L2/14-274

Title: **Proposed Math-Class Assignments for UTR #25 Revision 14**
Source: Laurențiu Iancu and Murray Sargent III – Microsoft Corporation
Status: Individual contribution
Action: For consideration by the Unicode Technical Committee
Date: 2014-10-24

1. Introduction

Revision 13 of UTR #25 [UTR25], published in April 2012, corresponds to Unicode Version 6.1 [TUS61]. As of October 2014, Revision 14 is in preparation, to update UTR #25 and its data files to the character repertoire of Unicode Version 7.0 [TUS70]. This document compiles a list of characters proposed to be assigned mathematical classes in Revision 14 of UTR #25.

In this document, the term *math-class* is being used to refer to the character classification in UTR #25. While functionally similar to a UCD character property, math-class is applicable only within the scope of UTR #25. Math-class is different from the UCD binary character property Math [UAX44]. The relation between Math and math-class is that the set of characters with the property Math=Yes is a proper subset of the set of characters assigned any math-class value in UTR #25.

As of Revision 13, the set relation between Math and math-class is invalidated by the collection of Arabic mathematical alphabetic symbols in the range U+1EE00 – U+1EEFF. This is a known issue [14-052], already discussed by the UTC [138-C12 in 14-026]. Once those symbols are added to the UTR #25 data files, the set relation will be restored.

This document proposes only UTR #25 math-class values, and not any UCD Math property values.

2. Character selection criteria

The characters being proposed for addition to the UTR #25 data files were selected by the following criteria:

- Arrows and geometric shapes that constitute sets with characters of the same kind which are already assigned math-class values
- Any newly encoded characters that have clear mathematical usage
- Characters affected by corrections of mapping errors

Unicode 7.0 added a large number of arrows and geometric shapes, but those symbols originated from the Wingdings and Webdings sets, so they are not necessarily suitable for mathematical usage. Only a small number of Unicode 7.0 symbols were selected, viz. those that constitute sets with characters which are assigned math-class values in Revision 13 of UTR #25.

3. Proposed characters

The characters proposed for addition to the UTR #25 data files (not including the Arabic mathematical alphabetic symbols) and their math-class values are listed in Table 1. The characters can be grouped as follows:

1. U+2717 BALLOT X, instead of U+2612 BALLOT BOX WITH X, to correct the mapping to the named ISOPUB entity *cross* [ISOPUB].
2. Geometric shapes introduced in Unicode Version 5.1 [TUS51], which are documented in Table 2.5 of UTR #25, but are not assigned math-class values. These characters, highlighted in green in Table 1, clearly constitute sets with characters that do have math-class values and, by that criterion, qualify for addition.
3. U+2B45 LEFTWARDS QUADRUPLE ARROW and U+2B46 RIGHTWARDS QUADRUPLE ARROW, also from Unicode 5.1, which form a set with U+27F0 UPWARDS QUADRUPLE ARROW and U+27F1 DOWNWARDS QUADRUPLE ARROW, already assigned.
4. U+2B95 RIGHTWARDS BLACK ARROW, which constitutes a set with U+2B05 LEFTWARDS BLACK ARROW through U+2B07 DOWNWARDS BLACK ARROW, already assigned.
5. U+2BC5 BLACK MEDIUM UP-POINTING TRIANGLE CENTRED through U+2BC8 BLACK MEDIUM RIGHT-POINTING TRIANGLE CENTRED. Although explicitly named CENTRED, they are otherwise generic triangles that fill the four open slots for medium-size black triangles in Table 2.5 of UTR #25. In terms of centering, it is expected for all geometric shapes to be centered in mathematical contexts, so the explicit character names should not prevent their addition. These characters are highlighted in purple in Table 1.
6. Polygons U+2BC2 TURNED BLACK PENTAGON through U+2BC4 BLACK OCTAGON, which follow the pattern of other polygons already assigned. These are highlighted in blue in Table 1.
7. U+2BCA TOP HALF BLACK CIRCLE and U+2BCB BOTTOM HALF BLACK CIRCLE, which constitute a set with U+25D6 LEFT HALF BLACK CIRCLE and U+25D7 RIGHT HALF BLACK CIRCLE, already assigned.
8. Symbols in the Geometric Shapes Extended block, which constitute sets with similar characters already assigned math-class values. These are also highlighted in blue in Table 1.

The symbols highlighted in Table 1 constitute sets that are better visualized in the manner of Table 2.5 of UTR #25. In the same manner, Tables 2 and 3 illustrate the filled (black) and, respectively, hollow (white) geometric shapes. The highlight color palette is the same as the one used in Table 1.

4. Characters not being proposed

None of the other arrows and geometric shapes introduced in Unicode 7.0 are being proposed. Examples include triangle-headed arrows, cusp shapes, and all of the symbols in the Geometric Shapes Extended block, U+1F780 – U+1F7FF. Some of those characters may have weakly mathematical usage. If evidence of such usage is found, they may qualify for addition to a future revision of UTR #25.

