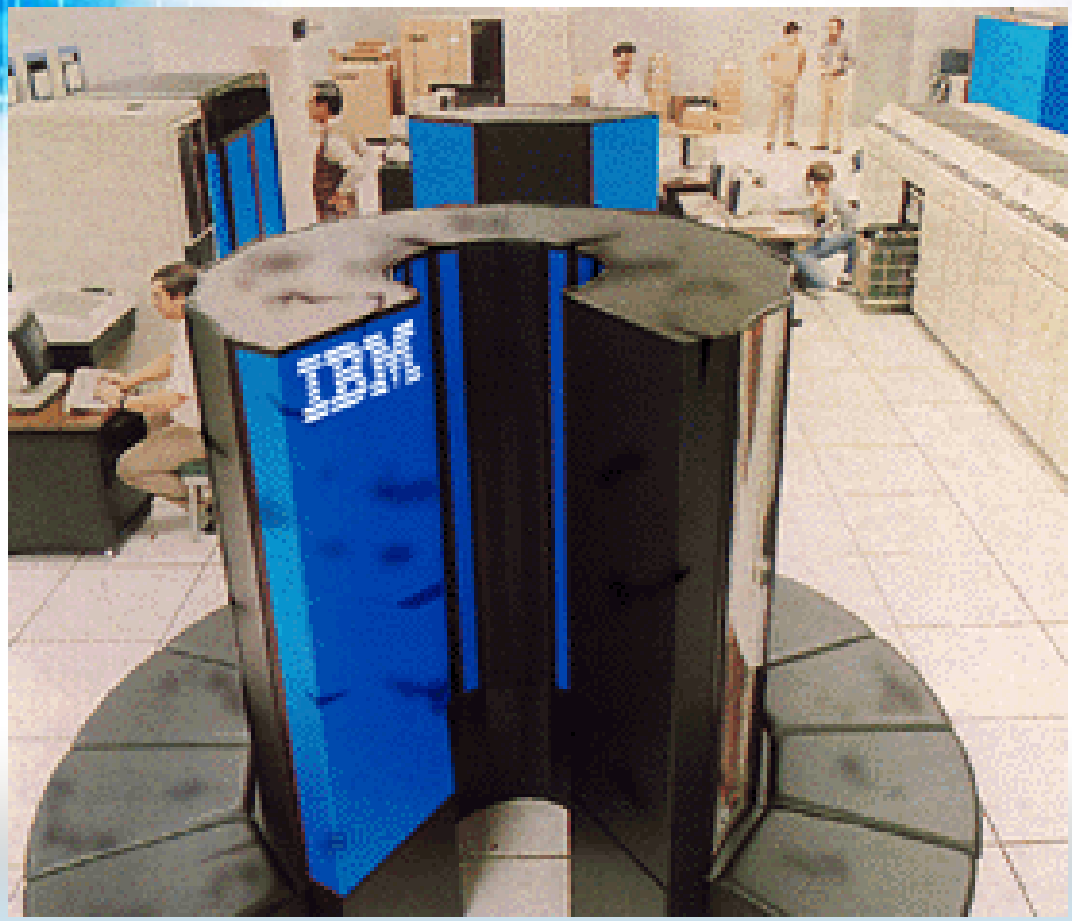
บทที่ 1 โครงสร้างข้อมูลและรหัสของข้อมูล

CT1701 ระบบคอมพิวเตอร์

ประเภทของคอมพิวเตอร์

Super Computer



Main-frame Computer

Mini Computer

Micro Computer / Personal Computer (PC)

