31268 Web Systems

UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Information Technology

Week 07: Computer Science 1

Part 2: Data Representation

Computer Science

- **Storage and processing of information**
  - History of computing
  - Computation
  - Memory
  - Coding
- **Under the bonnet**
  - Representation of information
  - Number systems
- **Logic and Mathematics**
  - Boolean Algebra
  - Binary Arithmetic

Data Representation

- You will need to understand:
  - how data exists in a computer as binary numbers
  - Types of character representation
  - How data is represented in files

Examples of Data…

- Documents
- Images
- Video
- Audio

It's all numbers to the computer

Character Representation

- CPU’s dont understand characters or images or Powerpoint documents etc
- ➔ O/S designers invented character encoding to represent alphabets, digits and symbols (e.g. punctuation)
- A-Z, a-z, 0-9, !@#$%^&*()_+-={}[]\;:’”/.,<>?

Data Representation

- CPU’s dont understand characters or images or Powerpoint documents etc

- ➔ O/S designers invented character encoding to represent alphabets, digits and symbols (e.g. punctuation)
- A-Z, a-z, 0-9, !@#$%^&*()_+-={}[]\;:’”/.,<>?

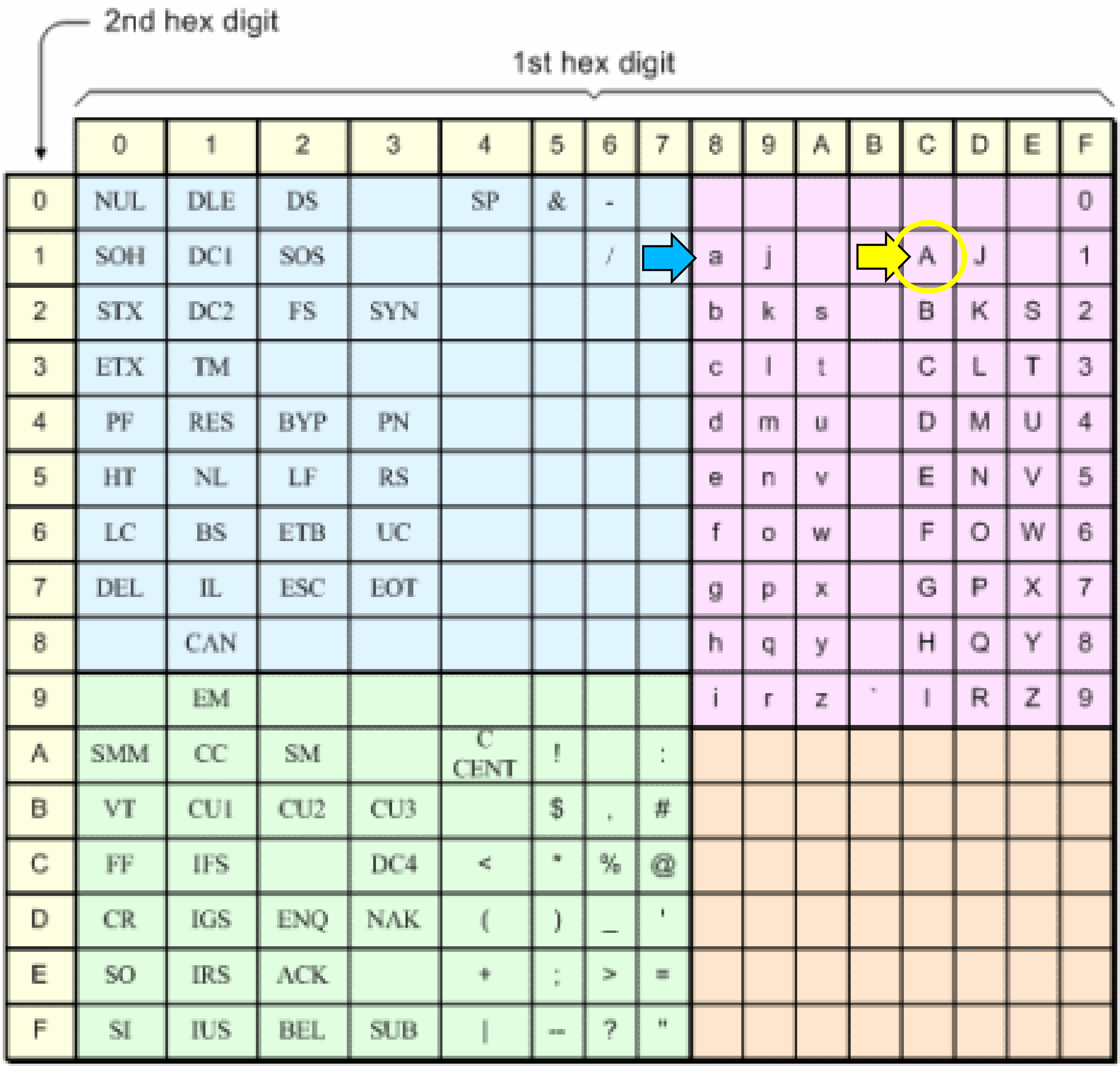
- ➔ later needed to extend to other languages eg European (Latin diacritics like áâãäæà) and other symbols (eg √©£¥✓)

