L2/06-034

0.   introduction

These ruminations are based on based on UTR 25 (Unicode and Mathematics), revision 7, section 2.11 (Geometric Shapes) in particular.

The conclusions are --

(i)   a plea for more consistency when naming characters;

(ii)   a request that characters based on abstract shapes should have some indicator of the size of the character relative to the rest of the font;

(iii)   that the code chart entry for U+2023 TRIANGULAR BULLET be amended to show
   "= 25B8 black right-pointing small triangle"
instead of
   "→ 25B8 black right-pointing small triangle";

(iv)   that the reference to U+2588 FULL BLOCK should be removed from Table 2.4 of UTR 25;

(v)   that the paragraph "*Actual Sizes*", in §2.11 Geometric Shapes, be amended to read:

   The shapes shown here for a given size all have the same area. The precise sizes and shapes chosen, however, are a matter for the font designer, provided only that the order of relative sizes is adhered to, and that shapes of identical size should, in some sense, have equal "visual impact".

(vi)   that, in the interests of consistency and completeness, the following characters be added to the set of abstract shapes:

LARGE BLACK CIRCLE
BLACK PENTAGON
LARGE BLACK SQUARE
LARGE WHITE SQUARE
SMALL BLACK DIAMOND
SMALL WHITE DIAMOND
SMALL BLACK LOZENGE
SMALL WHITE LOZENGE
SMALL WHITE ARABIC STAR

1.   circles

Table 2.4 of UTR 25 shows a nicely graduated set of circles:

tiny	very small		small		medium small		medium		normal		large
•	•	◦	●	○	●	○	●	○	●	○	○
22c5	2219 00b7	2218	2022	25e6	2981	26ac	26ab	26aa	25cf	25cb	25ef

The principal requirement of such a set is that they should be strictly increasing in size, but there are other requirements, implicit and explicit.

A striking thing about this set is that most of them are described as "small" or even "tiny". An initial attempt to make the diameters of the circles follow an arithmetic progression proved unusable. The diameters of those shown above follow a geometric progression that has the following convenient properties:

the "tiny" circle has a radius roughly equal to the stem width;

"medium" circles have diameters roughly equal to the x-height, and the white circle will enclose most mathematical operators comfortably, e.g:

⊕   U+2295 CIRCLED PLUS

"normal" sized circles are roughly equal in height to the caps-height and lowercase ascenders, and the white circle will enclose most mathematical operators with a little white space showing, which is convenient for N-ary operations, e.g:

⊕   U+2A01 N-ARY CIRCLED PLUS OPERATOR

the "large" circle is sufficiently large to enclose uppercase M.

That is fine for a font with a relatively large x-height, but other fonts will probably have to follow something very close to the above relative sizes, mainly because the shapes themselves are used elsewhere, to enclose other items, including circles, thus —

tiny	very small		small		medium small		medium		normal		large
●	●	○	●	○	●	○	●	○	●	○	○
22c5	2219 00b7	2218	2022	25e6	2981	26ac	26ab	26aa	25cf	25cb	25ef
⊙				◎							
2299				233e							
⊙	⊙	⊙	⊙	⊙	●	◎	●	◎			
2a00	29bf	229a		29be	25c9	25ca					

This set appears to work well for circled circles. If the bottom row is to be a graduated set, as described in the code tables, the enclosing circle has to be the "normal" circle. In addition, U+2299 CIRCLED DOT OPERATOR and U+2A00 N-ARY CIRCLED DOT OPERATOR are distinguished as mentioned above.

U+233E APL FUNCTIONAL SYMBOL CIRCLE JOT is something of a misfit. Most APL fonts stick very closely to the letter shapes originally used on 2741 golf-ball terminals. On that basis, the ISO APL standard stipulates U+25CB WHITE CIRCLE for the APL circle, and U+2218 RING OPERATOR for the APL jot. It is unlikely that the standard makers would have chosen those codepoints had it been clear at the time what their relative sizes were. The glyph shown above is closer in intent to the appearance of the original APL circle-jot (or "navel" sign), than to its parentage.

2. squares

Table 2.4 also shows a graduated set of squares, completed here by the inclusion of white squares:

■	◻	■	◻	■	◻	■	◻	■
25aa	25ab	25fe	25fd	25fc	25fb	25a0	25a1	

In Table 2.4, U+2588 is used for the large black square. Firstly, U+2588 is a connecting character. It will require overshoot on all sides to provide that connection, and prodigious amounts of overshoot if the character is to bridge the leading introduced by Word's "single" spacing, for instance. Secondly, U+2588 will not necessarily be square: its shape will depend on the body height of the font and the width chosen for Block Elements. U+2588 is governed by different rules. There is no codepoint yet for the glyph

shown above for the large black square, but (a) it is square, and (b) it has positive left- and right-bearings, and is therefore not a connecting character.

According to UTR 25, "the MEDIUM SMALL SQUARE should be used together with the MEDIUM size of the other basic shapes, and correspondingly for the other sizes".

