Title: Proposed changes to the UTR #50 character data
Source: Microsoft Corporation (Worachai Chaoweeraprasit, Stephen Chua, Dylan Deverill, Hiroshi Harada, Laurențiu Iancu, Matthew Isabel, Kiyohiro Maeda, Dwayne Robinson, Stuart Stuple)
Status: Individual contribution
Action: For consideration by the Unicode Technical Committee
Date: 2013-01-31

1. Abstract

Since the introduction of the *Proposed Draft UTR #50 Unicode Properties for Horizontal and Vertical Text Layout* [1] in 2011, the corresponding character data underwent several revisions. The data files corresponding to revision 8 of the draft were posted recently [2]. This proposal presents a number of suggested changes to that version of the character data. The proposed changes include the conclusions of the 2012-10-25 teleconference, which were not incorporated in revision 8 of the data files. A scoping of tailoring is also suggested.

2. Proposed character data changes

Table 1 summarizes the characters with MVO property values proposed for revision. Most of the sample glyphs in the second column of the table are rendered with the Windows 8 fonts Meiryo UI and Segoe UI Symbol [3]. The MVO property values in the fourth and fifth columns are those assigned as of revision 8 of UTR #50 [2] and, respectively, being proposed. The sixth column marks certain characters for which tailoring seems inevitable, due to a more diverse orientation encountered in practice.

Table 1: Proposed changes to MVO property values

Code point(s)	Glyph(s)	Character name(s)	MVO in [2]	Prop. MVO	Tailor	Notes
00A9	©	COPYRIGHT SIGN	U	R	Yes	Consistently with 00AE ®
00AE	®	REGISTERED SIGN	U	R	Yes	Consistently with 00A9 ©
00B6	¶	PILCROW SIGN	U	R		204B ⁋ is R
00BC 00BD 00BE	¼ ½ ¾	VULGAR FRACTION ONE QUARTER VULGAR FRACTION ONE HALF VULGAR FRACTION THREE QUARTERS	U	R		Consistently with other fractions and fraction slash 2044 ⁄
02E5 ..02E9	˥˦˧˨˩	MODIFIER LETTER EXTRA-HIGH TONE BAR ..MODIFIER LETTER EXTRA-LOW TONE BAR	U	R		Agreed U → R in 2012-10-25 teleconference
2018	'	LEFT SINGLE QUOTATION MARK	T	U	Yes	
2019	'	RIGHT SINGLE QUOTATION MARK	T	U	Yes	

201C	"	LEFT DOUBLE QUOTATION MARK	R	U	Yes	
201D	"	RIGHT DOUBLE QUOTATION MARK	R	U	Yes	
203D	‽	INTERROBANG	U	R		2E18 ⸘ is R
2044	/	FRACTION SLASH	U	R		Changed R → U in 2012-06-06 teleconference
2102 2102 210A 210B 210C 210D 210E 2110 2111 2112 2115 2119 211A 211B 211C 211D 2124 2128 212C 212D 212F 2130 2131 2133 2134	ℂ 𝑔 ℋ ℌ ℍ ℎ ℐ ℑ ℒ ℕ ℙ ℚ ℛ ℜ ℝ ℤ ℨ ℬ ℭ ℯ ℰ ℱ ℳ ℴ	DOUBLE-STRUCK CAPITAL C SCRIPT SMALL G SCRIPT CAPITAL H BLACK-LETTER CAPITAL H DOUBLE-STRUCK CAPITAL H PLANCK CONSTANT SCRIPT CAPITAL I BLACK-LETTER CAPITAL I SCRIPT CAPITAL L DOUBLE-STRUCK CAPITAL N DOUBLE-STRUCK CAPITAL P DOUBLE-STRUCK CAPITAL Q SCRIPT CAPITAL R BLACK-LETTER CAPITAL R DOUBLE-STRUCK CAPITAL R DOUBLE-STRUCK CAPITAL Z BLACK-LETTER CAPITAL Z SCRIPT CAPITAL B BLACK-LETTER CAPITAL C SCRIPT SMALL E SCRIPT CAPITAL E SCRIPT CAPITAL F SCRIPT CAPITAL M SCRIPT SMALL O	U	R		Same as the math alphanumeric symbols that they complete (Additional candidates for changing U → R are present in the Letterlike Symbols block)
2150 ..215F	⅐..⅟	VULGAR FRACTION ONE SEVENTH ..FRACTION NUMERATOR ONE	U	R		Consistently with ¼ etc.
2189	↉	VULGAR FRACTION ZERO THIRDS	U	R		Consistently with ¼ etc.
2322 2323	⌢ ⌣	FROWN SMILE	U	R		Same as the ties 203F ‿ and 2040 ⁀
2329 232A	〈 〉	LEFT-POINTING ANGLE BRACKET RIGHT-POINTING ANGLE BRACKET	U	R		Consistently with other brackets
23B4 23B5 23B6	⎴ ⎵ ⎶	TOP SQUARE BRACKET BOTTOM SQUARE BRACKET BOTTOM SQUARE BRACKET OVER TOP SQUARE BRACKET	U	R	Yes	Terminal emulation: U Stretching delimiters like 23DC ⏜..23E1 ⏡: R
23BA ..23BD	⎺⎻⎼	HORIZONTAL SCAN LINE-1 ..HORIZONTAL SCAN LINE-9	U	R		Same as box drawing characters
23CE	⏎	RETURN SYMBOL	U	R		
2423	␣	OPEN BOX	U	R		Agreed U → R in 2012-10-25 teleconference