- ➔ also needed to define file formats to contain this data …

Examples of Character encoding…

- Pre-Historic:
  - **EBCDIC** – used only IBM mainframes, punch cards

EBCDIC



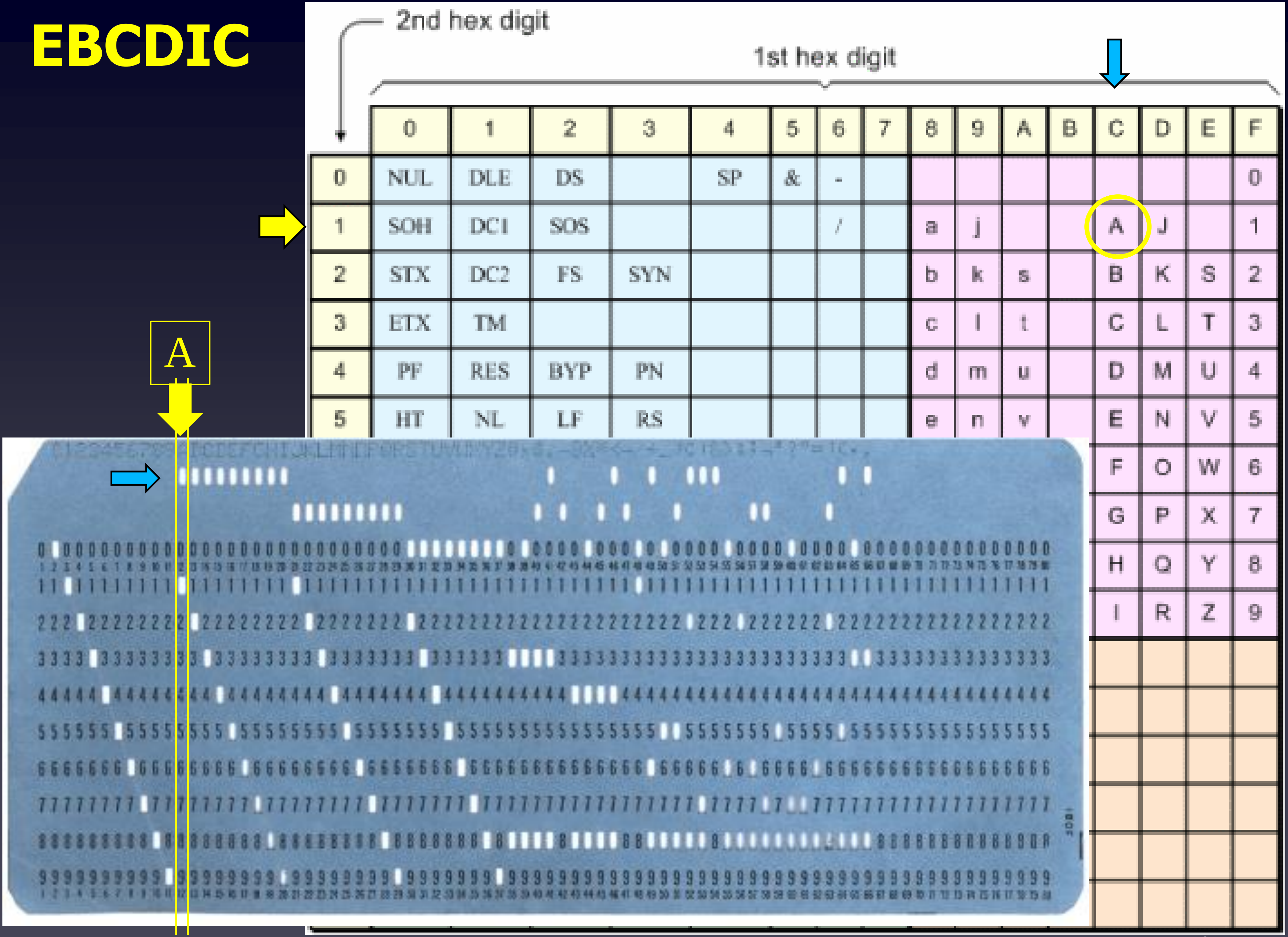
2nd hex digit

1st hex digit

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	DLE	DS		SP	&	-									0
1	SOH	DC1	SOS				/		a	j			A	J		1
2	STX	DC2	FS	SYN					b	k	s		B	K	S	2
3	ETX	TM							c	l	t		C	L	T	3
4	PF	RES	BYP	PN					d	m	u		D	M	U	4
5	HT	NL	LF	RS					e	n	v		E	N	V	5
6	LC	BS	ETB	UC					f	o	w		F	O	W	6
7	DEL	IL	ESC	EOT					g	p	x		G	P	X	7
8		CAN							h	q	y		H	Q	Y	8
9		EM							i	r	z	`	I	R	Z	9
A	SMM	CC	SM		C CENT	!		:								
B	VT	CU1	CU2	CU3		$	,	#								
C	FF	IFS		DC4	<	*	%	@								
D	CR	IGS	ENQ	NAK	(	)	_	'								
E	SO	IRS	ACK		+	;	>	=								
F	SI	IUS	BEL	SUB	|	¬	?	"								

http://m.eet.com/media/1119500/ebcdic-01.gif

EBCDIC



	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	DLE	DS		SP	&	-									0
1	SOH	DC1	SOS				/		a	j			A	J		1
2	STX	DC2	FS	SYN					b	k	s		B	K	S	2
3	ETX	TM							c	l	t		C	L	T	3
4	PF	RES	BYP	PN					d	m	u		D	M	U	4
5	HT	NL	LF	RS					e	n	v		E	N	V	5
													F	O	W	6
													G	P	X	7
													H	Q	Y	8
													I	R	Z	9

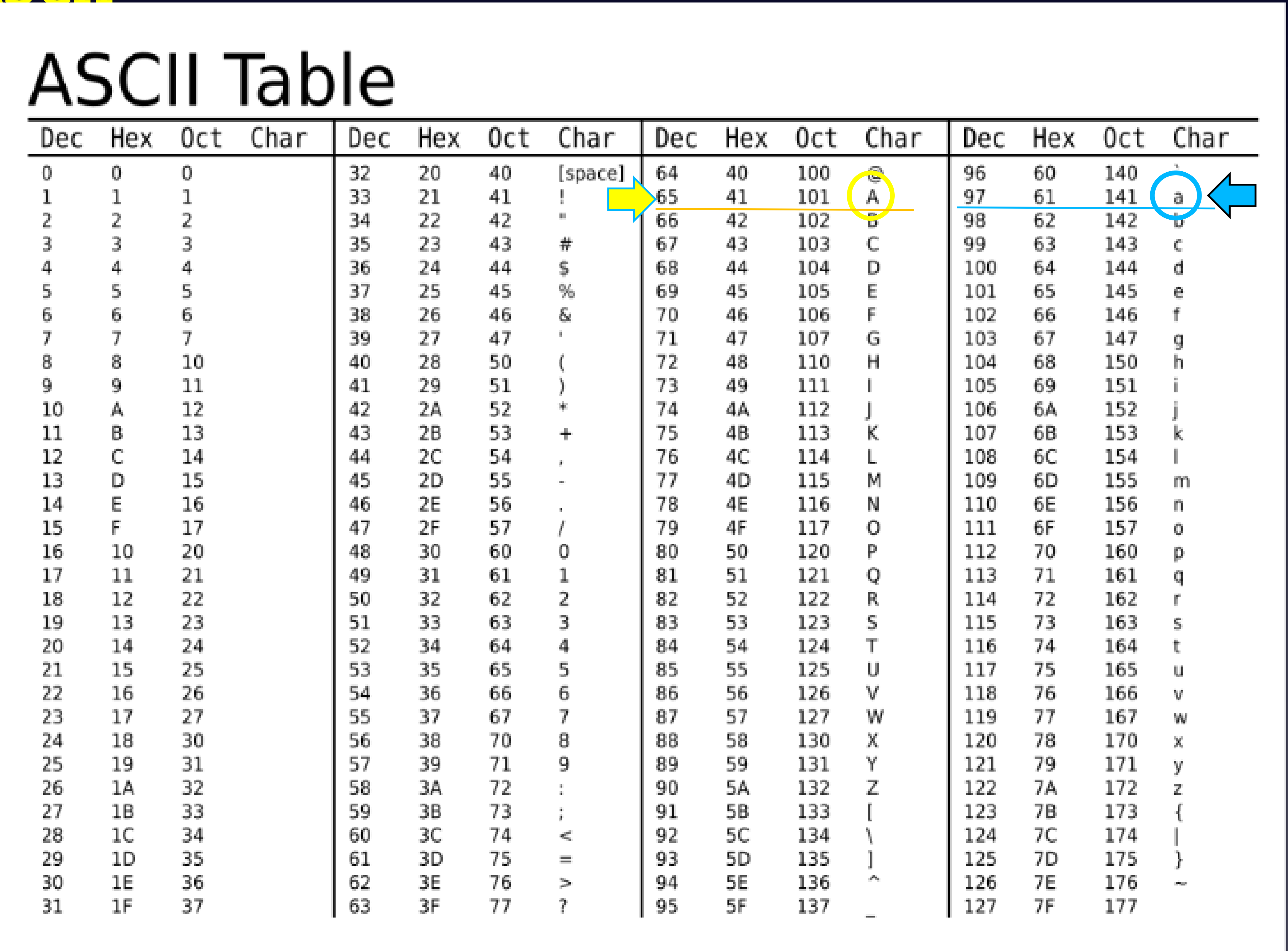
https://www.greenviewdata.com/img/ebcdic-punch-card.png

Examples of Character encoding…

- Historic:
  - **ASCII** – 7 bits – 128 different characters (add a 0 to the front to make it 8 bits or 1 byte per character)
  - **ISO-8859** – extension of ASCII to include Latin alphabets eg German, Scandinavian characters
    - 15 subsets to cover Latin, Cyrillic, Arabic, Hebrew, Greek, Thai

