L2/17-194

Proposal to create a new block for missing Block Element characters

Eduardo Marin Silva

16/06/2017

Introduction. Unicode added Block Elements very early on. Also called semi-graphics, these characters are meant to be units from which one can draw different things on the screen using only a text editor. They were very popular in the eighties since actual drawing applications were in their infancy, so this was a good replacement. As a result, we can find them on the character set of many computers of that period. They have specific rendering requirements, mainly all should be the same width and height regardless of the actual graphical part and when placed next to each other there should be no padding between characters at all, unlike characters like U+25A0 BLACK SQUARE. In Firefox the font is obviously flawed in that even the so called U+2588 FULL BLOCK looks like a rectangle and not a square, and other characters have their thickness depend on the graphical part, if one were to use such a font they could never get their "graphics" to align.

Sinclair ZX80 and ZX81. These computers were released in 1980 and 1981 respectively, when we compare their character sets (excluding controls) they are identical in repertoire. They contain many of the already encoded Block Elements, but also include some unencoded ones. The encoding of such characters is essential for applications that wish to emulate such computers and to also retrieve data in such legacy encodings cleanly.

Regular to negative mapping. If we look carefully at the code chart of ZX81, we can see that they actually intended for them to be a set of regular characters and a set of negative versions for those characters, including letters, punctuation and signs. To what end I'm not sure, maybe it served as a way to add emphasis to parts of text.

In this table I make the mapping between the distinct block elements clear, and in the same order as in the encoding of the ZX81. The proposed characters are highlighted in yellow.

Regular		Negative	
	U+0020 SPACE (same width as all of the blocks)		U+2588 FULL BLOCK
	U+2598 QUADRANT UPPER LEFT		U+259F QUADRANT UPPER RIGHT AND LOWER LEFT AND LOWER RIGHT
	U+259D QUADRANT UPPER RIGHT		U+2599 QUADRANT UPPER LEFT AND LOWER LEFT AND LOWER RIGHT
	U+2580 UPPER HALF BLOCK		U+2584 LOWER HALF BLOCK
	U+2596 QUADRANT LOWER LEFT		U+259C QUADRANT UPPER LEFT AND UPPER RIGHT AND LOWER RIGHT
	U+258C LEFT HALF BLOCK		U+2590 RIGHT HALF BLOCK
	U+259E QUADRANT UPPER RIGHT AND LOWER LEFT		U+259A QUADRANT UPPER LEFT AND LOWER RIGHT
	U+259B QUADRANT UPPER LEFT AND UPPER RIGHT AND LOWER RIGHT		U+2597 QUADRANT LOWER RIGHT
	U+2592 MEDIUM SHADE		NEGATIVE MEDIUM SHADE
	BOTTOM HALF BLOCK MEDIUM SHADE		FULL BLOCK WITH LOWER HALF MEDIUM SHADE
	UPPER HALF BLOCK MEDIUM SHADE		FULL BLOCK WITH UPPER HALF MEDIUM SHADE

Code position. Since block elements were encoded in the BMP it makes sense to also encode these characters there. I propose the range 2FE0-2FEF, although this may change if it is proven that more than 16 unencoded block elements exist. The name could be "Block Elements Extended".

Entries.

2FE0 ▒ NEGATIVE MEDIUM SHADE

→ 2592 ▒ MEDIUM SHADE

2FE1 ▄ BOTTOM HALF BLOCK MEDIUM SHADE

2FE2 ▄ FULL BLOCK WITH LOWER HALF MEDIUM SHADE

- Negative of the above character.

2FE3 ▀ UPPER HALF BLOCK MEDIUM SHADE

2FE4 ▀ FULL BLOCK WITH UPPER HALF MEDIUM SHADE

- Negative of the above character.

Discussion about information in Wikipedia. If one looks at the code chart for any of the two computers code-charts we see them reference the ZX81 character to their corresponding Unicode character when one exists. However, the character that corresponds to NEGATIVE MEDIUM SHADE is incorrectly mapped to U+2592, even though it is not encoded. Sure one could argue that the important bit is that it is a character intended to produce half shading and that making a negative out of it, just switches the black squares with the white and the white with the black, so the shading doesn't actually change and the difference that is there is barely visible. However, if one wants to truly emulate these computers, both characters must be encoded, it would otherwise introduce conflicts over what part of the screen is actually the negative part.

