

## **C0 Controls and Basic Latin**

### **Range: 0000–007F**

This file contains an excerpt from the character code tables and list of character names for the Unicode Standard, last updated for *The Unicode Standard, Version 4.0*.

This file may be updated as necessary to reflect errata without notice. For an up-to-date list of errata, see <http://www.unicode.org/errata/>

### **Disclaimer**

These charts are provided as the on-line reference to the character contents of the Unicode Standard, Version 4.0 but do not provide all the information needed to fully support individual scripts using the Unicode Standard. For a complete understanding of the use of the characters contained in this excerpt file, please consult the appropriate sections of *The Unicode Standard, Version 4.0* (ISBN 0-321-18578-1), as well as Unicode Standard Annexes #9, #11, #14, #15, #24 and #29, the other Unicode Technical Reports and the Unicode Character Database, which are available on-line.

See <http://www.unicode.org/Public/UNIDATA/UCD.html> and <http://www.unicode.org/reports/>

A thorough understanding of the information contained in these additional sources is required for a successful implementation.

### **Fonts**

The shapes of the reference glyphs used in these code charts are not prescriptive. Considerable variation is to be expected in actual fonts. The particular fonts used in these charts were provided to the Unicode Consortium by a number of different font designers, who own the rights to the fonts.

See <http://www.unicode.org/charts/fonts.html> for a list.

### **Terms of Use**

You may freely use these code charts for personal or internal business uses only. You may not incorporate them either wholly or in part into any product or publication, or otherwise distribute them without express written permission from the Unicode Consortium. However, you are welcome to provide links to these charts.

The fonts and font data used in production of these Code Charts may NOT be extracted or otherwise used in any commercial product without permission or license granted by the typeface owner(s).

The information in this file may be updated from time to time. The Unicode Consortium is not liable for errors or omissions in this excerpt file or the standard itself. Information on characters added to the Unicode Standard since the publication of Version 4.0 as well as on characters currently being considered for addition to the Unicode Standard can be found on the Unicode web site.



























See <http://www.unicode.org/pending/pending.html> and <http://www.unicode.org/alloc/Pipeline.html>.








Copyright © 1991-2003 Unicode, Inc. All rights reserved.

	000	001	002	003	004	005	006	007
0	NUL 0000	DLE 0010	SP 0020	0 0030	@ 0040	P 0050	` 0060	p 0070
1	SOH 0001	DC1 0011	! 0021	1 0031	A 0041	Q 0051	a 0061	q 0071
2	STX 0002	DC2 0012	" 0022	2 0032	B 0042	R 0052	b 0062	r 0072
3	ETX 0003	DC3 0013	# 0023	3 0033	C 0043	S 0053	c 0063	s 0073
4	EOT 0004	DC4 0014	\$ 0024	4 0034	D 0044	T 0054	d 0064	t 0074
5	ENQ 0005	NAK 0015	% 0025	5 0035	E 0045	U 0055	e 0065	u 0075
6	ACK 0006	SYN 0016	& 0026	6 0036	F 0046	V 0056	f 0066	v 0076
7	BEL 0007	ETB 0017	' 0027	7 0037	G 0047	W 0057	g 0067	w 0077
8	BS 0008	CAN 0018	( 0028	8 0038	H 0048	X 0058	h 0068	x 0078
9	HT 0009	EM 0019	) 0029	9 0039	I 0049	Y 0059	i 0069	y 0079
A	LF 000A	SUB 001A	* 002A	: 003A	J 004A	Z 005A	j 006A	z 007A
B	VT 000B	ESC 001B	+ 002B	; 003B	K 004B	[ 005B	k 006B	{ 007B
C	FF 000C	FS 001C	, 002C	< 003C	L 004C	\ 005C	l 006C	 007C
D	CR 000D	GS 001D	- 002D	= 003D	M 004D	] 005D	m 006D	} 007D
E	SO 000E	RS 001E	. 002E	> 003E	N 004E	^ 005E	n 006E	~ 007E
F	SI 000F	US 001F	/ 002F	? 003F	O 004F	_ 005F	o 006F	DEL 007F

**C0 controls**









*Alias names are those for ISO/IEC 6429:1992.  
Commonly used alternative aliases are also shown.*

