Title: Considerations concerning the publication of ISO/IEC 10646 6th Edition

Source: Michel Suignard, Project Editor
Status: Individual Contribution
Distribution: WG2
Replace: N4859

Summary: This document is requesting an endorsement from SC2 to publish ISO/IEC 10646 under a new format as requested by ISO/IEC new directives. While the current 5th edition already follow these directives, implementing recommendations presented in this document would provide an improvement by separating the code charts from the main text of the standard. It would also provide a permanent location for the ancillary data.

Context

The 5th Edition of ISO/IEC 10646 was significantly delayed because of changes made necessary by the new ISO publishing guidelines. Instead of using ancillary data and code chart files located in the same folder/directory as the main body of the text as done for prior editions, the new guidelines required to use regular hyperlinks with content hosted by ISO. This required several forth and back between the editor, the SC2 secretariat and ITTF to come up with an acceptable compromise. The end result was the creation of the following items (links provided):

- ISO/IEC 10646 5th edition (including code charts, a 2694 pages document, 109MB in compressed form, 115MB extracted), at:
  http://standards.iso.org/ittf/PubliclyAvailableStandards/c069119_ISO_IEC_10646_2017.zip
- List of Standardized Variation Sequences at: http://standards.iso.org/iso-iec/10646/ed-5/UCSVariants.txt,
- Source references for pictographic symbols at: http://standards.iso.org/iso-iec/10646/ed-5/EmojiSrc.txt ,
- Source references for CJK Ideographs at: http://standards.iso.org/iso-iec/10646/ed-5/CJKSrc.txt ,
- Source references for Tangut Ideographs at: http://standards.iso.org/iso-iec/10646/ed-5/TangutSrc.txt ,
- Source references for Nüshu characters at: http://standards.iso.org/iso-iec/10646/ed-5/NushuSrc.txt ,
- List of Named UCS Sequence Identifiers (NUSI) at http://standards.iso.org/iso-iec/10646/ed-5/NUSI.txt :
- Collection 374 at http://standards.iso.org/iso-iec/10646/ed-5/JIS-X-0213-FromPrevious.txt ,
- Collection 390 (Mojo-Joho-Kiban) at: http://standards.iso.org/iso-iec/10646/ed-5/JMJKI-2016.txt ,
- Alphabetically sorted list of character names: http://standards.iso.org/iso-iec/10646/ed-5/Allnames.txt ,
- Names of Hangul syllables at: http://standards.iso.org/iso-iec/10646/ed-5/HangulSy.txt .

An alternative that was considered before and was discussed in WG2 N4859 was to host all documents (except the standard text itself) in the Unicode web site. There were several reasons for that:

a) Many of these documents are slightly modified versions of identical files also published by the Unicode Consortium which would be for the 5th edition of 10646:

- Standardized Variants: https://www.unicode.org/Public/UCD/latest/ucd/StandardizedVariants.txt
- Pictographic Symbols: https://www.unicode.org/Public/UCD/latest/ucd/EmojiSources.txt
- Source references CJK: https://www.unicode.org/Public/UCD/latest/ucd/Unihan.zip (extract)
- Source references Tangut: https://www.unicode.org/Public/UCD/latest/ucd/TangutSources.txt
- Source references Nüshu: https://www.unicode.org/Public/UCD/latest/ucd/NushuSources.txt
- NUSI: https://www.unicode.org/Public/UCD/latest/ucd/NamedSequences.txt

b) Many comparable files hosted by Unicode are already referenced normatively in 10646 clause 2 (Normative references). It would be very simple to extend the concept to these files by adding them to the list of normative references in the same clause.

c) Unlike ISO, the Unicode Consortium maintains previous versions of the standards indefinitely, which ensure access to these resources in the future, even when future editions are published.