https://en.wikipedia.org/wiki/ISO/IEC_8859

ASCII



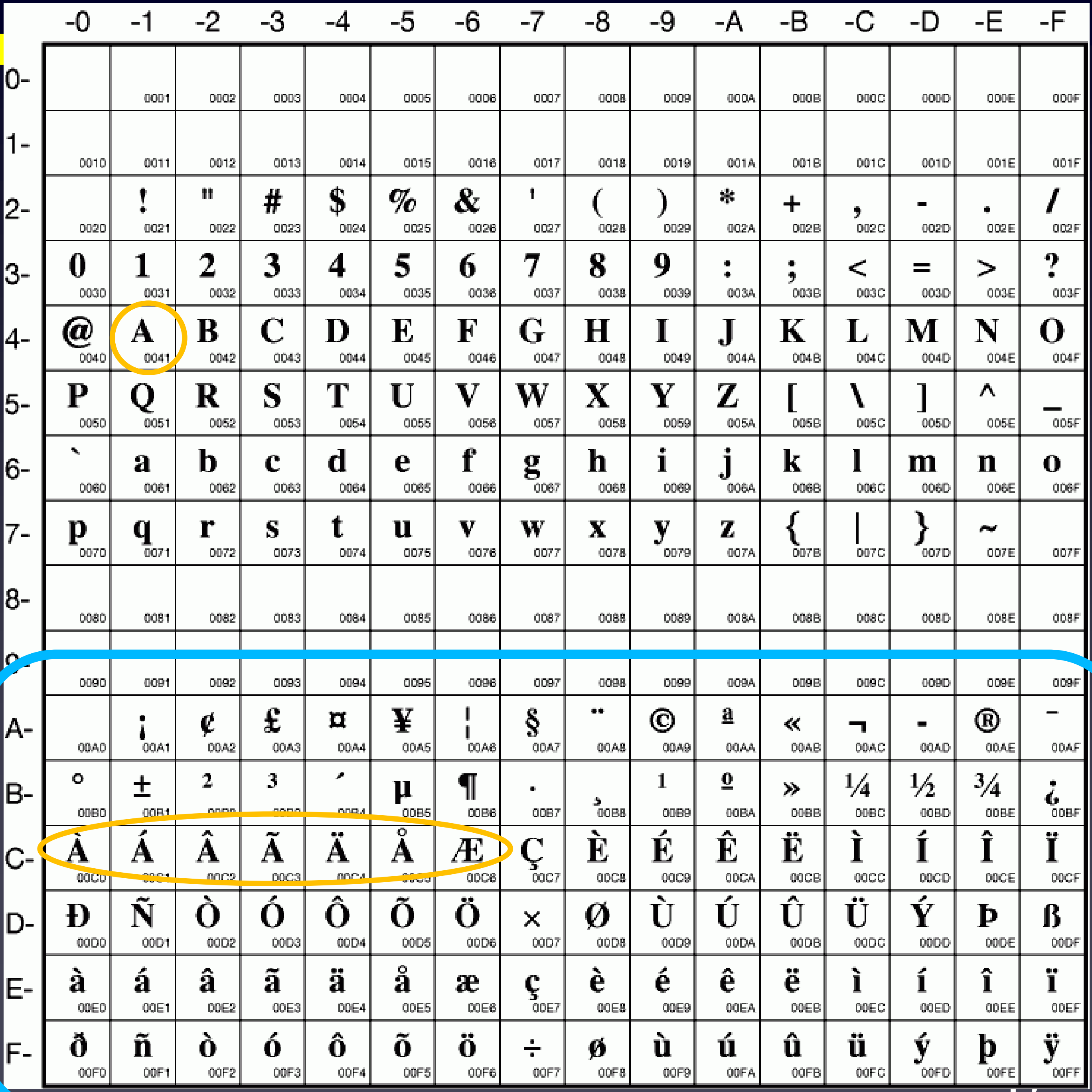
ASCII Table

Dec	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char	Dec	Hex	Oct	Char
0	0	0		32	20	40	[space]	64	40	100	@	96	60	140	`
1	1	1		33	21	41	!	65	41	101	A	97	61	141	a
2	2	2		34	22	42	"	66	42	102	B	98	62	142	b
3	3	3		35	23	43	#	67	43	103	C	99	63	143	c
4	4	4		36	24	44	$	68	44	104	D	100	64	144	d
5	5	5		37	25	45	%	69	45	105	E	101	65	145	e
6	6	6		38	26	46	&	70	46	106	F	102	66	146	f
7	7	7		39	27	47	'	71	47	107	G	103	67	147	g
8	8	10		40	28	50	(	72	48	110	H	104	68	150	h
9	9	11		41	29	51	)	73	49	111	I	105	69	151	i
10	A	12		42	2A	52	*	74	4A	112	J	106	6A	152	j
11	B	13		43	2B	53	+	75	4B	113	K	107	6B	153	k
12	C	14		44	2C	54	,	76	4C	114	L	108	6C	154	l
13	D	15		45	2D	55	-	77	4D	115	M	109	6D	155	m
14	E	16		46	2E	56	.	78	4E	116	N	110	6E	156	n
15	F	17		47	2F	57	/	79	4F	117	O	111	6F	157	o
16	10	20		48	30	60	0	80	50	120	P	112	70	160	p
17	11	21		49	31	61	1	81	51	121	Q	113	71	161	q
18	12	22		50	32	62	2	82	52	122	R	114	72	162	r
19	13	23		51	33	63	3	83	53	123	S	115	73	163	s
20	14	24		52	34	64	4	84	54	124	T	116	74	164	t
21	15	25		53	35	65	5	85	55	125	U	117	75	165	u
22	16	26		54	36	66	6	86	56	126	V	118	76	166	v
23	17	27		55	37	67	7	87	57	127	W	119	77	167	w
24	18	30		56	38	70	8	88	58	130	X	120	78	170	x
25	19	31		57	39	71	9	89	59	131	Y	121	79	171	y
26	1A	32		58	3A	72	:	90	5A	132	Z	122	7A	172	z
27	1B	33		59	3B	73	;	91	5B	133	[	123	7B	173	{
28	1C	34		60	3C	74	<	92	5C	134	\	124	7C	174	|
29	1D	35		61	3D	75	=	93	5D	135	]	125	7D	175	}
30	1E	36		62	3E	76	>	94	5E	136	^	126	7E	176	~
31	1F	37		63	3F	77	?	95	5F	137	_	127	7F	177	

ISO/IEC-8859-1

- http://www.unicodetools.com/

A= 41 (hex)



	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F
0-		0001	0002	0003	0004	0005	0006	0007	0008	0009	000A	000B	000C	000D	000E	000F
1-	0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F
2-	0020	! 0021	" 0022	# 0023	$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	- 002D	. 002E	/ 002F
3-	0 0030	1 0031	2 0032	3 0033	4 0034	5 0035	6 0036	7 0037	8 0038	9 0039	: 003A	; 003B	< 003C	= 003D	> 003E	? 003F
4-	@ 0040	A 0041	B 0042	C 0043	D 0044	E 0045	F 0046	G 0047	H 0048	I 0049	J 004A	K 004B	L 004C	M 004D	N 004E	O 004F
5-	P 0050	Q 0051	R 0052	S 0053	T 0054	U 0055	V 0056	W 0057	X 0058	Y 0059	Z 005A	[005B	\ 005C	] 005D	^ 005E	_ 005F
6-	` 0060	a 0061	b 0062	c 0063	d 0064	e 0065	f 0066	g 0067	h 0068	i 0069	j 006A	k 006B	l 006C	m 006D	n 006E	o 006F
7-	p 0070	q 0071	r 0072	s 0073	t 0074	u 0075	v 0076	w 0077	x 0078	y 0079	z 007A	{ 007B	| 007C	} 007D	~ 007E	007F
8-	0080	0081	0082	0083	0084	0085	0086	0087	0088	0089	008A	008B	008C	008D	008E	008F
9-	0090	0091	0092	0093	0094	0095	0096	0097	0098	0099	009A	009B	009C	009D	009E	009F
A-	00A0	¡ 00A1	¢ 00A2	£ 00A3	¤ 00A4	¥ 00A5	¦ 00A6	§ 00A7	¨ 00A8	© 00A9	ª 00AA	« 00AB	¬ 00AC	- 00AD	® 00AE	¯ 00AF
B-	° 00B0	± 00B1	² 00B2	³ 00B3	´ 00B4	µ 00B5	¶ 00B6	· 00B7	¸ 00B8	¹ 00B9	º 00BA	» 00BB	¼ 00BC	½ 00BD	¾ 00BE	¿ 00BF
C-	À 00C0	Á 00C1	Â 00C2	Ã 00C3	Ä 00C4	Å 00C5	Æ 00C6	Ç 00C7	È 00C8	É 00C9	Ê 00CA	Ë 00CB	Ì 00CC	Í 00CD	Î 00CE	Ï 00CF
D-	Ð 00D0	Ñ 00D1	Ò 00D2	Ó 00D3	Ô 00D4	Õ 00D5	Ö 00D6	× 00D7	Ø 00D8	Ù 00D9	Ú 00DA	Û 00DB	Ü 00DC	Ý 00DD	Þ 00DE	ß 00DF
E-	à 00E0	á 00E1	â 00E2	ã 00E3	ä 00E4	å 00E5	æ 00E6	ç 00E7	è 00E8	é 00E9	ê 00EA	ë 00EB	ì 00EC	í 00ED	î 00EE	ï 00EF
F-	ð 00F0	ñ 00F1	ò 00F2	ó 00F3	ô 00F4	õ 00F5	ö 00F6	÷ 00F7	ø 00F8	ù 00F9	ú 00FA	û 00FB	ü 00FC	ý 00FD	þ 00FE	ÿ 00FF

Latin-1 extensions→

Examples

- Since computers only know numbers, here is a the **ASCII** code is used to encode text.
- To print the word “Hello” on a printer, the computer sends the sequence of numbers:
-
- http://www.lookuptables.com/

Examples

- Since computers only know numbers, here is a the **ASCII** code is used to encode text.