0000		<control> = NULL
0001		<control> = START OF HEADING
0002		<control> = START OF TEXT
0003		<control> = END OF TEXT
0004		<control> = END OF TRANSMISSION
0005		<control> = ENQUIRY
0006		<control> = ACKNOWLEDGE
0007		<control> = BELL
0008		<control> = BACKSPACE
0009		<control> = CHARACTER TABULATION = horizontal tabulation (HT), tab
000A		<control> = LINE FEED (LF) = new line (NL), end of line (EOL)
000B		<control> = LINE TABULATION
000C		<control> = FORM FEED (FF)
000D		<control> = CARRIAGE RETURN (CR)
000E		<control> = SHIFT OUT
000F		<control> = SHIFT IN
0010		<control> = DATA LINK ESCAPE
0011		<control> = DEVICE CONTROL ONE
0012		<control> = DEVICE CONTROL TWO
0013		<control> = DEVICE CONTROL THREE
0014		<control> = DEVICE CONTROL FOUR
0015		<control> = NEGATIVE ACKNOWLEDGE
0016		<control> = SYNCHRONOUS IDLE
0017		<control> = END OF TRANSMISSION BLOCK
0018		<control> = CANCEL
0019		<control> = END OF MEDIUM

001A		<control> = SUBSTITUTE → FFFD  replacement character
001B		<control> = ESCAPE
001C		<control> = INFORMATION SEPARATOR FOUR = file separator (FS)
001D		<control> = INFORMATION SEPARATOR THREE = group separator (GS)
001E		<control> = INFORMATION SEPARATOR TWO = record separator (RS)
001F		<control> = INFORMATION SEPARATOR ONE = unit separator (US)

**ASCII punctuation and symbols**

*Based on ISO/IEC 646 from here.*

0020		SPACE <ul style="list-style-type: none"> <li>• sometimes considered a control code</li> <li>• other space characters: 2000  –200A </li> <li>→ 00A0  no-break space</li> <li>→ 200B  zero width space</li> <li>→ 2060  word joiner</li> <li>→ 3000  ideographic space</li> <li>→ FEFF  zero width no-break space</li> </ul>
0021	!	EXCLAMATION MARK <ul style="list-style-type: none"> <li>= factorial</li> <li>= bang</li> <li>→ 00A1 ¡ inverted exclamation mark</li> <li>→ 01C3 ! latin letter retroflex click</li> <li>→ 203C !! double exclamation mark</li> <li>→ 203D ? interrobang</li> <li>→ 2762 ● heavy exclamation mark ornament</li> </ul>
0022	"	QUOTATION MARK <ul style="list-style-type: none"> <li>• neutral (vertical), used as opening or closing quotation mark</li> <li>• preferred characters in English for paired quotation marks are 201C “ &amp; 201D ”</li> <li>→ 02BA " modifier letter double prime</li> <li>→ 030B ⸀ combining double acute accent</li> <li>→ 030E ⸂ combining double vertical line above</li> <li>→ 2033 " double prime</li> <li>→ 3003 " ditto mark</li> </ul>
0023	#	NUMBER SIGN <ul style="list-style-type: none"> <li>= pound sign, hash, crosshatch, octothorpe</li> <li>→ 266F # music sharp sign</li> </ul>
0024	\$	DOLLAR SIGN <ul style="list-style-type: none"> <li>= milreis, escudo</li> <li>• glyph may have one or two vertical bars</li> <li>• other currency symbol characters: 20A0 ₣ –20AF ₤</li> <li>→ 00A4 Ɱ currency sign</li> </ul>

0025	%	PERCENT SIGN
		→ 066A $\%$ arabic percent sign
		→ 2030 ‰ per mille sign
		→ 2031 ‰ per ten thousand sign
		→ 2052 $\%$ commercial minus sign
0026	&	AMPERSAND
0027	'	APOSTROPHE
		= APOSTROPHE-QUOTE
		= APL quote
		• neutral (vertical) glyph with mixed usage
		• 2019 ' is preferred for apostrophe
		• preferred characters in English for paired quotation marks are 2018 ‘ & 2019 ’
		→ 02B9 ' modifier letter prime
		→ 02BC ' modifier letter apostrophe
		→ 02C8 ' modifier letter vertical line
		→ 0301 ◌ combining acute accent
		→ 2032 ' prime
0028	(	LEFT PARENTHESIS
		= OPENING PARENTHESIS
0029	)	RIGHT PARENTHESIS
		= CLOSING PARENTHESIS
		• see discussion on semantics of paired bracketing characters
002A	*	ASTERISK
		= star (on phone keypads)
		→ 066D * arabic five pointed star
		→ 204E * low asterisk
		→ 2217 * asterisk operator
		→ 2731 * heavy asterisk
002B	+	PLUS SIGN
002C	,	COMMA
		= decimal separator
		→ 060C ، arabic comma
		→ 201A , single low-9 quotation mark
		→ 3001 、 ideographic comma
002D	-	HYPHEN-MINUS
		= hyphen or minus sign
		• used for either hyphen or minus sign
		→ 2010 - hyphen
		→ 2011 ⎵ non-breaking hyphen
		→ 2012 – figure dash
		→ 2013 – en dash
		→ 2212 – minus sign
002E	.	FULL STOP
		= PERIOD
		= dot, decimal point
		• may be rendered as a raised decimal point in old style numbers
		→ 06D4 - arabic full stop
		→ 3002 。 ideographic full stop
002F	/	SOLIDUS
		= SLASH
		= virgule, shilling (British)
		→ 01C0   latin letter dental click
		→ 0338 ∕ combining long solidus overlay
		→ 2044 / fraction slash
		→ 2215 / division slash