Publishing these files in the Unicode web site would be made with the understanding that the Unicode Consortium would follow the rules expressed in the ISO directives part 2 clause 10.2 (repeated below):

10.2 Permitted referenced documents

In principle, normatively referenced documents shall be documents published by ISO and/or IEC. In the absence of appropriate ISO and/or IEC documents, those published by other bodies may be referred to in a normative manner provided that

a) the referenced document is recognized by the committee as having wide acceptance and authoritative status,

b) the committee has the agreement of the authors or publishers (where known) of the referenced document for its inclusion as a reference,

c) the authors or publishers (where known) have also agreed to inform the committee of their intention to revise the referenced document and what points the revision will affect and

d) the document is available under commercial terms which are fair, reasonable and nondiscriminatory.

Informative reference may be made to any other type of document. Informative references shall be listed in the bibliography.

ISO and IEC normatively referenced documents shall have reached at least the enquiry stage (40.20 DIS or CDV).

The committees shall validate all referenced documents when a document is revised.

The normative references list shall not include the following:

- *referenced documents which are not publicly available (in this context, "publicly available" means published documents which are available free of charge, or available commercially under reasonable and non-discriminatory terms to any user);*
- *referenced documents which are cited only informatively as bibliographic or background material.*

A similar process can be done for the code charts. Instead of embedding them in the standard itself they could be offered as separate resources, with the same benefit as for the data files, i.e. making them available for current and future editions of the standard. The code charts would be publicly available under the usual Unicode copyright terms which allow liberal quoting. The link below gives terms of use allowing full usage of these code charts in the context of ISO/IEC 10646:

The Unicode Consortium specifically grants ISO the right to use those code charts with their associated character names list to show the repertoire of characters for that standard, as a normatively referenced, integral part of that standard.

Adopting this separation between the standard text and the code chart would reduce the standard text to less than 140 pages and few MB. There is already an implementation of these recommendations that can be accessed in http://www.unicode.org/charts/About.html#iso10646.

The following text describes the changes required in ISO/IEC 10646 to implement these considerations. These are described in the context of the 5th edition (because the related links are active and give an actual overview of the expected results), but could be extended to future editions and amendments.

Changes to ISO/IEC 10646:

(the following instructions refer to the 5th edition but can be extrapolated to future amendment and editions)

Clause 2 Normative references

Add the following text at the end:

Variation selectors and variation sequences:
https://www.unicode.org/wg2/iso10646/edition5/data/UCSVariants.txt

Source references for pictographic symbols
https://www.unicode.org/wg2/iso10646/edition5/data/EmojiSrc.txt

Source references for CJK ideographs:
https://www.unicode.org/wg2/iso10646/edition5/data/CJKSrc.txt

Source references for Nüshu characters:
https://www.unicode.org/wg2/iso10646/edition5/data/NushuSrc.txt

Source references for Tangut ideographs:
https://www.unicode.org/wg2/iso10646/edition5/data/TangutSrc.txt

Named UCS Sequence Identifiers:
https://www.unicode.org/wg2/iso10646/edition5/data/NUSI.txt

Collection MOJI-JOHO-KIBAN IDEOGRAPHS-2016:
https://www.unicode.org/wg2/iso10646/edition5/data/JMJKI-2016.txt

Collection JAPANESE JIS X 0213:2004 IDEOGRAPHS FROM PREVIOUS JIS STANDARDS:
https://www.unicode.org/wg2/iso10646/edition5/data/JIS-X-0213-FromPrevious.txt

Clause 33 Code charts and lists of character names, sub-clause 33.1 General

Replace first sentence by:

Detailed code charts and lists of character names for the BMP, SMP, SIP and SSP are introduced on the following link: http://www.unicode.org/wg2/iso10646/edition5/charts/.

Annex G

Update the link to: https://www.unicode.org/wg2/iso10646/edition5/data/Allnames.txt

Annex R

Update the link to: https://www.unicode.org/wg2/iso10646/edition5/data/HangulSy.txt

--