- To print the word "Hello" on a printer, the computer sends the sequence of numbers:

Bin	01001000	01100101	01101100	01101100	01101111
Dec	104	101	108	108	111
Char	**H**	**e**	**l**	**l**	**o**

- Hello = 01001000 01100101 01101100 01101100 01101111
- http://www.lookuptables.com/

Examples of Data…

- Current encoding
  - **Unicode** – expands ASCII to 16 bits or more to represent over 120,000 characters
  - UTF-8 – most common, compatible with ASCII
    - Variable length encoding:

– Character# 00-127	1 byte
– Character > 128	2 or more bytes

  - UTF-16 – every character (including A-Z) as 16 bits or 32 bit
  - The first 128 characters of Unicode are the same as ASCII for backward compatibility
  - See http://www.unicode.org/charts/

Examples of Data…

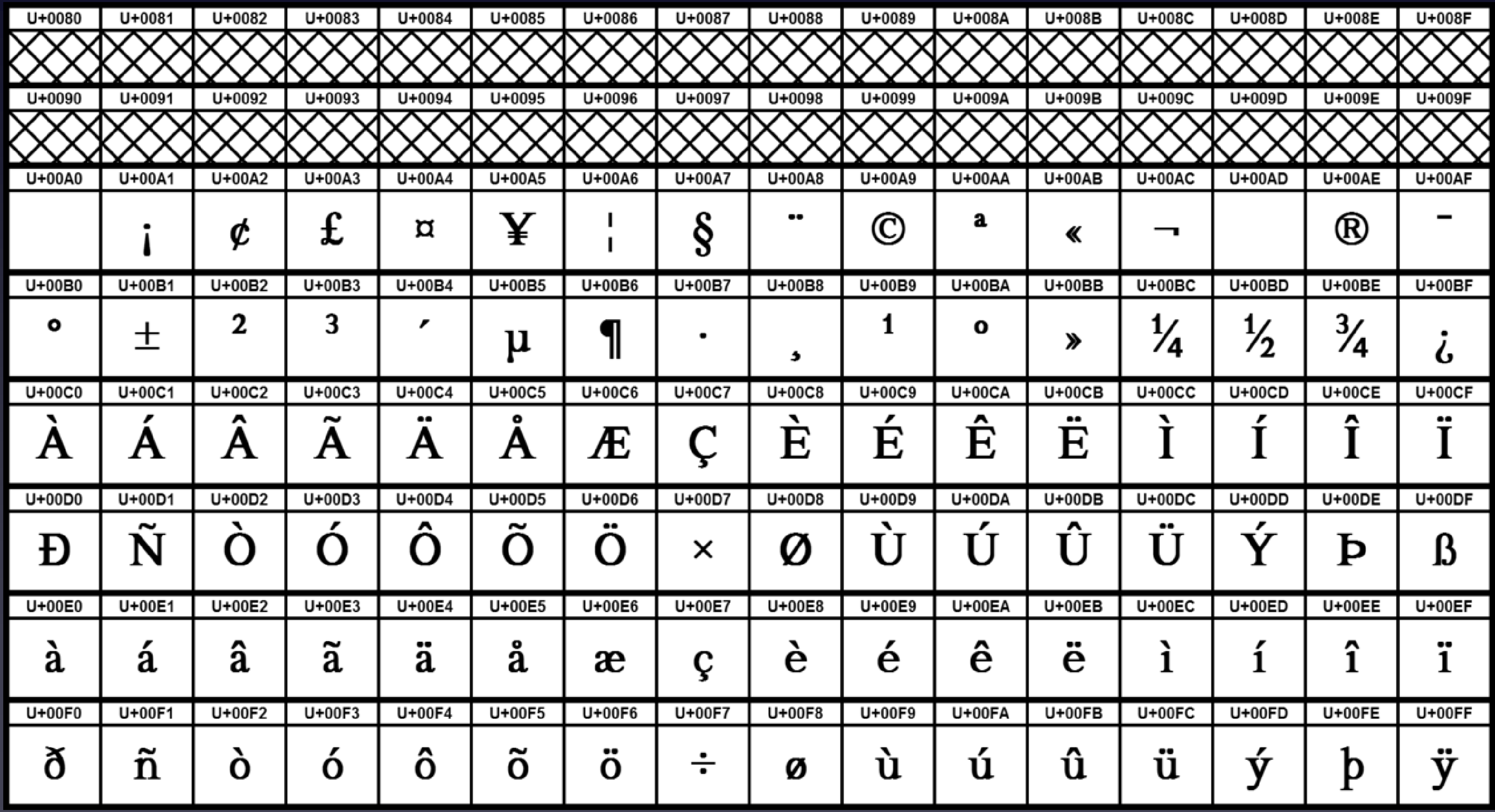
- Current encoding
  - **Unicode** – expands ASCII to 16 bits or more to represent over 120,000 characters
  - UTF-8 – most common, compatible with ASCII
    - Variable length encoding:

– Character# 00-127	1 byte
– Character > 128	2 or more bytes

HTML standard

  - UTF-16 – every character (including A-Z) as 16 bits or 32 bit
    - Commonly called a Double Byte Character Set (DBCS)
  - The first 128 characters of Unicode are the same as ASCII for backward compatibility
  - See http://www.unicode.org/charts/

