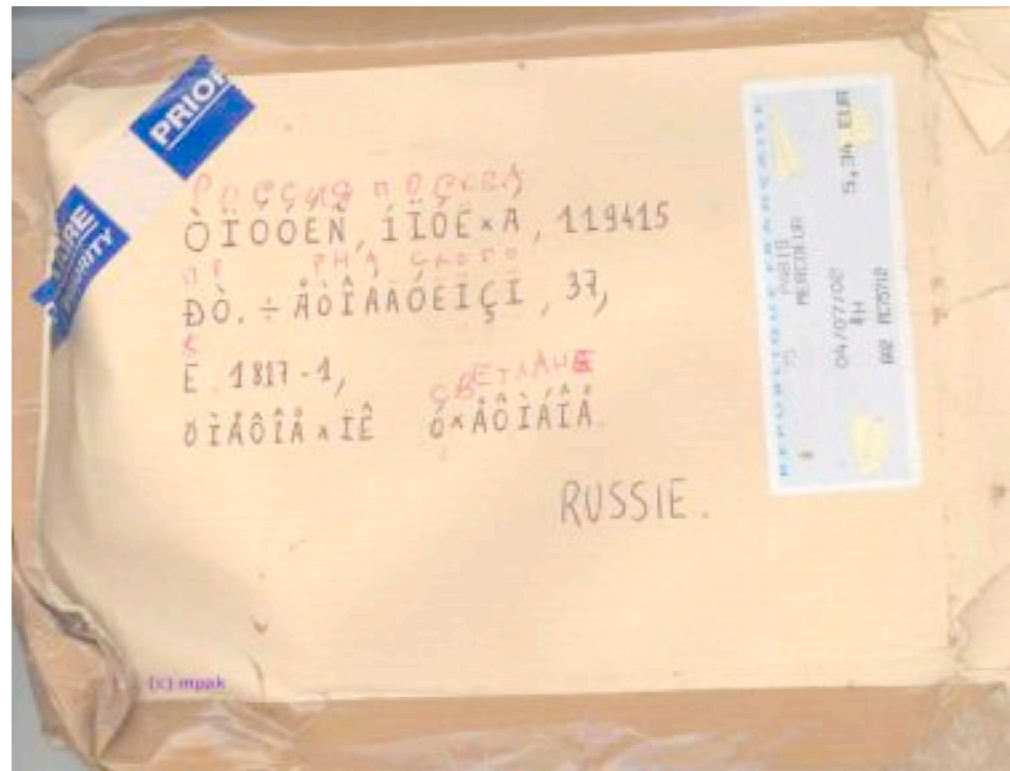
Ordre Logique

Unification

Composition

Convertibilité

- 文字化け
- *Moji* : Carácter
- *Baké* : Cambiado
- Ejemplo :



Unicode

ASCII (2)

Mojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- <http://hapax.qc.ca>
- Régler le problème du *Mojibaké*
 - Un caractère $\leftrightarrow$ un nombre
 - Pas de problèmes de copier / coller
- Tout dans le même logiciel

Definition

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Unicode est un code qui associe un identifiant unique à chaque caractère, indépendamment du système d'exploitation, du logiciel et de la langue.
- La conception de Unicode suit 10 principes.

Universalité

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Ensemble unique qui contient tous les caractères nécessaire du monde.
- Ψ:: ∫ ə Ў 𐄂 a z ٱ ٲ ٲ Å Ñ

Efficacité

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Tous les caractères sont également faciles d'accès
- Les caractères d'un même alphabet se retrouvent à des emplacements consécutifs
- Caractères de ponctuation

Caractères y non glyphes

SCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Unicode : codage et non visualisation ou impression

Glyphes (œils)	Caractères Unicode/ ISO 10646
A <i>Æ</i> À A Ä A	U+0041 lettre majuscule latine a
ffi <i>ffi</i> ffi ffi	U+0066 lettre minuscule latine f + U+0066 lettre minuscule latine f + U+0069 lettre minuscule latine i
ف ف	U+0647 lettre arabe fa'

Sémantique

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Casse du caractère
 - MAJUSCULE
 - minuscule
 - Casse de titre
- Direction :
 - Droite
 - Gauche
 - Faible
 - Neutre

Sémantique

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Sémantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Symétrie
 - Ajuster par rapport à la direction du texte :
 $\int$, $\leq$
- Nombres :
 - 1, 3, iv
- Caractères combinatoires
 - ٲ ٴ ڤ ښ

Texte brut

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Ni balisage, ni formattage

Ordre logique

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Utilise les propriétés de direction d'unicode

Unification

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Mêmes caractères pour gato, chat, cat...
 - Unification des écritures et pas des langues comme pour les idéogrammes chinois, japonais et coréens
- Unification des glyphes (Cf. caractères et glyphes)

Composition dynamique

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glyphes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- $e + \ddot{\circ} \Rightarrow \ddot{e}$
- $\ddot{\circ}$ peut être appliqué à des voyelles, mais aussi à des consonnes

Convertibilité

ASCII (2)

lojibaké

Unicode (12)

Définition

Universalité

Efficacité

Carac. y glypes

Semantique (2)

Texte brut

Ordre Logique

Unification

Composition

Convertibilité

- Definition de méthodes :
 - d'Importation
 - d'Exportation
- Avec la garantie de permettre des conversions sans perte