**ASCII digits**

0030	0	DIGIT ZERO
0031	1	DIGIT ONE
0032	2	DIGIT TWO
0033	3	DIGIT THREE
0034	4	DIGIT FOUR
0035	5	DIGIT FIVE
0036	6	DIGIT SIX
0037	7	DIGIT SEVEN
0038	8	DIGIT EIGHT
0039	9	DIGIT NINE

**ASCII punctuation and symbols**

003A	:	COLON
		→ 0589 ; armenian full stop
		→ 05C3 : hebrew punctuation sof pasuq
		→ 2236 : ratio
003B	;	SEMICOLON
		• this, and not 037E ; , is the preferred character for 'Greek question mark'
		→ 037E ; greek question mark
		→ 061B ؛ arabic semicolon
		→ 204F ; reversed semicolon
003C	<	LESS-THAN SIGN
		→ 2039 ‹ single left-pointing angle quotation mark
		→ 2329 ‹ left-pointing angle bracket
		→ 27E8 ‹ mathematical left angle bracket
		→ 3008 ‹ left angle bracket
003D	=	EQUALS SIGN
		• other related characters: 2241 ≠ –2263 ≡
		→ 2260 ≠ not equal to
		→ 2261 ≡ identical to
003E	>	GREATER-THAN SIGN
		→ 203A › single right-pointing angle quotation mark
		→ 232A › right-pointing angle bracket
		→ 27E9 › mathematical right angle bracket
		→ 3009 › right angle bracket
003F	?	QUESTION MARK
		→ 00BF ¿ inverted question mark
		→ 037E ; greek question mark
		→ 061F ؟ arabic question mark
		→ 203D ? interrobang
		→ 2048 ?! question exclamation mark
		→ 2049 !? exclamation question mark
0040	@	COMMERCIAL AT

**Uppercase Latin alphabet**

0041	A	LATIN CAPITAL LETTER A
0042	B	LATIN CAPITAL LETTER B
		→ 212C ℬ script capital b
0043	C	LATIN CAPITAL LETTER C
		→ 2102 Ɔ double-struck capital c
		→ 212D Ɔ black-letter capital c
0044	D	LATIN CAPITAL LETTER D

0045	E	LATIN CAPITAL LETTER E	→ 2107 $\mathcal{E}$ euler constant → 2130 $\mathcal{E}$ script capital e
0046	F	LATIN CAPITAL LETTER F	→ 2131 $\mathcal{F}$ script capital f → 2132 $\mathcal{F}$ turned capital f
0047	G	LATIN CAPITAL LETTER G	• invented circa 300 BCE by Spurius Carvilius Ruga, who added a stroke to the letter C
0048	H	LATIN CAPITAL LETTER H	→ 210B $\mathcal{H}$ script capital h → 210C $\mathcal{H}$ black-letter capital h → 210D $\mathcal{H}$ double-struck capital h
0049	I	LATIN CAPITAL LETTER I	• Turkish and Azerbaijani use 0131 $\mathcal{I}$ for lowercase → 0130 $\mathcal{I}$ latin capital letter i with dot above → 0406 $\mathcal{I}$ cyrillic capital letter byelorussian-ukrainian i → 04C0 $\mathcal{I}$ cyrillic letter palochka → 2110 $\mathcal{I}$ script capital i → 2111 $\mathcal{I}$ black-letter capital i → 2160 $\mathcal{I}$ roman numeral one
004A	J	LATIN CAPITAL LETTER J	
004B	K	LATIN CAPITAL LETTER K	→ 212A $\mathcal{K}$ kelvin sign
004C	L	LATIN CAPITAL LETTER L	→ 2112 $\mathcal{L}$ script capital l
004D	M	LATIN CAPITAL LETTER M	→ 2133 $\mathcal{M}$ script capital m
004E	N	LATIN CAPITAL LETTER N	→ 2115 $\mathcal{N}$ double-struck capital n
004F	O	LATIN CAPITAL LETTER O	
0050	P	LATIN CAPITAL LETTER P	→ 2119 $\mathcal{P}$ double-struck capital p
0051	Q	LATIN CAPITAL LETTER Q	→ 211A $\mathcal{Q}$ double-struck capital q
0052	R	LATIN CAPITAL LETTER R	→ 211B $\mathcal{R}$ script capital r → 211C $\mathcal{R}$ black-letter capital r → 211D $\mathcal{R}$ double-struck capital r
0053	S	LATIN CAPITAL LETTER S	
0054	T	LATIN CAPITAL LETTER T	
0055	U	LATIN CAPITAL LETTER U	
0056	V	LATIN CAPITAL LETTER V	
0057	W	LATIN CAPITAL LETTER W	
0058	X	LATIN CAPITAL LETTER X	
0059	Y	LATIN CAPITAL LETTER Y	
005A	Z	LATIN CAPITAL LETTER Z	→ 2124 $\mathcal{Z}$ double-struck capital z → 2128 $\mathcal{Z}$ black-letter capital z