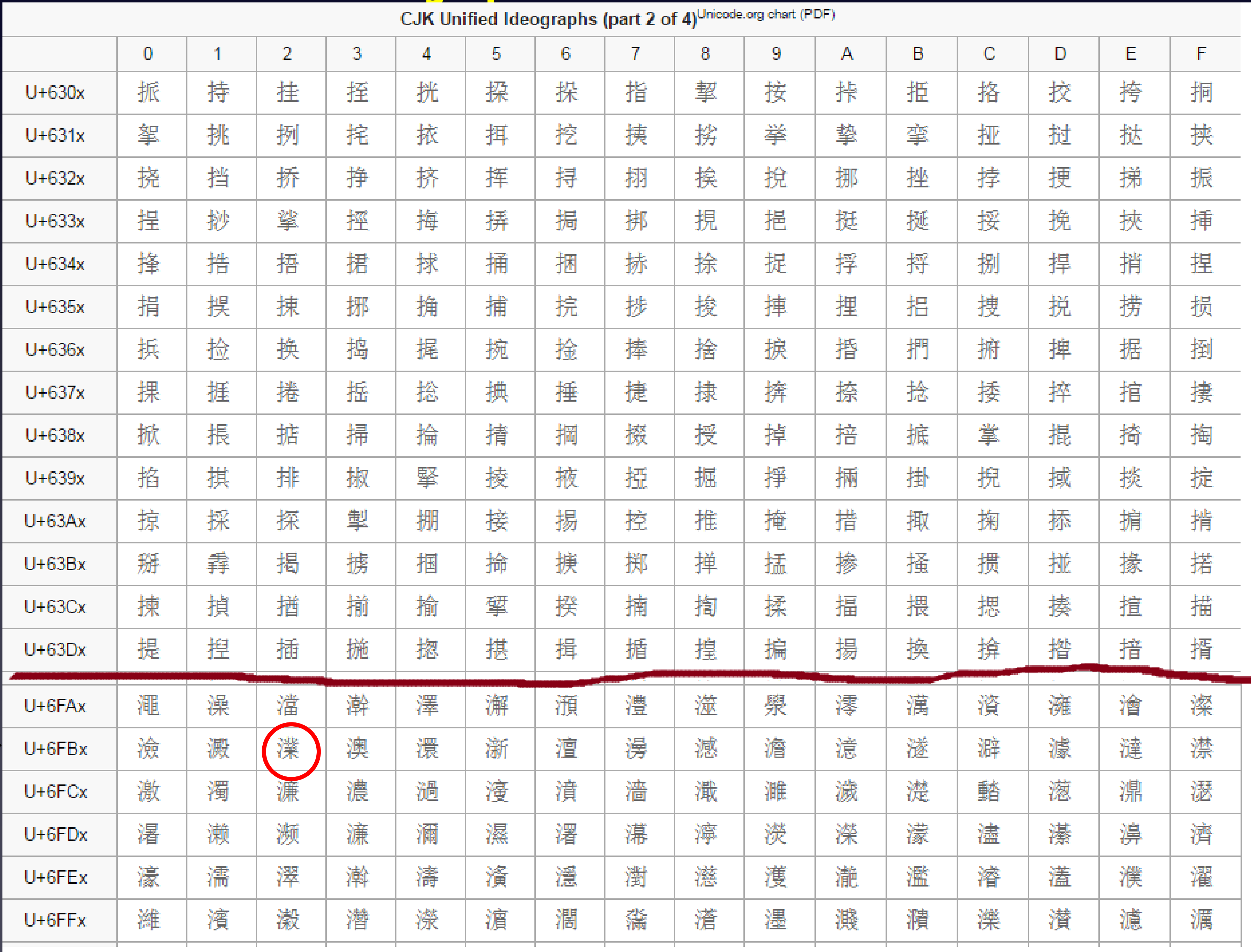
Unicode - Latin-1_Supplement



U+0080	U+0081	U+0082	U+0083	U+0084	U+0085	U+0086	U+0087	U+0088	U+0089	U+008A	U+008B	U+008C	U+008D	U+008E	U+008F
▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩
U+0090	U+0091	U+0092	U+0093	U+0094	U+0095	U+0096	U+0097	U+0098	U+0099	U+009A	U+009B	U+009C	U+009D	U+009E	U+009F
▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩	▩
U+00A0	U+00A1	U+00A2	U+00A3	U+00A4	U+00A5	U+00A6	U+00A7	U+00A8	U+00A9	U+00AA	U+00AB	U+00AC	U+00AD	U+00AE	U+00AF
	¡	¢	£	¤	¥	¦	§	¨	©	ª	«	¬		®	¯
U+00B0	U+00B1	U+00B2	U+00B3	U+00B4	U+00B5	U+00B6	U+00B7	U+00B8	U+00B9	U+00BA	U+00BB	U+00BC	U+00BD	U+00BE	U+00BF
°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
U+00C0	U+00C1	U+00C2	U+00C3	U+00C4	U+00C5	U+00C6	U+00C7	U+00C8	U+00C9	U+00CA	U+00CB	U+00CC	U+00CD	U+00CE	U+00CF
À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
U+00D0	U+00D1	U+00D2	U+00D3	U+00D4	U+00D5	U+00D6	U+00D7	U+00D8	U+00D9	U+00DA	U+00DB	U+00DC	U+00DD	U+00DE	U+00DF
Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
U+00E0	U+00E1	U+00E2	U+00E3	U+00E4	U+00E5	U+00E6	U+00E7	U+00E8	U+00E9	U+00EA	U+00EB	U+00EC	U+00ED	U+00EE	U+00EF
à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
U+00F0	U+00F1	U+00F2	U+00F3	U+00F4	U+00F5	U+00F6	U+00F7	U+00F8	U+00F9	U+00FA	U+00FB	U+00FC	U+00FD	U+00FE	U+00FF
ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

https://commons.wikimedia.org/wiki/File:UCB_Latin-1_Supplement.png

Unicode – CJK Unified Ideographs



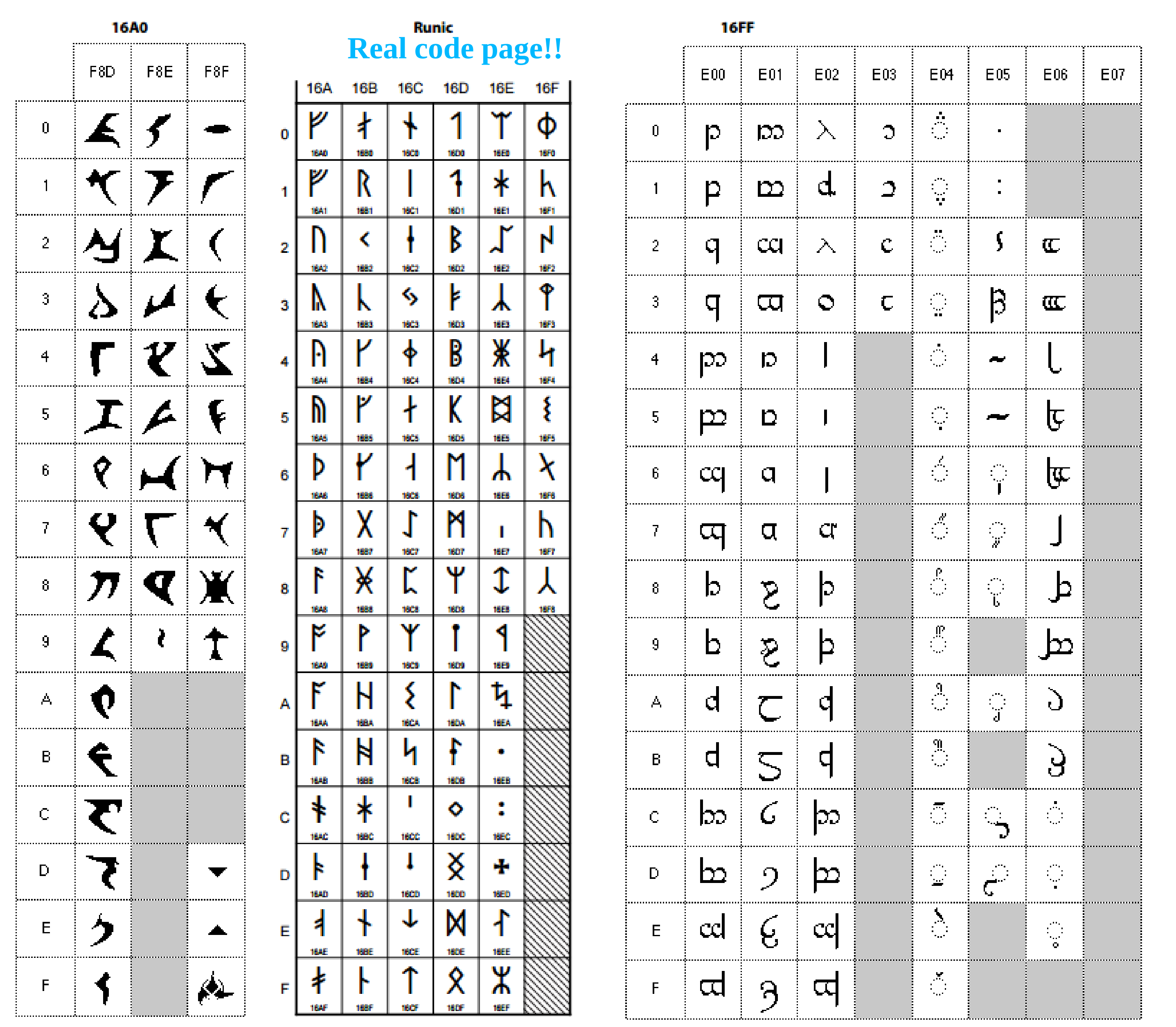
CJK Unified Ideographs (part 2 of 4)[Unicode.org chart (PDF)]

