

## APPENDIX E

# UNICODE CHARACTER SET

Unicode, adopted as an international standard in 1992, is a standardized set of characters. It uses a 16-bit code to represent each character. This allows for 65,536 ( $2^{16}$ ) different characters to be represented. Almost all character codes (letters, special characters, and other linguistic symbols) covering most of the languages in the world are included. Unicode even represents the Japanese character set Kanji, despite its thousands of characters.

Unicode is intended to eventually replace its 8-bit predecessor called the ASCII (American Standard Code for Information Interchange) alphanumeric character set. Only 256 characters can be represented by ASCII, which is restricted to representing only our familiar Latin alphabet, standard numbers, and commonly used symbols, such as ?, ,, and !

ASCII is a subset of the Unicode character set. The first 256 characters of the Unicode set are identical to the ASCII set. This appendix includes the first 128 characters from this Unicode subset.

It is easy to convert from the 8-bit ASCII representation to the 16-bit representation of Unicode. Unicode simply leaves each of the 8 leftmost bits to be 0. For example, the character @ is in ASCII represented as 0100 0000, whereas Unicode represents it as 0000 0000 0100 0000.

More information about Unicode can be found at <http://www.unicode.org>.

The first 32 characters of the ASCII character set are control functions; they are not relevant to any of the content in this book but are included here for completeness.

Decimal	Hexadecimal	Character	Control Function
0	00	null	NUL
1	01	😊	SOH
2	02	●	STX
3	03	♥	ETX
4	04	♦	EOT
5	05	♣	ENQ
6	06	♠	ACK
7	07	•	BEL (Bell Sound)
8	08		BS (Back Space)
9	09		TAB
10	0A		LF (Line Feed)
11	0B		VT (Vertical Tab)
12	0C		FF (Form feed)
13	0D		CR (Carriage Return)
14	0E		SO
15	0F	☒	SI
16	10		DLE
17	11		DC1
18	12		DC2
19	13		DC3
20	14		DC4
21	15		NAK
22	16		SYN
23	17		ETB
24	18		CAN
25	19		EM
26	1A		SUB
27	1B		ESC (Escape)
28	1C		FS

Decimal	Hexadecimal	Character	Control Function
29	1D		GS
30	1E		RS
31	1F		US
32	20		
33	21	!	
34	22	"	
35	23	#	
36	24	\$	
37	25	%	
38	26	&	
39	27	'	
40	28	(	
41	29	)	
42	2A	*	
43	2B	+	
44	2C	,	
45	2D	-	
46	2E	.	
47	2F	/	
48	30	0	
49	31	1	
50	32	2	
51	33	3	
52	34	4	
53	35	5	
54	36	6	
55	37	7	
56	38	8	
57	39	9	

Decimal	Hexadecimal	Character	Control Function
58	3A	:	
59	3B	;	
60	3C	<	
61	3D	=	
62	3E	>	
63	3F	?	
64	40	@	
65	41	A	
66	42	B	
67	43	C	
68	44	D	
69	45	E	
70	46	F	
71	47	G	
72	48	H	
73	49	I	
74	4A	J	
75	4B	K	
76	4C	L	
77	4D	M	
78	4E	N	
79	4F	O	
80	50	P	
81	51	Q	
82	52	R	
83	53	S	
84	54	T	
85	55	U	

Decimal	Hexadecimal	Character	Control Function
86	56	V	
87	57	W	
88	58	X	
89	59	Y	
90	5A	Z	
91	5B	[	
92	5C	\	
93	5D	]	
94	5E	^	
95	5F	_	
96	60	`	
97	61	a	
98	62	b	
99	63	c	
100	64	d	
101	65	e	
102	66	f	
103	67	g	
104	68	h	
105	69	i	
106	6A	j	
107	6B	k	
108	6C	l	
109	6D	m	
110	6E	n	
111	6F	o	
112	70	p	
113	71	q	

Decimal	Hexadecimal	Character	Control Function
114	72	r	
115	73	s	
116	74	t	
117	75	u	
118	76	v	
119	77	w	
120	78	x	
121	79	y	
122	7A	z	
123	7B	{	
124	7C		
125	7D	}	
126	7E	~	
127	7F	DEL (Delete)	