

| | |
|---------|--|
| Title: | Comments accompanying the US positive vote on the CD ISO/IEC 10646 |
| Source: | NCITS/L2 |
| Action: | Forward to NCITS |

The US is voting YES on the Committee Draft ISO/IEC 10646-2 with the following comments:

Technical Comments:

Secondary Multilingual Plane for scripts and symbols: Western Musical Symbols

T.1 The list of combining characters as specified by the Annex B.1 is incorrect; it should be amended as follows (encoding within Plane 1): D110-D117, D11A-D120, D134-D135, D157-D15A.

Furthermore the chart pages should show a dotted circle as customary for these characters.

T.2 The Annex E (Usage of musical symbols) illustrates ‘equivalence’ between pre-composed note values and primitives from which complete notes may be constructed. However these characters and sequences are not strictly equivalent. We suggest replacing the last phrase of the first column by:

‘Although these characters are not formally equivalent, in practice conforming devices may not want to distinguish among the following sequence pairs:’

The ‘=’ symbol used in the following lists should be replaced by a symbol reflecting a looser association than an equivalence. The character ‘~’ is suggested.

Supplementary Plane for CJK Ideographs:

T.3 The CD should reflect the changes suggested in the IRG editors report following the 14th IRG meeting in Singapore.

Collections

Annex A (page 368) or a possible Amendment to part 1

T.4 A new collection needs to be added to cover all planes from 0 to 16. The collection number, name and positions (character range) are suggested as follows:

10646 UNICODE 0000 0000 - 0010 FFFF

As the collection spans many plane, it is suggested that the text should be added to the part 1 of the standard. As the formal introduction of the part 2 within ISO/IEC 10646 will require a change in the first part of the standard, it could be part of the necessary amendment.

Although the US understands that the text appearing at the very end of the CD ballot (page 386) looking like a part 1 amendment is not part of the CD ballot itself, it is nevertheless an informative document to be used as a template for the eventual amendment to part 1.

T.5 The clause A.1 contains a definition for CJK UNIFIED IDEOGRAPHS EXTENSION B specifying characters from Part 1 (like 0FA0E, 0FA0F,...FA29). These characters should not be part of Extension B but documented as being part of a larger collection including all CJK UNIFIED IDEOGRAPHS from all planes. The intent is to make sure that these 10 ideographic characters are part of that collection. The US asks that these characters be removed from collection 1006 and be added to a new collection in the amendment to part 1 specified above. Source information for these 10 characters should also be provided in the same amendment.

Editorial Comments:

Secondary Multilingual Plane for scripts and symbols: Western Musical Symbols

Clause 6 (page 2)

E.1 The end of the last sentence (the SMP cannot be used as a two-octet coded character set) is confusing, a reference to UCS-2 should be used instead.

E.2 In the first sentence of the 2nd paragraph, replace ‘them’ by ‘CJK Ideographs’.

Clause 9 (page 3)

E.3 The note is unclear. Please clarify.

Annex E (page 383 and 384):

E.4 Remove the numbering in all lists in the Annex clauses E.2 and E.3.

Supplementary Plane for CJK Ideographs:

Clause 7 (page 3)

E.5 The end of the last sentence (the SIP cannot be used as a two-octet coded character set) is confusing, a reference to UCS-2 should be used instead.

Clause 10.2 (page 36)

E.6 The clause is imprecise in many cases about the correlation between the sources as described in the first part and as described in the explanation of the text file source document. For example, it is unclear to which source the Unique Hanzi abbreviated names (like G_4K, G_BK, etc.) refers to. The descriptive section (first part) should mention explicitly these abbreviated names in correlation with the full name (like G_CY for Ci Yuan).

E.7 The Hong Kong Hanzi source should be updated to reflect the new naming for the sources: Government Common Character Set (GCCS) and Hong Kong Supplementary Character Set (HKSCS) instead of GCCS_1 and GCCS_2.

E.8 Furthermore, the Hong Kong Hanzi source specifies two groups (GCCS) and Industry Collection. But the source document only references Industry Collection (HA-, HB-, HC-). It should be made clear whether or not the GCCS and HSCS collections are fully contained within the Industry Collection.

E.9 The Kanji source should explicitly refer to the newly adopted JIS-x-213, as it is now known as the official source for these characters.

E.10 The second Editor’s note can be addressed by developing the explanation of the plain text file. The explanation basis should be based on the following: The Kangxi Dictionary is really used for two purposes: true entries using the ‘dddd.dd0’ notation (‘dddd’ is the page number, and ‘dd’ is the order within the page in decimal notation), and duplicates noted as ‘dddd.dd1’. The duplicates are ordered within the single ‘dddd.dd1’ entry as determined by the IRG committee. That internal order bears no significance.

End of US comments