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+630x	挀	持	挂	挃	挄	挅	挆	指	挈	按	挊	挋	挌	挍	挎	挏
U+631x	挐	挑	挒	挓	挔	挕	挖	挗	挘	挙	挚	挛	挜	挝	挞	挟
U+632x	挠	挡	挢	挣	挤	挥	挦	挧	挨	挩	挪	挫	挬	挭	挮	振
U+633x	挰	挱	挲	挳	挴	挵	挶	挷	挸	挹	挺	挻	挼	挽	挾	挿
U+634x	捀	捁	捂	捃	捄	捅	捆	捇	捈	捉	捊	捋	捌	捍	捎	捏
U+635x	捐	捑	捒	捓	捔	捕	捖	捗	捘	捙	捚	捛	捜	捝	捞	损
U+636x	捠	捡	换	捣	捤	捥	捦	捧	捨	捩	捪	捫	捬	捭	据	捯
U+637x	捰	捱	捲	捳	捴	捵	捶	捷	捸	捹	捺	捻	捼	捽	捾	捿
U+638x	掀	掁	掂	掃	掄	掅	掆	掇	授	掉	掊	掋	掌	掍	掎	掏
U+639x	掐	掑	排	掓	掔	掕	掖	掗	掘	掙	掚	掛	掜	掝	掞	掟
U+63Ax	掠	採	探	掣	掤	接	掦	控	推	掩	措	掫	掬	掭	掮	掯
U+63Bx	掰	掱	掲	掳	掴	掵	掶	掷	掸	掹	掺	掻	掼	掽	掾	掿
U+63Cx	揀	揁	揂	揃	揄	揅	揆	揇	揈	揉	揊	揋	揌	揍	揎	描
U+63Dx	提	揑	插	揓	揔	揕	揖	揗	揘	揙	揚	換	揜	揝	揞	揟
U+6FAx	澠	澡	澢	澣	澤	澥	澦	澧	澨	澩	澪	澫	澬	澭	澮	澯
U+6FBx	澰	澱	澲	澳	澴	澵	澶	澷	澸	澹	澺	澻	澼	澽	澾	澿
U+6FCx	激	濁	濂	濃	濄	濅	濆	濇	濈	濉	濊	濋	濌	濍	濎	濏
U+6FDx	濐	濑	濒	濓	濔	濕	濖	濗	濘	濙	濚	濛	濜	濝	濞	濟
U+6FEx	濠	濡	濢	濣	濤	濥	濦	濧	濨	濩	濪	濫	濬	濭	濮	濯
U+6FFx	濰	濱	濲	濳	濴	濵	濶	濷	濸	濹	濺	濻	濼	濽	濾	濿

Unicode – *fun!*

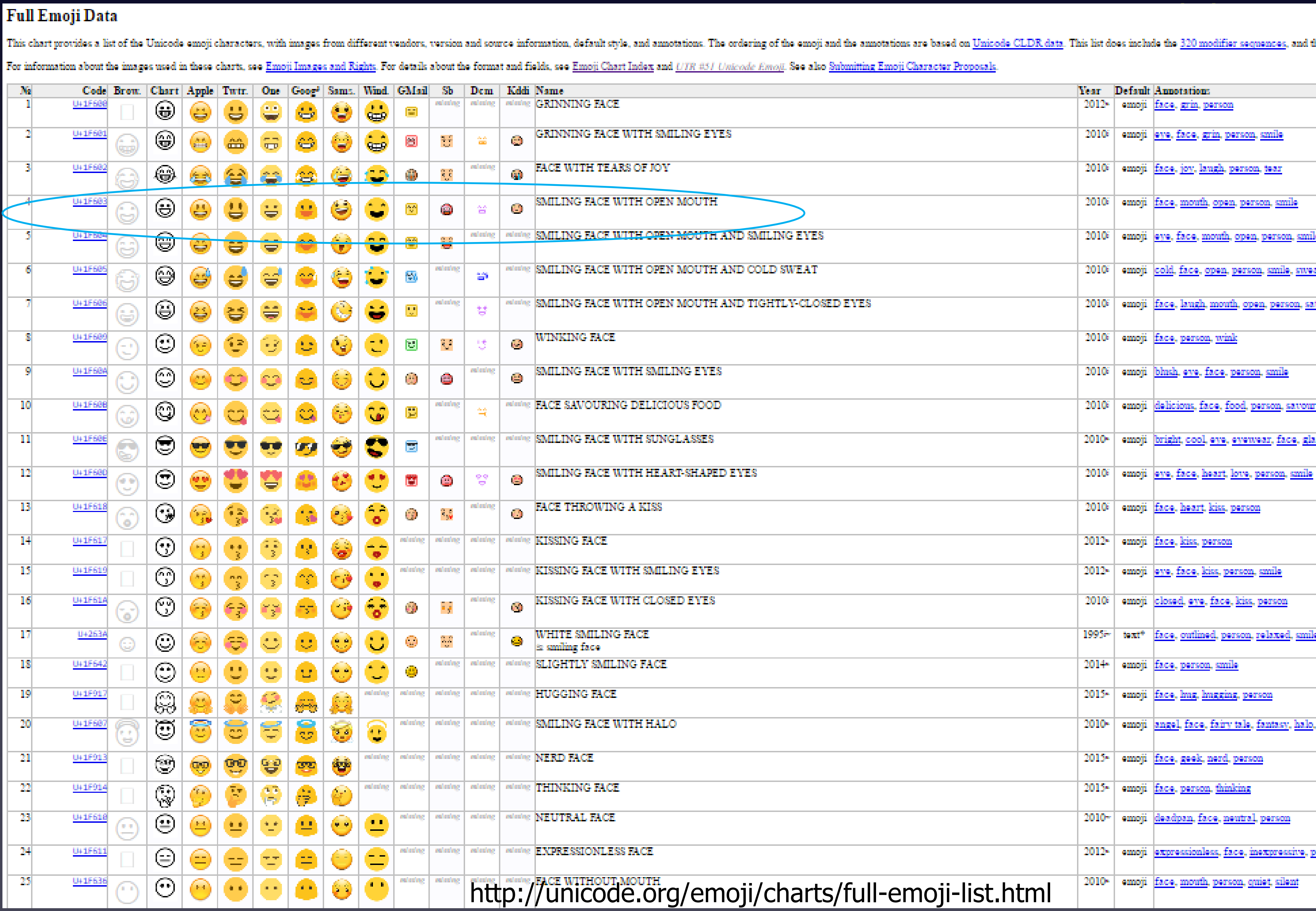
- "Unicode Private Use Area (E000-F8FF and 000F0000-0010FFFF)

http://www.evertype.com/standards/csur/


16A0
F8D
F8E
F8F
Runic
16A
16B
16C
16D
16E
16F
16FF
E00
E01
E02
E03
E04
E05
E06
E07

Real code page!!

Unicode –Emoji!



Full Emoji Data

This chart provides a list of the Unicode emoji characters, with images from different vendors, version and source information, default style, and annotations. The ordering of the emoji and the annotations are based on Unicode CLDR data. This list does include the 320 modifier sequences, and th

For information about the images used in these charts, see Emoji Images and Rights. For details about the format and fields, see Emoji Chart Index and UTR #51 Unicode Emoji. See also Submitting Emoji Character Proposals.