Discussion on the flawed font of Firefox. As I mentioned before, Firefox's default font for the Block Elements is incorrectly made, however one may try to refute such idea by showing the rendition of the name "Wikipedia" using the same font. Don't be fooled however, that rendition uses mostly just the character FULL BLOCK, and one instance of LEFT HALF BLOCK and one final one of RIGHT HALF BLOCK, if they tried to enter character like QUADRANT UPPER LEFT AND UPPER RIGHT, it would leave some space in blank, breaking the entire purpose. Such a font could work if they stretched all characters equally but they didn't. They tried to make the full block about the width and height of a typical Latin letter (in effect stretching it) and only cared to scale some other characters. This is why it is important to stress that Block Elements that go beyond one line should not interact with regular fonts, instead a monospaced font should be used with the same width as the FULL BLOCK (including space). Furthermore, LEFT ONE EIGHT BLOCK and RIGHT ONE EIGHT BLOCK, have different widths which makes no particular sense.

With respect to other negative characters. Since the negative letters, punctuations and signs occupy their own code space in the ZX81 it is tempting to say to encode that set of those characters however I believe that is better left for a discussion in another document. For now, I want to focus in the Block Elements.

Case for encoding. There is an active community around these two computers, and they are forced to use graphics when dealing with text documents in the interface between the old and the new. http://www.sinclairzxworld.com/viewtopic.php?f=3&t=2434 There is simply no reason not to encode these. One might argue that it would open the gate to many more semi-graphics from a plethora of computers, however this doesn't need to be the case. If we just restrict the encoding to platforms that still have a user community, then we are only talking about a few popular computers and their clones (that share a character set anyway). Furthermore, it increases the possibilities of artistic expression using just text.

The consortium should not just arbitrarily decide that some semi-graphics are worth encoding while others are not.

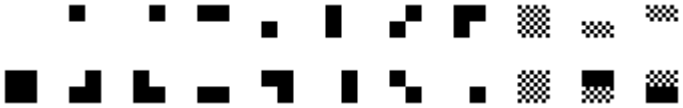

Figure 1. ZX80/ZX81 Block Elements (notice the contrastive usage of MEDIUM SHADE and NEGATIVE MEDIUM SHADE.

Figure 2. “Wikipedia” rendered with the flawed font

https://en.wikipedia.org/wiki/Block_Elements

Figure 3. The entire character set rendered on the computer itself.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** ***Proposal to create a new block for missing Block Element characters***
2. Requester's name: *Eduardo Marin Silva*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *16/06/2017*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Complete*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *Not a script but an extension to existing characters*
   Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *No*
   Name of the existing block:
2. Number of characters in proposal: *5*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary ✓ B.1-Specialized (small collection) B.2-Specialized (large collection)
   C-Major extinct D-Attested extinct E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided? *Yes*
   a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? *Yes*
   b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
   Ken Lunde
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   lunde@adobe.com
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes in the forum post.*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?
   It does address current issues with the fonts for block elements.

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

[1] Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *No*
   If YES explain
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? *Yes*
   If YES, with whom? *A couple of users in the relevant forum.*
   If YES, available relevant documents: *http://www.sinclairzxworld.com/viewtopic.php?f=3&t=2434*
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *Yes*
   Reference: *Although the size is unknown the applications include emulators, text converters for printing and virtual compilers*
4. The context of use for the proposed characters (type of use; common or rare) *Rare*
   Reference: *Translation of text between old and new computers*
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference: *http://www.sinclairzxworld.com/viewtopic.php?f=3&t=2434*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *Yes*
   If YES, is a rationale provided? *Yes*
   If YES, reference: *The already encoded Block Elements are in the BMP*
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? *Yes*
    If YES, is a rationale for its inclusion provided? *Yes*
    If YES, reference: *Discussion about information in Wikipedia*
11. Does the proposal include use of combining characters and/or use of composite sequences? *No*
    If YES, is a rationale for such use provided?
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *No*
    If YES, describe in detail (include attachment if necessary)
13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: