

Character Set Support

Developing comprehensive fonts can be quite a challenge. There are over 340 characters in the Arabic Unicode blocks (not including the presentations forms). For many of those characters, there are up to 4 different forms. An Arabic script font can require many, many glyphs.

Not counting presentation forms, the four Unicode blocks that include Arabic script characters are:

- [Arabic \(U+0600..U+06FF\)](#) was added to Unicode 1.0.
- [Arabic Supplement \(U+0750..U+077F\)](#) was added to Unicode 4.1
- [Arabic Extended-A \(U+08A0..U+08FF\)](#) was added to Unicode 6.1
- [Arabic Extended-B \(U+0870..U+089F\)](#) was added to Unicode 14.0
- [Arabic Extended-C \(U+10EC0..U+10EFF\)](#) was added to Unicode 15.0

There are still some Arabic script characters being added to Unicode, so it is possible that not all languages using the Arabic script are fully represented in Unicode.

For optimal compatibility with a variety of operating systems, all Non-Roman fonts should also include a set of glyphs for basic Roman characters and punctuation. Ideally this should include all the following characters, although some depend on other considerations (see the notes). The basis for this list is a union of the Windows Codepage 1252 and MacRoman character sets plus additional useful characters.

- [Recommended characters for Non-Roman fonts](#)

In addition, there are other characters which are specifically recommended for Arabic script.

These are characters often required in regular Arabic text (not necessarily mixed script). Because these are normally used in Arabic text, these characters should be designed to match the design style of the Arabic script characters listed above.

0020	SPACE
0021	EXCLAMATION MARK
0022	QUOTATION MARK
0026	AMPERSAND
0027	APOSTROPHE
0028	LEFT PARENTHESIS
0029	RIGHT PARENTHESIS
002B	PLUS SIGN
002C	COMMA
002D	HYPHEN-MINUS
002E	FULL STOP
002F	SOLIDUS
0030	DIGIT ZERO
0031	DIGIT ONE
0032	DIGIT TWO
0033	DIGIT THREE
0034	DIGIT FOUR
0035	DIGIT FIVE
0036	DIGIT SIX
0037	DIGIT SEVEN
0038	DIGIT EIGHT
0039	DIGIT NINE
003A	COLON
003B	SEMICOLON
003C	LESS-THAN SIGN
003D	EQUALS SIGN
003E	GREATER-THAN SIGN
003F	QUESTION MARK
0040	COMMERCIAL AT
005B	LEFT SQUARE BRACKET
005C	REVERSE SOLIDUS
005D	RIGHT SQUARE BRACKET
007B	LEFT CURLY BRACKET
007C	VERTICAL LINE
007D	RIGHT CURLY BRACKET
00AB	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
00BB	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
00D7	MULTIPLICATION SIGN

00F7	DIVISION SIGN
2000	EN QUAD
2001	EM QUAD
2002	EN SPACE
2003	EM SPACE
2004	THREE-PER-EM SPACE
2005	FOUR-PER-EM SPACE
2006	SIX-PER-EM SPACE
2007	FIGURE SPACE
2008	PUNCTUATION SPACE
2009	THIN SPACE
200A	HAIR SPACE
200B	ZERO WIDTH SPACE
200C	ZERO WIDTH NON-JOINER
200D	ZERO WIDTH JOINER
200F	RIGHT-TO-LEFT MARK
2018	LEFT SINGLE QUOTATION MARK
2019	RIGHT SINGLE QUOTATION MARK
201C	LEFT DOUBLE QUOTATION MARK
201D	RIGHT DOUBLE QUOTATION MARK
2027	HYPHENATION POINT
2028	LINE SEPARATOR
2029	PARAGRAPH SEPARATOR
202A	LEFT-TO-RIGHT EMBEDDING
202B	RIGHT-TO-LEFT EMBEDDING
202C	POP DIRECTIONAL FORMATTING
202D	LEFT-TO-RIGHT OVERRIDE
202E	RIGHT-TO-LEFT OVERRIDE
202F	NARROW NO-BREAK SPACE
2039	SINGLE LEFT-POINTING ANGLE QUOTATION MARK
203A	SINGLE RIGHT-POINTING ANGLE QUOTATION MARK
204F	REVERSED SEMICOLON
2060	WORD JOINER
2066	LEFT-TO-RIGHT ISOLATE
2067	RIGHT-TO-LEFT ISOLATE
2068	FIRST STRONG ISOLATE
2069	POP DIRECTIONAL ISOLATE
206C	INHIBIT ARABIC FORM SHAPING

USV	Name
206D	ACTIVATE ARABIC FORM SHAPING
2212	MINUS SIGN
2219	BULLET OPERATOR
25CC	DOTTED CIRCLE
2E41	REVERSED COMMA

This guide is from the [font-arab-tools project](#) and is copyright © 2022-2023 SIL International.