№	Code	Brow.	Chart	Apple	Twtr.	One	Goog¹	Sams.	Wind.	GMail	Sb	Dcm	Kddi	Name	Year	Default	Annotations
1	U+1F600	□	😀	😀	😀	😀	😀	😀	😀	😀	missing	missing	missing	GRINNING FACE	2012	emoji	face, grin, person
2	U+1F601	😁	😁	😁	😁	😁	😁	😁	😁	😁	😁	😁	😁	GRINNING FACE WITH SMILING EYES	2010	emoji	eye, face, grin, person, smile
3	U+1F602	😂	😂	😂	😂	😂	😂	😂	😂	😂	😂	missing	😂	FACE WITH TEARS OF JOY	2010	emoji	face, joy, laugh, person, tear
4	U+1F603	😃	😃	😃	😃	😃	😃	😃	😃	😃	😃	😃	😃	SMILING FACE WITH OPEN MOUTH	2010	emoji	face, mouth, open, person, smile
5	U+1F604	😄	😄	😄	😄	😄	😄	😄	😄	😄	😄	missing	missing	SMILING FACE WITH OPEN MOUTH AND SMILING EYES	2010	emoji	eye, face, mouth, open, person, smil
6	U+1F605	😅	😅	😅	😅	😅	😅	😅	😅	😅	missing	😅	missing	SMILING FACE WITH OPEN MOUTH AND COLD SWEAT	2010	emoji	cold, face, open, person, smile, swea
7	U+1F606	😆	😆	😆	😆	😆	😆	😆	😆	😆	missing	😆	missing	SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES	2010	emoji	face, laugh, mouth, open, person, sa
8	U+1F609	😉	😉	😉	😉	😉	😉	😉	😉	😉	😉	😉	😉	WINKING FACE	2010	emoji	face, person, wink
9	U+1F60A	😊	😊	😊	😊	😊	😊	😊	😊	😊	😊	missing	😊	SMILING FACE WITH SMILING EYES	2010	emoji	blush, eye, face, person, smile
10	U+1F60B	😋	😋	😋	😋	😋	😋	😋	😋	😋	missing	😋	missing	FACE SAVOURING DELICIOUS FOOD	2010	emoji	delicious, face, food, person, savour
11	U+1F60E	😎	😎	😎	😎	😎	😎	😎	😎	😎	missing	missing	missing	SMILING FACE WITH SUNGLASSES	2010	emoji	bright, cool, eye, eyewear, face, gla
12	U+1F60D	😍	😍	😍	😍	😍	😍	😍	😍	😍	😍	😍	😍	SMILING FACE WITH HEART-SHAPED EYES	2010	emoji	eye, face, heart, love, person, smile
13	U+1F618	😘	😘	😘	😘	😘	😘	😘	😘	😘	😘	missing	😘	FACE THROWING A KISS	2010	emoji	face, heart, kiss, person
14	U+1F617	□	😗	😗	😗	😗	😗	😗	😗	missing	missing	missing	missing	KISSING FACE	2012	emoji	face, kiss, person
15	U+1F619	□	😙	😙	😙	😙	😙	😙	😙	missing	missing	missing	missing	KISSING FACE WITH SMILING EYES	2012	emoji	eye, face, kiss, person, smile
16	U+1F61A	😚	😚	😚	😚	😚	😚	😚	😚	😚	😚	missing	😚	KISSING FACE WITH CLOSED EYES	2010	emoji	closed, eye, face, kiss, person
17	U+263A	☺	☺	☺️	☺️	☺️	☺️	☺️	☺️	☺️	☺️	missing	☺️	WHITE SMILING FACE ≊ smiling face	1995	text*	face, outlined, person, relaxed, smile
18	U+1F642	□	🙂	🙂	🙂	🙂	🙂	🙂	🙂	🙂	missing	missing	missing	SLIGHTLY SMILING FACE	2014	emoji	face, person, smile
19	U+1F917	□	🤗	🤗	🤗	🤗	🤗	🤗	missing	missing	missing	missing	missing	HUGGING FACE	2015	emoji	face, hug, hugging, person
20	U+1F607	😇	😇	😇	😇	😇	😇	😇	😇	missing	missing	missing	missing	SMILING FACE WITH HALO	2010	emoji	angel, face, fairy tale, fantasy, halo,
21	U+1F913	□	🤓	🤓	🤓	🤓	🤓	🤓	missing	missing	missing	missing	missing	NERD FACE	2015	emoji	face, geek, nerd, person
22	U+1F914	□	🤔	🤔	🤔	🤔	🤔	🤔	missing	missing	missing	missing	missing	THINKING FACE	2015	emoji	face, person, thinking
23	U+1F610	😐	😐	😐	😐	😐	😐	😐	😐	missing	missing	missing	missing	NEUTRAL FACE	2010	emoji	deadpan, face, neutral, person
24	U+1F611	□	😑	😑	😑	😑	😑	😑	😑	missing	missing	missing	missing	EXPRESSIONLESS FACE	2012	emoji	expressionless, face, inexpressive, p
25	U+1F636	😶	😶	😶	😶	😶	😶	😶	😶	missing	missing	missing	missing	FACE WITHOUT MOUTH	2010	emoji	face, mouth, person, quiet, silent

http://unicode.org/emoji/charts/full-emoji-list.html

Continued

- How much storage space is required for the string
  - Hello hero?

Continued

- How much storage space is required for the string
  - Hello hero?

H	e	l	l	o		h	e	r	o	?
48	65	6C	6C	6F	20	68	65	72	6F	3F

Continued

- How much storage space is required for the string
  - Hello hero?

H	e	l	l	o		h	e	r	o	?
48	65	6C	6C	6F	20	68	65	72	6F	3F

11

Continued

- How much storage space is required for the string
  - Hello hero?

H	e	l	l	o		h	e	r	o	?
48	65	6C	6C	6F	20	68	65	72	6F	3F

  - 澳大利亚

Continued

- How much storage space is required for the string
  - Hello hero?

H	e	l	l	o		h	e	r	o	?
48	65	6C	6C	6F	20	68	65	72	6F	3F

  - 澳大利亚

澳		大		利		亚	
6F	B3	59	27	55	29	4E	9A

Continued

- How much storage space is required for the string
  - Hello hero?

H	e	l	l	o		h	e	r	o	?
48	65	6C	6C	6F	20	68	65	72	6F	3F

  - 澳大利亚

澳		大		利		亚	
6F	B3	59	27	55	29	4E	9A

8

Continued

- How much storage space is required for the string
  - Hello hero?

H	e	l	l	o		h	e	r	o	?
48	65	6C	6C	6F	20	68	65	72	6F	3F

  - 澳大利亚

澳		大		利		亚	
6F	B3	59	27	55	29	4E	9A

  - ☺

1, 2 or 3?

Continued

- How much storage space is required for the string
  - Hello hero?

H	e	l	l	o		h	e	r	o	?
48	65	6C	6C	6F	20	68	65	72	6F	3F

  - 澳大利亚

澳		大		利		亚	
6F	B3	59	27	55	29	4E	9A

  - ☺

01	F6	03

3!

What about images, etc

- Encoded in special format
  - Images: GIF, JPEG, TIFF etc
  - Movies: AVI, MOV etc
  - Documents: DOC, DOCX, PDF etc

What about images, etc

- Often file metadata (eg extensions) tell applications what format the file is in.

- Sometimes the 1st few bytes tells you:

  See FourCC or "Magic Number" in Wikipedia
- Examples:

  "GIF89a" for GIF, "JFIF" for JPEG

  "ID3" for MP3

  "%PDF-1.5" for PDF

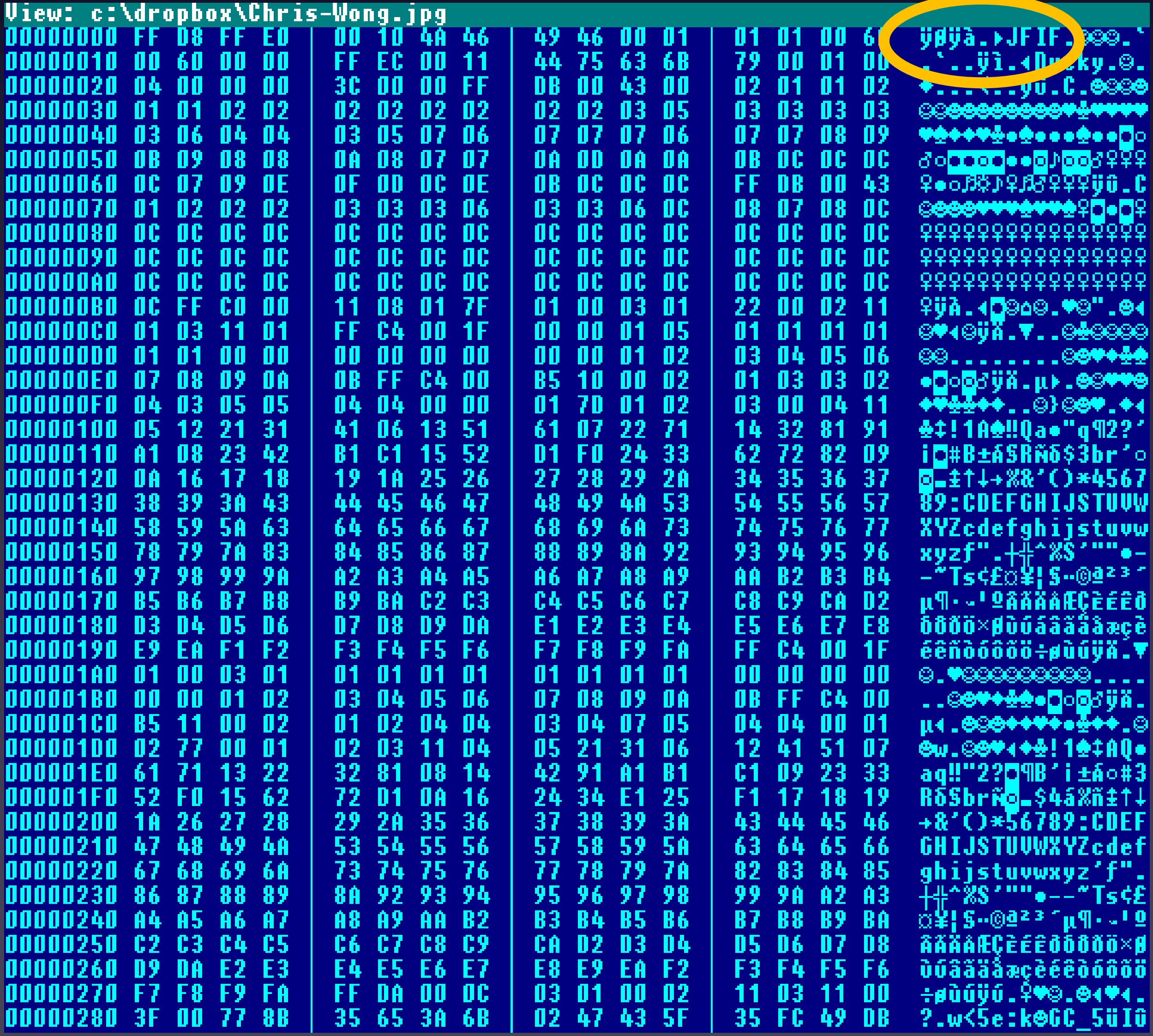
  "#!" for shell scripts

  "PK" for zip files

https://en.wikipedia.org/wiki/Magic_number_(programming)