The Pahawh Hmong arithmetic symbols U+16B3C PAHAWH HMONG SIGN XYEEM NTXIV through U+16B3F PAHAWH HMONG SIGN XYEEM FAIB are also not being proposed, for the reasons provided in [14-060].

Table 1. Characters proposed for assignment of math-class values

Character code point and name		Age	Math-class
2612	BALLOT BOX WITH X	1.1	(blank)
2717	BALLOT X	1.1	N
2B1B	BLACK LARGE SQUARE	5.1	N
2B1C	WHITE LARGE SQUARE	5.1	N
2B1D	BLACK VERY SMALL SQUARE	5.1	N
2B1E	WHITE VERY SMALL SQUARE	5.1	N
2B1F	BLACK PENTAGON	5.1	N
2B24	BLACK LARGE CIRCLE	5.1	N
2B25	BLACK MEDIUM DIAMOND	5.1	N
2B26	WHITE MEDIUM DIAMOND	5.1	N
2B27	BLACK MEDIUM LOZENGE	5.1	N
2B28	WHITE MEDIUM LOZENGE	5.1	N
2B29	BLACK SMALL DIAMOND	5.1	N
2B2A	BLACK SMALL LOZENGE	5.1	N
2B2B	WHITE SMALL LOZENGE	5.1	N
2B2C	BLACK HORIZONTAL ELLIPSE	5.1	N
2B2D	WHITE HORIZONTAL ELLIPSE	5.1	N
2B2E	BLACK VERTICAL ELLIPSE	5.1	N
2B2F	WHITE VERTICAL ELLIPSE	5.1	N
2B45	LEFTWARDS QUADRUPLE ARROW	5.1	R
2B46	RIGHTWARDS QUADRUPLE ARROW	5.1	R
2B50	WHITE MEDIUM STAR	5.1	N
2B51	BLACK SMALL STAR	5.1	N
2B52	WHITE SMALL STAR	5.1	N
2B53	BLACK RIGHT-POINTING PENTAGON	5.1	N
2B54	WHITE RIGHT-POINTING PENTAGON	5.1	N
2B95	RIGHTWARDS BLACK ARROW	7.0	R
2BC2	TURNED BLACK PENTAGON	7.0	N
2BC3	HORIZONTAL BLACK OCTAGON	7.0	N
2BC4	BLACK OCTAGON	7.0	N
2BC5	BLACK MEDIUM UP-POINTING TRIANGLE CENTRED	7.0	N
2BC6	BLACK MEDIUM DOWN-POINTING TRIANGLE CENTRED	7.0	N
2BC7	BLACK MEDIUM LEFT-POINTING TRIANGLE CENTRED	7.0	N
2BC8	BLACK MEDIUM RIGHT-POINTING TRIANGLE CENTRED	7.0	N
2BCA	TOP HALF BLACK CIRCLE	7.0	N
2BCB	BOTTOM HALF BLACK CIRCLE	7.0	N
1F784	BLACK SLIGHTLY SMALL CIRCLE	7.0	N
1F78C	BLACK TINY SQUARE	7.0	N
1F78D	BLACK SLIGHTLY SMALL SQUARE	7.0	N
1F797	BLACK TINY DIAMOND	7.0	N
1F798	BLACK VERY SMALL DIAMOND	7.0	N
1F799	BLACK MEDIUM SMALL DIAMOND	7.0	N
1F79D	BLACK TINY LOZENGE	7.0	N
1F79E	BLACK VERY SMALL LOZENGE	7.0	N
1F79F	BLACK MEDIUM SMALL LOZENGE	7.0	N

Table 2. Simple filled (black) shapes

Table cells show code point, math-class value, and character age (Unicode version).
Entries 2605, 2B25, and 2B27 are shifted w.r.t. Table 2.5 of UTR #25 Revision 13.