Notebook / Laptop

Workstation Computer

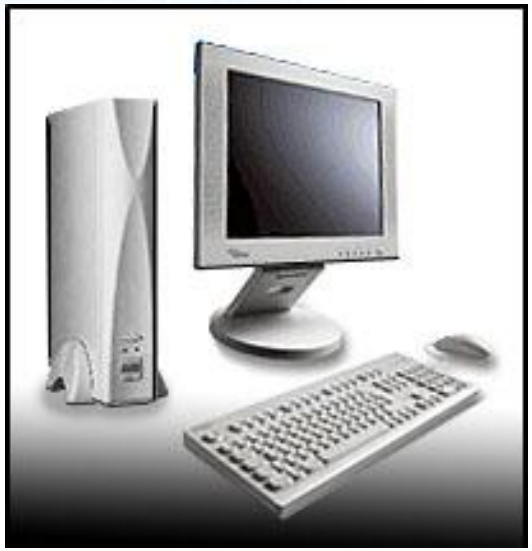
Tablet PC

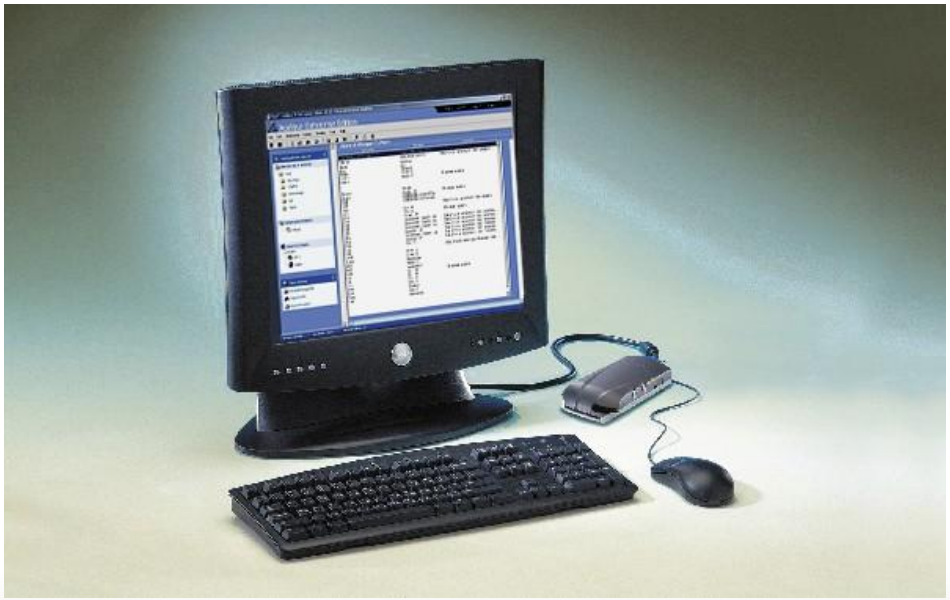
Personal Digital Assistant (PDA)

Pocket PC

Palm

Network Computer (NC) / Thin Client





กำเนิดคอมพิวเตอร์

- คอมพิวเตอร์จะทำงานเกี่ยวกับตัวเลขและการคำนวณ
- การออกแบบระบบคอมพิวเตอร์มาจากการพยายามจำลองรูปแบบการทำงานของตัวเลข และคิดหาหนทางนำมาใช้งานกับข้อมูลต่างๆ ให้ได้
- นักวิชาการคอมพิวเตอร์ได้ค้นพบหลักทฤษฎีของ “เลขฐานสอง (Binary arithmetic)” เพื่อนำมาใช้กับระบบคอมพิวเตอร์

ระบบเลขฐานสอง (Binary)

- ระบบเลขฐานสองประกอบไปด้วยตัวเลขสองตัวคือ 0 และ 1
- ระบบเลขฐานสองเหมาะสำหรับระบบคอมพิวเตอร์ เพราะวงจรไฟฟ้าและอิเล็กทรอนิกส์มีสองสภาวะ คือ การปิด (0) และ การเปิด (1)

การแปลงจากเลขฐานสิบให้เป็นเลขฐานสอง

- 8
- 20
- 55
- 70
- 129

การแปลงจากเลขฐานสองให้เป็นเลขฐานสิบ

- 11
- 101
- 1111
- 11001
- 100111

ระบบเลขฐานสิบหก (Hexadecimal)

- ระบบเลขฐานสิบหกประกอบไปด้วยตัวเลขและตัวอักษร 16 ตัว ได้แก่ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
- การนำเลขฐานสิบหกมาใช้เป็นวิธีการแก้ปัญหาการเขียนข้อมูลลงไปในคอมพิวเตอร์
- ตัวเลขในฐานสิบหก 1 ตัว สามารถใช้แทนตัวเลขฐานสองได้ 4 ตัว
- การแปลงระบบเลขฐานระหว่างฐานสองและฐานสิบหกทำได้ง่าย จึงถูกนำมาใช้ในระบบคอมพิวเตอร์