So we would expect to see the following glyphs with a "similar" size:

> U+25FE   BLACK MEDIUM SMALL SQUARE
> U+25FD   WHITE MEDIUM SMALL SQUARE
> U+26AB   MEDIUM BLACK CIRCLE
> U+26AA   MEDIUM WHITE CIRCLE

which would mean the table of glyphs aligns thus —

tiny	very small		small		medium small		medium		normal		large	
					■ 25aa	□ 25ab	■ 25fe	□ 25fd	■ 25fc	□ 25fb	■ 25a0	□ 25a1
• 22c5	• 2219 (00b7)	◦ 2218	● 2022	○ 25e6	● 2981	○ 26ac	● 26ab	○ 26aa	● 25cf	○ 25cb		○ 25ef

but . . .

3.   triangles

Part of the justification for the above alignment is that "the basic shapes from the Zapf Dingbats font match the unmarked size for triangle, diamond, and circle, and the MEDIUM size for the square". That's fine for Zapf Dingbats, but a little inconvenient for our purposes.

And there are other considerations. The code chart shows U+2022 BULLET as a "black small circle". Table 2.4 shows both U+2022 and U+25E6 WHITE BULLET as "small" circles, and U+25AA BLACK SMALL SQUARE as a small shape of similar size.

Other small shapes are

> U+25C2   BLACK LEFT-POINTING SMALL TRIANGLE
> U+25C3   WHITE LEFT-POINTING SMALL TRIANGLE
> U+25B8   BLACK RIGHT-POINTING SMALL TRIANGLE
> U+25B9   WHITE RIGHT-POINTING SMALL TRIANGLE
> U+25B4   BLACK UP-POINTING SMALL TRIANGLE

U+25B5   WHITE UP-POINTING SMALL TRIANGLE
U+25BE   BLACK DOWN-POINTING SMALL TRIANGLE
U+25BF   WHITE DOWN-POINTING SMALL TRIANGLE

If we make the rule that all "small" shapes should be of similar size, the circles, squares and triangles align thus —

tiny	very small		small (bullets)		medium small		medium		normal		large
			◀ 25c2	◁ 25c3			◀ 25c0	◁ 25c1			
			▶ 25b8	▷ 25b9			▶ 25b6	▷ 25b7			
			▲ 25b4	△ 25b5			▲ 25b2	△ 25b3			
			▼ 25be	▽ 25bf			▼ 25bc	▽ 25bd			
			■ 25aa	◻ 25ab	■ 25fe	◻ 25fd	■ 25fc	◻ 25fb	■ 25a0	◻ 25a1	
• 22c5	• 2219 00b7	◦ 2218	● 2022	○ 25e6	● 2981	○ 26ac	● 26ab	○ 26aa	● 25cf	○ 25cb	◯ 25ef

We can now designate all "small" shapes as bullets, and *vice versa*. (This is a little presumptuous. Although U+2022 BULLET is shown as "= black small circle", no such guide is given for U+2023 TRIANGULAR BULLET, which is simply shown as "→ 25B8 black right-pointing small triangle". It assumed that this is an oversight.)

And since the table includes both "medium" and "medium small" squares, users are free to choose which size best fits their purposes, knowing that all "medium" shapes are of similar size.

In addition to their use as bullets, triangles are also used in mathematical notation, selection functions being particularly popular. These triangles are all vertically centred on the same horizontal line as most other mathematical symbols. The results look slightly odd for up- and down-pointing triangles, but experiments aligning the centre of mass on the said horizontal looked even worse.

U+2206 INCREMENT and U+2207 NABLA should presumably have similar sizes to other symbols, but may show elements of the parent

font, such as stem widths, contrast and stress, and may therefore be deemed to be outside the remit of this section of UTR 25.

U+2A64 Z-NOTATION DOMAIN ANTIRESTRICTION and U+2A65 Z-NOTATION RANGE ANTIRESTRICTION look fine when based on "medium" sized triangles.

2A1E LARGE LEFT TRIANGLE OPERATOR is a curiosity. A glyph that is truly "large" in the sense of Table 2.4 is going to dominate any expression it appears in, and look ugly. Perhaps some explanation could be included in the relevant code chart?

4.   diamonds and lozenges

For the purposes of this document, a diamond is a square, rotated 45° about it centre point.

U+22C4 DIAMOND OPERATOR is shown as a lozenge in Table 2.4, but is here shown as a diamond.

Also, since it is described as being an operator, it shown as a "medium-small" character, so that it is closer in size to the arithmetical signs of plus, minus, times, etc.

For the purposes of this document, a lozenge is composed of two equilateral triangles. In the absence of any specific indicators from the code charts, U+29EB BLACK LOZENGE and U+25CA LOZENGE default to "medium" sized shapes.

There are no "small" lozenges, so far. Diamonds and lozenges therefore align with the circles and squares thus —

tiny	very small		small		medium small		medium		normal		large	
			■ 25aa	◫ 25ab	■ 25fe	◻ 25fd	■ 25fc	◻ 25fb	■ 25a0	◻ 25a1		
						◇ 22c4	◆ 25c6	◇ 25c7				
							◆ 29eb	◇ 25ca				
• 22c5	• 2219 00b7	◦ 2218	● 2022	◦ 25e6	● 2981	○ 26ac	● 26ab	○ 26aa	● 25cf	○ 25cb		○ 25ef

5.   polygons

If *n* points are placed equidistantly on the circumference of circle, and adjacent points connected by straight lines, the result is an inscribed regular polygon. As *n* increases, the inscribed polygon tends towards a circle.

