

INCITS/L2/09- 305

Date: August 15, 2009

Title: Comments accompanying the U.S. positive vote on ISO/IEC CD 10646 (SC2 N4079)

Source: INCITS/L2

Action: Forward to INCITS

The U.S. National Body is voting Yes with comments on the following SC2 ballot: SC2 N4079: CD 10646, Information technology -- Universal Coded Character Set (UCS).

Technical Comments

T.1.

p 2159: "A.6 Unicode collections"

Because the reference in Annex A to Unicode collections unnecessarily duplicates information that can be derived from values specified in the Unicode character database, we suggest those collections instead be defined by reference to the Age values for Unicode characters as specified in the Unicode character database file DerivedAge.txt. Because the Age values define exactly the same lists of characters as the Unicode collections in Annex A, explicitly listing these characters in Annex A represents a maintenance burden and a potential source of error for the 10646 standard.

T.2.

The U.S. welcomes the publication of the whole repertoire of CJK unified ideographs in multi-column format in the code charts, and we would like to thank the national bodies who provided fonts for that purpose. However, we are concerned that in the process of migrating to these new fonts, accidental glyph changes may have been introduced (in particular compared to the already published URO and Extension A charts). We would like WG2 to provide an assessment of the possible differences.

T.3.

We would like the CJK compatibility ideographs code charts (both BMP and SIP) to also be presented in multi-column format. As most characters have only one or two sources, the organization adopted for Extension C would be appropriate. Furthermore, we would like the necessary fonts for those glyphs to be provided by the national bodies as well, so that they are harmonized with those of the corresponding unified ideographs.

In addition, such a change would thereby also provide multi-column charts for the 12 unified ideographs present in the CJK Compatibility Ideographs block.

T.4.

The font data for the glyphs for H source (3B19, 549E) and KX source (2054B, 238A7, 24C36, 2597C, 270D2, 29B30) have problems with their contour orientation (in the PDFs of the code charts, rendered on screen, some of the stroke crossings are white).

T.5.

The U.S. requests that the font for Sundanese be updated in the CD, using the font as provided for the Sundanese proposal N3666 (L2/09-251).

Editorial Comments

E.1.

p 9, note , "The Unicode Standard, Version 5.2"

This should probably have an editor's note to remember to change "5.2" to the appropriate value. We suggest an update to "6.0."

E.2.

p 21, "The full syntax of the notation of a short identifier, in Backus-Naur form, is { U | u } [{+}(xxxx | xxxxx | xxxxxx)]"

We recommend the removal of the "[" and "]".

E.3.

p 33, note 5: "The variation selector only selects a different *appearance* of an already encoded character."

We recommend the wording be changed to: "The variation selector only selects a specific *appearance* among those acceptable for an encoded character."

E.4.

p 34, 20.1: "If a combining character is to be regarded as a composite sequence in its own right, it shall be coded as a composite sequence by association with the character 00AD NO-BREAK SPACE. For example, grave accent can be composed as 00AD NO-BREAK SPACE followed by 0300 COMBINING GRAVE ACCENT."

Change 00AD to 00A0.

E.5.

p 34, 20.1, note: "NOTE – Indic matras form a special category of combining characters, since the presentation can depend on more than one of the surrounding characters. Thus it might not be desirable to associate Indic matra with the character SPACE"

Change the last word from SPACE to NO-BREAK SPACE.

E.6.

p 37, last paragraph of 22.2: "A 'unique-spelling' rule is defined as follows. According to this rule, no coded character from a table for Rows 09 to 0D or 0F, or for the MYANMAR block in Row 10, shall be regarded as equivalent to a sequence of two or more other coded characters taken from the same table.

The text as it currently reads is misleading about two-part vowels. Reword the section with correct examples.

E.7.

p. 2180 Annex I, Ideographic description characters.

The USNB requests removing the restriction on Ideographic Description Sequences that they contain only CJK ideographs and radicals to allow them to be used for other East Asian ideographic scripts.

E.8.

The U.S. requests the following Kaithi decompositions be added to the nameslist for the Kaithi characters in the CD:

110AB KAITHI LETTER VA: 110A5 KAITHI LETTER BA 110BA KAITHI SIGN NUKTA

1109C KAITHI LETTER RHA: 1109B KAITHI LETTER DDHA 110BA KAITHI SIGN NUKTA

1109A KAITHI LETTER DDDHA: 11099 KAITHI LETTER DDA 110BA KAITHI SIGN NUKTA

Note: Kaithi was added in Amendment 6.