

Characters and Glyphs

Our alphabet is not just 26 caps, 26 lowercase letters, figures, and punctuation; there are many special characters, plus thousands more throughout the world's lan-

guages. These are available to computer users through option keys and expert fonts. Shown below are some commonly used ones.

The Unicode Consortium,¹ in concert with the International Standards Organization (ISO), is defining standards for a wide variety of applications, to allow for

common usage of data across computer platforms, industries, and languages. While the space here is too limited to convey the magnitude of the effort, the following definitions have been extracted from its work.

1. A **character** conveys distinctions in meaning or sounds.

A character is a concept, not the shape or appearance of an image on paper or video screen. For example, to represent the unit of information meaning "one," ISO/IEC 10646-1 encodes a large number of characters that could represent this basic "unit" (opposite).

' acute	â circumflex a	ð eth, Icelandic	« left angle quote, guillemet
Á acute A	Ê circumflex E	Ð eth, Icelandic	{ left curly brace
á acute a	ê circumflex e	! exclamation mark	< less than
É acute E	Î circumflex I	^a feminine ordinal	a-z lowercase letters
é acute e	î circumflex i	fi f i ligature	- macron
í acute i	Ô circumflex O	fl f l ligature	° masculine ordinal
Ó acute O	ô circumflex o	f florin	µ mu
ó acute o	Û circumflex U	` grave	· middle dot
Ú acute U	û circumflex u	À grave A	× multiply
ú acute u	: colon	à grave a	¬ not
Ý acute Y	† dagger	È grave E	# number
ý acute y	, comma	è grave e	Ω omega
& ampersand	@ commercial at	Ì grave I	½ one-half
' apostrophe	© copyright	ì grave i	¼ one-quarter
* asterisk	° degree	Ò grave O	¶ paragraph
\ backslash	Δ delta	ò grave o	(parenthesis left
(reverse solidus)	0-9 digits	Û grave U) parenthesis right
broken vertical bar	Æ diphthong ligature	ù grave u	% percent
A-Z capitals	æ diphthong ligature	> greater than	‰ per mil (parts per thousand)
^ caret	÷ division	horizontal tab	· period
, cedilla	\$ dollar	- hyphen	π pi
Ç cedilla C	‡ double dagger	" inches	Π pi
ç cedilla c	... ellipsis	∞ infinity	+ plus
¢ cent	- en dash	¡ inverted exclamation mark	± plus or minus
^ circumflex	— em dash	¿ inverted question mark	
Â circumflex A	= equals		

2. A **glyph** conveys distinctions in form or appearance and is a recognizable abstract graphic symbol which is independent of any specific design, *i.e.*, the letter “A,” whether it is bold, italic, Helvetica, or Times Roman.

3. A **glyph shape** is an image of a glyph, as displayed on a pre-

sentation surface. This defines the visual differences between Helvetica and Times Roman.

4. **Glyph metrics** are the set of information in a glyph representation used for defining the dimensions and positioning of the glyph shape (character spacing, kerning, etc.). Other aspects

of the problem that a designer may encounter include the relationship between coded characters and glyph identifiers; they may be one-to-one, one-to-many, many-to-one, or many-to-many. Spell checking, sorting, etc. present special problems, as does presentation

direction: a text stream of Arabic written from right to left could contain a Latin quote written from left to right.

1. *The Unicode Standard, Version 2.0, The Unicode Consortium* (Reading, Massachusetts: Addison-Wesley Developers Press, 1991-96).

' prime	Ð thorn, Icelandic
? question mark	¾ three-fourths
” quotation mark, closing	~ tilde
“ quotation mark, opening	ã tilde a
® registered trademark	Ã tilde A
» right angle quote, guillemet	Ñ tilde N
} right curly brace	ñ tilde n
] right square bracket	Õ tilde O
Å ring A	õ tilde o
å ring a	¨ umlaut
§ section	Ä umlaut A
; semi-colon	ä umlaut a
Ø slashed O	Ë umlaut E
ø slashed o	ë umlaut e
ß German eszett (sz)	Ï umlaut I
– soft hyphen	ï umlaut i
/ solidus (slash), space	Ö umlaut O
[square bracket left	ö umlaut o
£ sterling	Ü umlaut u
√ square root	ü umlaut u
¹ superscript one	ÿ umlaut Y
² superscript two	– underscore
	vertical bar (caesura)
	¥ yen

Consider for a moment the case with the unit of information meaning “one.” ISO not only encodes a large number of characters that conceivably “represent” this “unit of information,” but also encodes a number of “characters” that represent a particular form associated with this meaning. The characters that may be said to “represent” the “unit of information” designated by “one” are (at least):

0031	Digit One “1”
00B9	Superscript One “ ¹ ”
0661	Arabic-Indic Digit One “١”
06F1	Extended Arabic-Indic Digit One “ا١”
0967	Devanagari Digit One “१”
09E7	Bengali Digit One “১”
09f4	Bengali Currency Numerator One “৳”
0A67	Gurmukhi Digit One “ੴ”
0AE7	Gujarati Digit One “૧”
0b67	Oriya Digit One “୧”
0BE7	Tamil Digit One “௧”
0c67	Telugu Digit One “౧”
0CE7	Kannada Digit One “೧”
0d67	Malayalam Digit One “൧”
0E51	Thai Digit One “๑”
0ed1	Lao Digit One “໑”
2081	Subscript One “ ₁ ”
215f	Fraction Numerator One “ ¹ ”
2160	Roman Numeral One “I”
2170	Small Roman Numeral One “i”
2460	Circled Digit One “①”
2474	Parenthesized Digit One “(1)”
2488	Digit One Full Stop “1.”
2776	Dingbat Negative Circled Digit One “Ⓜ”
2780	Dingbat Circled Sans-Serif Digit One “Ⓢ”
278a	Dingbat Negative Circled Sans-Serif Digit One “Ⓜ”
3021	Hangzhou Numeral One “一”
3192	Ideographic Annotation One Mark “一”
3220	Parenthesized Ideograph One “(一)”
3280	Circled Ideograph One “⊖”
4E00	Cjk Unified Ideograph-4E00 “一”
58f9	Cjk Unified Ideograph-58f9 “壹”
FF11	Fullwidth Digit One “1”

The example shown above is excerpted from the ISO/IEC 10646-1, and demonstrates the complexity of one aspect of the font problem; assigning a numeric

code to a character. There are 23 additional characters not shown. (From Unicode Standard, Version 2.0.)