ระบบเลขฐานสอง ฐานสิบ และฐานสิบหก

ฐานสิบ	ฐานสอง	ฐานสิบหก
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7

ฐานสิบ	ฐานสอง	ฐานสิบหก
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

การแปลงจากเลขฐานสิบให้เป็นเลขฐานสิบหก

- 18
- 54
- 100
- 125
- 158

การแปลงจากเลขฐานสิบหกให้เป็นเลขฐานสิบ

- D
- FF
- 2F2
- FDA
- BC8A

การแปลงจากเลขฐานสองให้เป็นเลขฐานสิบหก

- 1011
- 11001
- 1110001
- 10111111
- 100010001

จงแปลงเลขฐานสิบหกให้เป็นเลขฐานสอง และฐานสิบ

- D
- FF
- 2F2
- FDA
- BC8A

หน่วยนับข้อมูลของระบบคอมพิวเตอร์

- 1 Bit (b)= 0, 1
- 8 Bit (b) = 1 Byte (B)
- 2 Byte = 1 Word (16 Bit)

Binary prefix

Name	Symbol	Quantity
Kilobyte	KB	2^{10} (10^{3})
Megabyte	MB	2^{20} (10^{6})
Gigabyte	GB	2^{30} (10^{9})
Terabyte	TB	2^{40} (10^{12})
Petabyte	PB	2^{50} (10^{15})
Exabyte	EB	2^{60} (10^{18})
Zettabyte	ZB	2^{70} (10^{21})
Yottabyte	YB	2^{80} (10^{24})

จำนวนเลขมาตรฐาน

- คอมพิวเตอร์สามารถทำงานกับเลขจำนวนเต็มได้เลย
- จำนวนเต็มแบ่งออกเป็นสองประเภทคือ จำนวนเต็มขนาด 1 Byte และจำนวนเต็มขนาด 2 Byte หรือ 1 Word
- ตัวเลขขนาด 8 Bit ใช้แทนตัวเลขได้ตั้งแต่ 0 ถึง 255 (จำนวนเต็มบวก) และ –128 ถึง 127 (รวมจำนวนเต็มลบด้วย)
- ตัวเลขขนาด 16 Bit ใช้แทนตัวเลขได้ตั้งแต่ 0 ถึง 65,535 (จำนวนเต็มบวก) และ –32,768 ถึง 32,767 (รวมจำนวนเต็มลบด้วย)

เลขขนาดใหญ่

- ในการใช้งานจริง การใช้ตัวเลขจำนวนเต็มเพียงอย่างเดียวไม่เป็นการเพียงพอ เพราะจะต้องประมวลผลเลขจำนวนจริงด้วย
- คอมพิวเตอร์มีสองวิธีที่จะทำให้ตัวเลขที่จะใช้มีค่ากว้างออกไป
  - การเพิ่มขนาดของตัวเลขจำนวนเต็มให้มากขึ้น เช่น เพิ่มเป็น 4 Byte ก็จะให้ค่ามากกว่าถึงประมาณ 2,000 ล้านค่า
  - แบ่งตัวเลขออกเป็นสองส่วน คือ ส่วนของจำนวนเต็มและส่วนของจุดทศนิยม
- โดยปกติตัวแปรภาษา (Compiler หรือ Interpreter) จะมีรูทีนพิเศษที่ทำหน้าที่ดูแลจัดการเกี่ยวกับตัวเลขเหล่านี้

การทำงานกับข้อความ

- ตัวอักษรแต่ละตัวจะแทนด้วยกลุ่มของ Bit เฉพาะตัวของมันเอง
- การนำตัวอักษรแต่ละตัวมาเรียงต่อกันเป็นข้อความเรียกว่า String
- ขนาดของข้อความไม่มีความแน่นอน ในการใช้งานบางอย่างก็ต้องกำหนดขนาดของข้อความให้แน่นอนไว้เสียก่อน

การกำหนดจุดสิ้นสุดของข้อความ

- เก็บความยาวไว้ที่จุดเริ่มต้นของตัวมันเอง

4This2is1a6string2of5words

- ใส่สัญลักษณ์บางอย่างลงไปตอนท้ายของข้อความ เรียกว่า ตัวแยก (Delimeter)”

This*is*a*string*of*words

กลุ่มอักขระในคอมพิวเตอร์

- ตัวอักษรที่อยู่บนเครื่องคอมพิวเตอร์โดยทั่วไปจะมีขนาด 8 Bit หรือ 1 Byte จึงสามารถมีตัวอักษรต่างกันได้ถึง 256 แบบ
- ประกอบไปด้วยตัวอักษรธรรมดา ตัวเลข สัญลักษณ์ต่าง ๆ รวมทั้งอักขระควบคุม