Filled shape	Tiny	Very small	Slightly small	Small	Medium small	Medium	Regular	Large
Triangle left ◀				25C2 B 1.1		2BC7 N 7.0	25C0 B 1.1	
Triangle right ▶				25B8 B 1.1		2BC8 N 7.0	25B6 B 1.1	
Triangle up ▲				25B4 B 1.1		2BC5 N 7.0	25B2 B 1.1	
Triangle down ▼				25BE B 1.1		2BC6 N 7.0	25BC B 1.1	
Square ■	1F78C N 7.0	2B1D N 5.1	1F78D N 7.0	25AA N 1.1	25FE B 3.2	25FC B 3.2	25A0 N 1.1	2B1B N 5.1
Diamond ◆	1F797 N 7.0	1F798 N 7.0		2B29 N 5.1	1F799 N 7.0	2B25 N 5.1	25C6 N 1.1	
Lozenge ⧫	1F79D N 7.0	1F79E N 7.0		2B2A N 5.1	1F79F N 7.0	2B27 N 5.1	29EB B 3.2	
Pentagon up ⬟							2B1F N 5.1	
Pentagon right ⬟							2B53 N 5.1	
Pentagon down ⬟							2BC2 N 7.0	
Hexagon horizontal ⬢							2B23 N 5.0	
Hexagon vertical ⬢							2B22 N 5.0	
Octagon horizontal ⯃							2BC3 N 7.0	
Octagon vertical ⯃							2BC4 N 7.0	
Arabic star ★				2B51 N 5.1	22C6 B 1.1		2605 B 1.1	
Ellipse horizontal ⬬							2B2C N 5.1	
Ellipse vertical ⬮							2B2E N 5.1	
Circle ●	22C5 B 1.1	2219 B 1.1	1F784 N 7.0	2022 B 1.1	2981 N 3.2	26AB N 4.1	25CF N 1.1	2B24 N 5.1

Table 3. Simple hollow (white) shapes

Table cells show code point, math-class value, and character age (Unicode version).
Entries 2606, 2B26, 2B28, and 2B50 are shifted w.r.t. Table 2.5 of UTR #25 Revision 13.

Hollow shape	Tiny	Very small	Slightly small	Small	Medium small	Medium	Regular	Large
Triangle left ◁				25C3 B 1.1			25C1 B 1.1	
Triangle right ▷				25B9 B 1.1			25B7 B 1.1	
Triangle up △				25B5 B 1.1			25B3 B 1.1	
Triangle down ▽				25BF B 1.1			25BD B 1.1	
Square □		2B1E N 5.1		25AB N 1.1	25FD B 3.2	25FB B 3.2	25A1 N 1.1	2B1C N 5.1
Diamond ◇				22C4 B 1.1		2B26 N 5.1	25C7 N 1.1	
Lozenge ◊				2B2B N 5.1		2B28 N 5.1	25CA B 1.1	
Pentagon up ⬠							2B20 N 5.0	
Pentagon right ⬠							2B54 N 5.1	
Pentagon down								
Hexagon horizontal ⬡							2394 N 3.0	
Hexagon vertical ⬡							2B21 N 5.0	
Octagon horizontal								
Octagon vertical								
Arabic star ☆				2B52 N 5.1		2B50 N 5.1	2606 B 1.1	
Ellipse horizontal ⬭							2B2D N 5.1	
Ellipse vertical ⬯							2B2F N 5.1	
Circle ○		2218 B 1.1		25E6 B 1.1	26AC N 4.1	26AA N 4.1	25CB B 1.1	25EF N 1.1

5. References

[ISOPUB] David Carlisle and Patrick Ion, ISOPUB entity set at http://www.w3.org/TR/xml-entity-names/isopub.html defined in the W3C Recommendation *XML Entity Definitions for Characters*, 2nd Edition, April 2014, http://www.w3.org/TR/2014/REC-xml-entity-names-20140410/.

[TUS51] The Unicode Consortium, *The Unicode Standard, Version 5.1.0*, April 2008, http://www.unicode.org/versions/Unicode5.1.0/.

[TUS61] The Unicode Consortium, *The Unicode Standard, Version 6.1.0*, January 2012, http://www.unicode.org/versions/Unicode6.1.0/.

[TUS70] The Unicode Consortium, *The Unicode Standard, Version 7.0.0*, June 2014, http://www.unicode.org/versions/Unicode7.0.0/.

[UAX44] Mark Davis, Laurențiu Iancu, and Ken Whistler, *Unicode Standard Annex #44, Unicode Character Database, Version 7.0.0*, June 2014, http://www.unicode.org/reports/tr44/tr44-14.html.

[UTR25] Barbara Beeton, Asmus Freytag, and Murray Sargent III, *Unicode Technical Report #25, Unicode Support for Mathematics, Revision 13*, April 2012, http://www.unicode.org/reports/tr25/tr25-13.pdf, and corresponding data files at http://www.unicode.org/Public/math/revision-13/.

[14-026] The Unicode Technical Committee, *Minutes of UTC Meeting #138*, L2/14-026, February 2014, http://www.unicode.org/L2/L2014/14026.htm.

[14-052] Laurențiu Iancu and Murray Sargent III, *Proposal to assign math-class property values to the Arabic mathematical alphabetic symbols*, L2/14-052, January 2014, http://www.unicode.org/L2/L2014/14052-math-class-arabic.pdf.

[14-060] Laurențiu Iancu and Ken Whistler, *Proposed properties for incorrectly or insufficiently documented characters in the Unicode 7.0 repertoire*, L2/14-060, February 2014, http://www.unicode.org/L2/L2014/14060-u70-properties.pdf.