For *n*=1 and *n*=2, the inscribed polygons are degenerate cases. We have covered *n*=3 and *n*=4. For *n*=5, Unicode v4 has a white hexagon at U+2394 SOFTWARE-FUNCTION SYMBOL, while v5 proposes to add

> U+2B20   WHITE PENTAGON
> U+2B21   WHITE HEXAGON
> U+2B22   BLACK HEXAGON
> U+2B23   HORIZONTAL BLACK HEXAGON

Using a default size of "medium" for these characters is not a good idea, because they are so difficult to distinguish from circles at small display sizes. The problem is alleviated a little if "normal" is made the default size for convex polygons of order greater than 4.

(It should be remarked that, while the white vertical hexagon shown in the final table may be OK as a mathematical symbol, chemists (including mathematical chemists) are likely to prefer something much larger, and perhaps a little lighter, to represent a benzene ring.)

6.   stars

If *n* points are placed equidistantly on the circumference of circle, and alternate points connected by straight lines, the result is an inscribed star. As *n* increases, the inscribed star tends towards a circle.

These shapes are still polygons, and regular polygons at that – stars are our first examples of non-convex regular polygons.

For *n*=1, *n*=2 and *n*=4, the inscribed stars are degenerate cases; for *n*=3, the inscribed star is a triangle with negative area – a concept perhaps best explored elsewhere.

For *n*=5, the result is known as an "Arabic Star". The two characters U+2605 BLACK STAR and U+2606 WHITE STAR look consistent with (i.e, have the same "visual impact" as) the other shapes, when drawn at "medium" size.

The character U+22C6 STAR OPERATOR is, like the diamond operator, shown as a "medium-small" glyph.

The sample glyph shown in the code charts for U+066D ARABIC FIVE-POINTED STAR is small enough to function as a bullet, and is shown as such in the final table.

For n=6, the result is known as the "Star of David". The only such star currently included is in the Dingbats block, and this is not an appropriate point to consider the more general class of star shapes found therein. U+2721 STAR OF DAVID will presumably be drawn in a manner consistent with other members of that block, and as such may not be appropriate for mathematical work.

7.   the final table

The Final Table shows the set of shapes currently available, plus those demanded by consistency and completeness. Where a codepoint is known, it is shown below the glyph. Where no codepoint is shown, the character is a candidate for addition to Unicode.

Going from north-west to south-east, the glyphs increase in size and complexity. (For this purpose, a circle may be regarded as a regular convex polygon with an infinite number of sides.)

Although the Unicode standard tries not to be prescriptive any more than necessary, it is quite striking how few options there are to a font designer, if the font is to be consistent.

For a start, it is unlikely that a monospaced font could comfortably include both tiny and large circles. The circles must be strictly increasing in size, and of sufficient size to enclose their smaller brethren, so there is a degree of latitude in the choice of sizes, but not much.

The visual impact of the triangles is quite startling, especially the left- and right-pointing ones, so there is a good case for scaling some shapes down a little, although this table shows all the shapes at a given size as having the same area.

As mentioned in §5 above, the "default" size for higher-order convex polygons is "normal" (a curious label, that – how does a glyph size become "normal"?), while that for all others is "medium".

There is currently no demonstrable need to populate the table more fully than is shown.

	tiny	very small		small (bullets)		medium small		medium (default1)		"normal" (default2)		large	
triangle left				◄ 25c2	◁ 25c3			◀ 25c0	◁ 25c1				
triangle right				► 25b8 2023	▷ 25b9			▶ 25b6	▷ 25b7				
triangle up				▲ 25b4	△ 25b5			▲ 25b2	△ 25b3				
triangle down				▼ 25be	▽ 25bf			▼ 25bc	▽ 25bd				
square				▪ 25aa	▫ 25ab	■ 25fe	□ 25fd	■ 25fc	□ 25fb	■ 25a0	□ 25a1	■	□
diamond				◆	◇		◇ 22c4	◆ 25c6	◇ 25c7				
lozenge				◆	◇			◆ 29eb	◇ 25ca				
pentagon										⬠	⬠ 2b20		
hexagon horizontal										⬡ 2b23	⬡ 2394		
hexagon vertical										⬢ 2b22	⬡ 2b21		
arabic star				★ 066d	☆	★ 22c6		★ 2605	☆ 2606				
circle	· 22c5	• 2219 00b7	◦ 2218	● 2022	○ 25e6	● 2981	○ 26ac	● 26ab	○ 26aa	● 25cf	○ 25cb	●	○ 25ef
circled circles	⊙ 2299				◎ 233a								
	⊙ 2a00	⊙ 29bf	⊙ 229a	⊙	⊙ 29be	⊙ 25c9	◎ 25ce	◉	◎				