**ASCII punctuation and symbols**

005B	[	LEFT SQUARE BRACKET	= OPENING SQUARE BRACKET	• other bracket characters: 3008 $\langle$ –301B $\rangle$
005C	\	REVERSE SOLIDUS	= BACKSLASH	→ 20E5 $\backslash$ combining reverse solidus overlay → 2216 $\backslash$ set minus
005D	]	RIGHT SQUARE BRACKET	= CLOSING SQUARE BRACKET	
005E	^	CIRCUMFLEX ACCENT	• this is a spacing character	→ 02C4 $\hat{}$ modifier letter up arrowhead → 02C6 $\hat{}$ modifier letter circumflex accent → 0302 $\hat{}$ combining circumflex accent → 2303 $\hat{}$ up arrowhead
005F	_	LOW LINE	= SPACING UNDERSCORE	• this is a spacing character
				→ 02CD $\_$ modifier letter low macron → 0331 $\_$ combining macron below → 0332 $\_$ combining low line → 2017 $\_$ double low line
0060	`	GRAVE ACCENT	• this is a spacing character	→ 02CB $\`$ modifier letter grave accent → 0300 $\`$ combining grave accent → 2035 $\`$ reversed prime

**Lowercase Latin alphabet**


0061	a	LATIN SMALL LETTER A	
0062	b	LATIN SMALL LETTER B	
0063	c	LATIN SMALL LETTER C	
0064	d	LATIN SMALL LETTER D	
0065	e	LATIN SMALL LETTER E	→ 212E $e$ estimated symbol → 212F $e$ script small e
0066	f	LATIN SMALL LETTER F	
0067	g	LATIN SMALL LETTER G	→ 0261 $g$ latin small letter script g → 210A $g$ script small g
0068	h	LATIN SMALL LETTER H	→ 04BB $h$ cyrillic small letter shha → 210E $h$ planck constant
0069	i	LATIN SMALL LETTER I	• Turkish and Azerbaijani use 0130 $\mathcal{I}$ for uppercase → 0131 $\mathcal{I}$ latin small letter dotless i
006A	j	LATIN SMALL LETTER J	
006B	k	LATIN SMALL LETTER K	
006C	l	LATIN SMALL LETTER L	→ 2113 $\mathcal{l}$ script small l
006D	m	LATIN SMALL LETTER M	
006E	n	LATIN SMALL LETTER N	→ 207F $^n$ superscript latin small letter n

006F	o	LATIN SMALL LETTER O → 2134 <i>ø</i> script small o
0070	p	LATIN SMALL LETTER P
0071	q	LATIN SMALL LETTER Q
0072	r	LATIN SMALL LETTER R
0073	s	LATIN SMALL LETTER S
0074	t	LATIN SMALL LETTER T
0075	u	LATIN SMALL LETTER U
0076	v	LATIN SMALL LETTER V
0077	w	LATIN SMALL LETTER W
0078	x	LATIN SMALL LETTER X
0079	y	LATIN SMALL LETTER Y
007A	z	LATIN SMALL LETTER Z → 01B6 <i>z</i> latin small letter z with stroke

### ASCII punctuation and symbols

007B	{	LEFT CURLY BRACKET = OPENING CURLY BRACKET = opening brace
007C		VERTICAL LINE = VERTICAL BAR • used in pairs to indicate absolute value → 01C0   latin letter dental click → 05C0   hebrew punctuation paseq → 2223   divides → 2758   light vertical bar
007D	}	RIGHT CURLY BRACKET = CLOSING CURLY BRACKET = closing brace
007E	~	TILDE • this is a spacing character → 02DC ~ small tilde → 0303 ¨ combining tilde → 223C ~ tilde operator → FF5E ~ fullwidth tilde

### Control character

007F		<control> = DELETE
------	---	-----------------------