10980 .. 1099F	[glyph] .. [glyph]	MEROITIC HIEROGLYPHIC LETTER A ..MEROITIC HIEROGLYPHIC SYMBOL VIDJ-2	R	U		Same as Egyptian hieroglyphs
F0000 .. 10FFFF		PRIVATE USE CHARACTER-F0000 ETC.	-	U		Supplementary PUA-A, B, including noncharacters

3. Tailoring

As of revision 8, UTR #50 defines tailoring in a loose manner that allows overriding the MVO property values of the entire Unicode codespace, which is also a consequence of the conformance requirements associated with an informative property. However, in practice, implementations are likely to tailor only a subset of characters. Although in principle any character may be subject to tailoring, in practice certain characters have clear orientation, with no need for tailoring, while others are inherently ambiguous, and tailoring is strictly necessary for those. Examples of the former kind are ideographs, which appear upright in all realistic usage scenarios, and examples of the latter kind include U+00AE REGISTERED SIGN, which is known to be found in either orientation in Japanese vertical layout and therefore cannot be accommodated by either single orientation.

A lax definition of tailoring makes the default MVO property values, and the property itself, arguably less useful. Partitioning the codespace into non-tailorable and tailorable subsets can better express, at the plain text level, the orientation semantics for characters without or, respectively, with variance in orientation, and therefore scope the applicability of overriding accordingly.

Possible approaches of defining a partitioning of the codespace into characters with non-tailorable vs. tailorable orientation include the following:

- Introduce a binary property that identifies the characters of either set
- Enumerate a list of characters of the set which can be described in a more concise way

An example of a candidate set of characters with non-tailorable orientation, i.e., non-tailorable MVO property values, may be the set of characters whose Ideographic property value is Yes, which accounts for 75,633 characters as of Unicode Version 6.2. An example of a candidate set of characters with tailorable orientation would be the list of arrows given in Table 3 of UTR #50 [1].

The adequate approach can be determined after validating the merits of the suggested scoping of tailoring and partitioning of the codespace.

4. References

[1] Koji Ishii, *Proposed Draft UTR #50 Unicode Properties for Horizontal and Vertical Text Layout*, Revision 8, http://www.unicode.org/reports/tr50/tr50-8.html, October 2012. Earlier revisions authored by Eric Muller.

[2] Eric Muller et al., *UTR #50 character data files*, Revision 8, http://www.unicode.org/reports/tr50/ tr50-8.Orientation.txt and tr50-8.Orientation.html, posted January 2013.

[3] Microsoft Corporation, *Microsoft Typography – Fonts and Products – Font Families*, http://www.microsoft.com/typography/fonts/family.aspx.