Example of binary file

- chris-wong.jpg



```
View: c:\dropbox\Chris-Wong.jpg
00000000 FF D8 FF E0 | 00 10 4A 46 | 49 46 00 01 | 01 01 00 60  ÿØÿà.►JFIF.☺☺☺.`
00000010 00 60 00 00 | FF EC 00 11 | 44 75 63 6B | 79 00 01 00  .`..ÿì.◄Ducky.☺.
00000020 04 00 00 00 | 3C 00 00 FF | DB 00 43 00 | 02 01 01 02  ♦...<..ÿÛ.C.☻☺☺☻
00000030 01 01 02 02 | 02 02 02 02 | 02 02 03 05 | 03 03 03 03  ☺☺☻☻☻☻☻☻☻☻♥♣♥♥♥♥
00000040 03 06 04 04 | 03 05 07 06 | 07 07 07 06 | 07 07 08 09  ♥♠♦♦♥♣•♠•••♠••◘○
00000050 0B 09 08 08 | 0A 08 07 07 | 0A 0D 0A 0A | 0B 0C 0C 0C  ♂○◘◘◙◘••◙♪◙◙♂♀♀♀
00000060 0C 07 09 0E | 0F 0D 0C 0E | 0B 0C 0C 0C | FF DB 00 43  ♀•○♫☼♪♀♫♂♀♀♀ÿÛ.C
00000070 01 02 02 02 | 03 03 03 06 | 03 03 06 0C | 08 07 08 0C  ☺☻☻☻♥♥♥♠♥♥♠♀◘•◘♀
00000080 0C 0C 0C 0C | 0C 0C 0C 0C | 0C 0C 0C 0C | 0C 0C 0C 0C  ♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀
00000090 0C 0C 0C 0C | 0C 0C 0C 0C | 0C 0C 0C 0C | 0C 0C 0C 0C  ♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀
000000A0 0C 0C 0C 0C | 0C 0C 0C 0C | 0C 0C 0C 0C | 0C 0C 0C 0C  ♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀
000000B0 0C FF C0 00 | 11 08 01 7F | 01 00 03 01 | 22 00 02 11  ♀ÿÀ.◄◘☺⌂☺.♥☺".☻◄
000000C0 01 03 11 01 | FF C4 00 1F | 00 00 01 05 | 01 01 01 01  ☺♥◄☺ÿÄ.▼..☺♣☺☺☺☺
000000D0 01 01 00 00 | 00 00 00 00 | 00 00 01 02 | 03 04 05 06  ☺☺........☺☻♥♦♣♠
000000E0 07 08 09 0A | 0B FF C4 00 | B5 10 00 02 | 01 03 03 02  •◘○◙♂ÿÄ.µ►.☻☺♥♥☻
000000F0 04 03 05 05 | 04 04 00 00 | 01 7D 01 02 | 03 00 04 11  ♦♥♣♣♦♦..☺}☺☻♥.♦◄
00000100 05 12 21 31 | 41 06 13 51 | 61 07 22 71 | 14 32 81 91  ♣↕!1A♠‼Qa•"q¶2?‘
00000110 A1 08 23 42 | B1 C1 15 52 | D1 F0 24 33 | 62 72 82 09  ¡◘#B±Á§RÑð$3br‚○
00000120 0A 16 17 18 | 19 1A 25 26 | 27 28 29 2A | 34 35 36 37  ◙▬↨↑↓→%&'()*4567
00000130 38 39 3A 43 | 44 45 46 47 | 48 49 4A 53 | 54 55 56 57  89:CDEFGHIJSTUVW
00000140 58 59 5A 63 | 64 65 66 67 | 68 69 6A 73 | 74 75 76 77  XYZcdefghijstuvw
00000150 78 79 7A 83 | 84 85 86 87 | 88 89 8A 92 | 93 94 95 96  xyzƒ„…†‡ˆ‰Š’“”•–
00000160 97 98 99 9A | A2 A3 A4 A5 | A6 A7 A8 A9 | AA B2 B3 B4  —˜™š¢£¤¥¦§¨©ª²³´
00000170 B5 B6 B7 B8 | B9 BA C2 C3 | C4 C5 C6 C7 | C8 C9 CA D2  µ¶·¸¹ºÂÃÄÅÆÇÈÉÊÒ
00000180 D3 D4 D5 D6 | D7 D8 D9 DA | E1 E2 E3 E4 | E5 E6 E7 E8  ÓÔÕÖ×ØÙÚáâãäåæçè
00000190 E9 EA F1 F2 | F3 F4 F5 F6 | F7 F8 F9 FA | FF C4 00 1F  éêñòóôõö÷øùúÿÄ.▼
000001A0 01 00 03 01 | 01 01 01 01 | 01 01 01 01 | 00 00 00 00  ☺.♥☺☺☺☺☺☺☺☺☺....
000001B0 00 00 01 02 | 03 04 05 06 | 07 08 09 0A | 0B FF C4 00  ..☺☻♥♦♣♠•◘○◙♂ÿÄ.
000001C0 B5 11 00 02 | 01 02 04 04 | 03 04 07 05 | 04 04 00 01  µ◄.☻☺☻♦♦♥♦•♣♦♦.☺
000001D0 02 77 00 01 | 02 03 11 04 | 05 21 31 06 | 12 41 51 07  ☻w.☺☻♥◄♦♣!1♠↕AQ•
000001E0 61 71 13 22 | 32 81 08 14 | 42 91 A1 B1 | C1 09 23 33  aq‼"2?◘¶B‘¡±Á○#3
000001F0 52 F0 15 62 | 72 D1 0A 16 | 24 34 E1 25 | F1 17 18 19  Rð§brÑ◙▬$4á%ñ↨↑↓
00000200 1A 26 27 28 | 29 2A 35 36 | 37 38 39 3A | 43 44 45 46  →&'()*56789:CDEF
00000210 47 48 49 4A | 53 54 55 56 | 57 58 59 5A | 63 64 65 66  GHIJSTUVWXYZcdef
00000220 67 68 69 6A | 73 74 75 76 | 77 78 79 7A | 82 83 84 85  ghijstuvwxyz‚ƒ„…
00000230 86 87 88 89 | 8A 92 93 94 | 95 96 97 98 | 99 9A A2 A3  †‡ˆ‰Š’“”•–—˜™š¢£
00000240 A4 A5 A6 A7 | A8 A9 AA B2 | B3 B4 B5 B6 | B7 B8 B9 BA  ¤¥¦§¨©ª²³´µ¶·¸¹º
00000250 C2 C3 C4 C5 | C6 C7 C8 C9 | CA D2 D3 D4 | D5 D6 D7 D8  ÂÃÄÅÆÇÈÉÊÒÓÔÕÖ×Ø
00000260 D9 DA E2 E3 | E4 E5 E6 E7 | E8 E9 EA F2 | F3 F4 F5 F6  ÙÚâãäåæçèéêòóôõö
00000270 F7 F8 F9 FA | FF DA 00 0C | 03 01 00 02 | 11 03 11 00  ÷øùúÿÚ.♀♥☺.☻◄♥◄.
00000280 3F 00 77 8B | 35 65 3A 6B | 02 47 43 5F | 35 FC 49 DB  ?.w‹5e:k☻GC_5üIÛ
```