Title: CJK chart and source references update
Source: Michel Suignard, ISO/IEC 10646 Project Editor
Distribution: UTC, WG2

Summary: This document describes current issues in CJK encoding and representations and proposes changes to remedy these issues.

1. Extension C code point U+2B08F

In ISO/IEC 10646 (4th edition), that code point is displayed in the chart as follows:

Proposed change: Remove UCI reference and modify glyph for TD-6162 to look like the one for the UCI reference. The new entry is visible below (in the 'details' section).

Details:
In the original introduction (Amendment 5 to ISO/IEC 10646:2003) it was shown only with the T-source and the glyph corresponding to the UCI source. This was still the case in ISO/IEC 10646:2011 (2nd edition).

Following IRG#36, it was determined that the original unique source (TD-6162) did not look like the glyph shown in the chart. Therefore, because the source was the only one defining the code point entry, a UCI was created to denote a code point without identified source reference. For production issues, the original and new sources persisted in the 3rd and 4th edition of ISO/IEC 10646.

For what was seen as a production error, two documents (IRG N2021 and IRG N2037) requested the removal of the extraneous T–source. However since then, Taiwan has determined that the glyph for TD-6162 actually looks like the one originally shown in Amendment 5 and has provided a glyph for it.

It does not seem that the glyph shown for TD-6162 in the 3rd and 4th edition is yet encoded.

2. Extension C code point U+2AD12

In ISO/IEC 10646 (4th edition), that code point is displayed in the chart as follows:

1

2AD12
木 75.15
UCI-00938

Proposed change: Replace UCI reference by T-source TE-353F and modify glyph it to look like the one introduced in the original amendment (5):

2AD12
木 75.15
TE-353F

Details:
In the original introduction (Amendment 5 to ISO/IEC 10646:2003) it was shown as follows:

2AD12
TE-353F

(Note the amendment specific format)
But in the following 2nd edition, it was shown as follows:

2AD12
木 75.15
TE-353F

Following IRG#36, it was determined that the original unique source (TE-353F) was supposedly looking like the one in the 2nd edition chart and not like in amendment 5. Furthermore, it had the incorrect radical; it should be the tree (木) and the glyph shown is already encoded at U+22E16. Therefore, because the source was the only one defining the code point entry, a UCI was created to denote a code point without identified source reference. Following this, the 3rd and 4th edition showed that entry as follows:

2AD12
木 75.15
UCI-00938

This is obviously incorrect because while having a new UCI number is expected, the glyph should have been reversed to the glyph shown in Amendment 5. This is noted in document IRG N2027.

However, Taiwan has now determined that the glyph for TE-353F actually looks like the one originally shown in Amendment 5 and has provided a glyph for it.

3. Main Block (URO) code point U+7921

In ISO/IEC 10646 (4th edition), that code point is displayed in the chart as follows:

7921
石 112.14
GE-3559   HB2-F2A1   T2-662B

Proposed change: Replace glyphs for H-source and T-source to look similar to the G-source:

7921
石 112.14
GE-3559   HB2-F2A1   T2-662B
(Layout is an approximation)

2

Details:

This is originally a T-source error for U+7921. The glyph should have the same component structure as the G-Source glyph, ⿰石尃, though the top portion of the right-side should be the split (four-stroke) form of Radical #140. CNS 11643-1992 is correct, but CNS 11643-2007 introduced this glyph error, which then propagated into the Code Charts.

The problem is already present in ISO/IEC 10646:2003 (first edition)

It was then also propagated to the H source when it was introduced in ISO/IEC 10646:2012 3rd edition. Note that the H source is part of Big-5, not HKSCS.

4. CJK Compatibility code point U+2F949

In ISO/IEC 10646 (4th edition), that code point is displayed in the chart as follows:

Proposed change: Replace T-source by a new UCI source:

(According to http://www.unicode.org/Public/7.0.0/ucd/USourceData.txt the next available value is 01199)

Details:

Two characters U+4039 and U+2F949 have currently the same T-source value: T6-4B7A. This is the result of a complex dis-unification of U+4039 with a new character encoded at U+9FC3 and affecting two compatibility characters (U+FAD4 and U+2F89F). See WG2 N3196 for the details.

However while U+FAD4 correctly lost its T-source value (replaced by UCI-00947), the same thing did not happen for U+2F949. This is an error and should be fixed by using as reference the next UCI value available which is currently UCI-01199.

5. Main Block (URO) code point U+9FCF

In the proposed 2nd amendment to ISO/IEC 10646 (4th edition), that code point is displayed in the chart as follows:

Proposed change: Replace the radical value/stroke count from 167.7 to 167'.7.

3

Details:

The radical of U+9FCF should be 167' (钅, simplified gold), not 167 (金, traditional gold), like that of U+9FD4.

6. Extension A code point U+4CA4, and new code point (U+9FD0)

In ISO/IEC 10646 (4th edition), that code point is displayed in the chart as follows:

Proposed change: Remove the H-source reference (both glyph and value) and replace by a new character as follows:

(Proposed in amendment 2 to the 4th edition)

Details:

The H-source character uses a traditional radical and therefore should not be unified with the G-source character. This problem has existed since the H-source was added to U+4CA4 (ISO/IEC 10646 2nd edition).

7. Extension A code point U+3D1D

In ISO/IEC 10646 (4th edition), that code point is displayed in the chart as follows:

Proposed change: Move the H-source reference (both glyph and value) to U+2A3ED which currently shows as:

(See alternative change in details)

Details:

The radical shape show for H-91B5 (lower half of the character) is an acceptable variant of the water radical (水). However the character was unfortunately dis-unified into U+2A3ED which provides a better mapping for the H-source character. The code point U+2A3ED also uses a different radical (黍). An alternative is to leave the sources as they are, add the alternate radical/stroke at both code points, and possible an entry in Annex P to explain the context.