ASCII Character

- ASCII ย่อมาจาก The American Standard Code for Information Interchange)
- ASCII ขนาด 1 byte จะมี 8 bit ใช้พื้นที่แค่ 7 bit ก็สามารถ Encode ข้อมูลได้ครบแล้ว ส่วน bit ที่ 8 ก็เติม 0 ลงไป (เลขฐานสอง)
- โปรแกรมที่ใช้สำหรับเรียงลำดับข้อมูลโดยทั่วไปจะเรียงตามลำดับที่ปรากฏในตาราง ASCII

ASCII Table

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	&#127;	DEL

Source: www.pubblinet.com

Control Character

- ตัวอักษร 32 ตัวแรกไม่ได้เป็นข้อมูลแต่จะทำหน้าที่เป็นคำสั่งบางอย่าง
  - ควบคุมการทำงานของเครื่องพิมพ์
  - ควบคุมการทำงานของอุปกรณ์คอมพิวเตอร์อื่นๆ
  - ควบคุมการสื่อสารระหว่างเครื่องคอมพิวเตอร์

Control Character

Dec	Hx	Oct	Char	
0	0	000	NUL	(null)
1	1	001	SOH	(start of heading)
2	2	002	STX	(start of text)
3	3	003	ETX	(end of text)
4	4	004	EOT	(end of transmission)
5	5	005	ENQ	(enquiry)
6	6	006	ACK	(acknowledge)
7	7	007	BEL	(bell)
8	8	010	BS	(backspace)
9	9	011	TAB	(horizontal tab)
10	A	012	LF	(NL line feed, new line
11	B	013	VT	(vertical tab)
12	C	014	FF	(NP form feed, new page
13	D	015	CR	(carriage return)
14	E	016	SO	(shift out)
15	F	017	SI	(shift in)
16	10	020	DLE	(data link escape)
17	11	021	DC1	(device control 1)
18	12	022	DC2	(device control 2)
19	13	023	DC3	(device control 3)
20	14	024	DC4	(device control 4)
21	15	025	NAK	(negative acknowledge)
22	16	026	SYN	(synchronous idle)
23	17	027	ETB	(end of trans. block)
24	18	030	CAN	(cancel)
25	19	031	EM	(end of medium)
26	1A	032	SUB	(substitute)
27	1B	033	ESC	(escape)
28	1C	034	FS	(file separator)
29	1D	035	GS	(group separator)
30	1E	036	RS	(record separator)
31	1F	037	US	(unit separator)

Extended ASCII Table

128	Ç	144	É	161	í	177	▒	193	┴	209	╤	225	ß	241	±
129	ü	145	æ	162	ó	178	▓	194	┬	210	╥	226	Γ	242	≥
130	é	146	Æ	163	ú	179	│	195	├	211	╙	227	π	243	≤
131	â	147	ô	164	ñ	180	┤	196	─	212	╘	228	Σ	244	⌠
132	ä	148	ö	165	Ñ	181	╡	197	┼	213	╒	229	σ	245	⌡
133	à	149	ò	166	ª	182	╢	198	╞	214	╓	230	µ	246	÷
134	å	150	û	167	º	183	╖	199	╟	215	╫	231	τ	247	≈
135	ç	151	ù	168	¿	184	╕	200	╚	216	╪	232	Φ	248	°
136	ê	152	ÿ	169	⌐	185	╣	201	╔	217	┘	233	Θ	249	∙
137	ë	153	Ö	170	¬	186	║	202	╩	218	┌	234	Ω	250	·
138	è	154	Ü	171	½	187	╗	203	╦	219	█	235	δ	251	√
139	ï	156	£	172	¼	188	╝	204	╠	220	▄	236	∞	252	ⁿ
140	î	157	¥	173	¡	189	╜	205	═	221	▌	237	φ	253	²
141	ì	158	₧	174	«	190	╛	206	╬	222	▐	238	ε	254	■
142	Ä	159	ƒ	175	»	191	┐	207	╧	223	▀	239	∩	255	
143	Å	160	á	176	░	192	└	208	╨	224	α	240	≡		

www.LookupTables.com

Extended ASCII Table (Thai)

	128	136	144	152	160	168	176	184	192	200	208	216	224	232	240	248
0	□	ฆ	ฌ	ฯ		จ	ฐ	ธ	ภ	ศ	ะ	ุ	เ	็	๐	๘
1	ค	+	‘	°	ก	ฉ	ฑ	น	ม	ษ	ั	ู	แ	่	๑	๙
2	ฅ	ๆ	’	ๅ	ข	ช	ฒ	บ	ย	ส	า	ฺ	โ	้	๒	๚
3	ฆ	'	“	ำ	ฃ	ซ	ณ	ป	ร	ก	ิ	฿	ใ	๊	๓	๛
4	ง	ฯ	”	ฆ	ค	ฌ	ด	ผ	ฤ	พ	ี	เ	ไ	๋	๔	
5	…	ฆ	•		ฅ	ญ	ต	ฝ	ล	อ	ึ	□	ๅ	์	๕	๎
6	'	□	–	□	ฆ	ฏ	ถ	พ	ฦ	ฮ	ื	□	ๆ	๎	๖	๏
7	ฯ	ๆ	—	๎	ง	ฏ	ท	ฟ	ว	ฯ	ุ	฿	็	◉	๗	□

Unicode

- Unicode ต่างจาก ASCII คือ ASCII เก็บ byte เดียว แต่ Unicode เก็บ 2 byte (เก็บตัวอักษรได้ 65,536 ตัวอักษร)
- สามารถเก็บข้อมูลได้มากมายหลายภาษาในโลก
- ดังนั้นรหัสภาษาไทยเอาไปเปิดในภาษาจีน ก็ยังเป็นภาษาไทยอยู่ ไม่ออกมาเป็นภาษาจีน เพราะว่ามี code ตายตัวอยู่ว่า code นี้จองไว้สำหรับภาษาไทย แล้ว code ตรงช่วงนี้เป็นภาษาจีน ตรงนี้เป็นภาษาญี่ปุ่น จะไม่ใช้ที่ซ้ำกัน เป็นต้น
- Unicode จะใช้เลขฐานสิบหกแทนรหัสของตัวอักษรเนื่องจากถ้าใช้เลขฐานสองจะต้องเขียนรหัสยาวมาก (16 ตัว)

Unicode Table (English)

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	NUL 0000	STX 0001	SOT 0002	ETX 0003	EOT 0004	ENQ 0005	ACK 0006	BEL 0007	BS 0008	HT 0009	LF 000A	VT 000B	FF 000C	CR 000D	SO 000E	SI 000F
10	DLE 0010	DC1 0011	DC2 0012	DC3 0013	DC4 0014	NAK 0015	SYN 0016	ETB 0017	CAN 0018	EM 0019	SUB 001A	ESC 001B	FS 001C	GS 001D	RS 001E	US 001F
20	SP 0020	! 0021	" 0022	# 0023	$ 0024	% 0025	& 0026	' 0027	(0028	) 0029	* 002A	+ 002B	, 002C	- 002D	. 002E	/ 002F
30	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
40	@ 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
50	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
60	` 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
70	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	DEL 007F
80	€ 20AC					… 2026										
90		‘ 2018	’ 2019	“ 201C	” 201D	• 2022	– 2013	— 2014								

Unicode Table (Thai)

A0	NBSP 00A0	ก 0E01	ข 0E02	ฃ 0E03	ค 0E04	ฅ 0E05	ฆ 0E06	ง 0E07	จ 0E08	ฉ 0E09	ช 0E0A	ซ 0E0B	ฌ 0E0C	ญ 0E0D	ฎ 0E0E	ฏ 0E0F
B0	ฐ 0E10	ฑ 0E11	ฒ 0E12	ณ 0E13	ด 0E14	ต 0E15	ถ 0E16	ท 0E17	ธ 0E18	น 0E19	บ 0E1A	ป 0E1B	ผ 0E1C	ฝ 0E1D	พ 0E1E	ฟ 0E1F
C0	ภ 0E20	ม 0E21	ย 0E22	ร 0E23	ฤ 0E24	ล 0E25	ฦ 0E26	ว 0E27	ศ 0E28	ษ 0E29	ส 0E2A	ห 0E2B	ฬ 0E2C	อ 0E2D	ฮ 0E2E	ฯ 0E2F
D0	ะ 0E30	ั 0E31	า 0E32	ำ 0E33	ิ 0E34	ี 0E35	ึ 0E36	ื 0E37	ุ 0E38	ู 0E39	ฺ 0E3A					฿ 0E3F
E0	เ 0E40	แ 0E41	โ 0E42	ใ 0E43	ไ 0E44	ๅ 0E45	ๆ 0E46	็ 0E47	่ 0E48	้ 0E49	๊ 0E4A	๋ 0E4B	์ 0E4C	ํ 0E4D	๎ 0E4E	๏ 0E4F
F0	๐ 0E50	๑ 0E51	๒ 0E52	๓ 0E53	๔ 0E54	๕ 0E55	๖ 0E56	๗ 0E57	๘ 0E58	๙ 0E59	๚ 0E5A	๛ 0E5B				