

Universal Multiple-Octet Coded Character Set
 International Organization for Standardization
 Organisation Internationale de Normalisation
 Международная организация по стандартизации

Doc Type: Ideographic Rapporteur Group Document
Title: Feedback on IRGN2369
Source: Eiso Chan(陈永聪)
Status: Individual Contribution to IRG #52, Hong Kong
Action: For consideration by IRG
Date: 2019-05-06

UTC has submitted IRGN2369 (aka L2/19-185) to do the horizontal extensions for 150 CJK Unified Ideographs. This document shows something questionable as below.

1. UTC-00081 for U+9FB8

U	UTC-00081	𠄎	𠄎	U+9FB8
---	-----------	---	---	--------

9FB8 𠄎
 小 42.2
 G9-FE6D

The current UTC-00081 glyph doesn't match the G9-FE6D glyph under U+9FB8, but it's as the same as the UK-02966 glyph under *U+300EE in WG2 N5032.



When we check USourceData-12.0.0.txt, the Field 6 shows the source of UTC-00081 is GB 18030-2000, which is as the same as G9-FE6D.

In L2/03-411, Deborah Goldsmith and Eric Muller pointed out that there are several unencoded characters in GB 18030 and HKSCS. In WG2 N2807R and WG2 N2808R, Michel and John provided the encoding method of these unencoded characters in GB 18030-2000.

9FB8	𠄎	FE6D	E832
------	---	------	------

So, the UTC-00081 glyph must match the source glyph in GB 18030-2000.

Solution:

Modify the UTC-00081 glyph to match GB 18030-2000 and keep the horizontal extension.

2. UTC-00111 for U+9083

U	UTC-00111	邃	邃	U+9083
---	-----------	---	---	--------

9083
辵 162.13
 邃 邃 邃 邃 邃
G0-6564 HB1-C2E3 T1-774E J0-6364 K0-6245

The current UTC-00111 doesn't match any existing source under U+9083, KP0-E4AC included.

As Ming Fan's comment in the email to Unihan list, UTC-00111 should be unified with U+2E11C as below.

2E11C
穴 116.13
 邃
JMJ-058251

Solution:

1) Change the request as below.

F	UTC-00111	邃	邃	U+2E11C
---	-----------	---	---	---------

2) Change the information in UAX #45 as below.

UTC-00111;F;U+2E11C;116.13;0869.071;邃穴邃;kMatthews 5434;

3. UCI-00942 for U+24FB9

B	UCI-00942	擘	擘	U+24FB9
---	-----------	---	---	---------

The source reference has just been changed back to the G-Source in WG2 #67 in London according my document IRGN2245 and there is no need to add the UCI-prefix source reference to U+24FB9.

Solution:

Remove this request.

4. UCI-00938 for U+2AD12

C	UCI-00938	攄	攄	U+2AD12
---	-----------	---	---	---------

The current UCI-00938 doesn't match the TE-353F glyph for U+2AD12 and the UCI glyph is wrong.

The TE-353F glyph was wrong in Unicode, 5.2.0.

2AD12
木 75.15
 攄
TE-353F

However, TCA had modified the glyph since Unicode, 8.0.0.

2AD12
木 75.15 榦
TE-353F

Please see the details in Section 2 of WG2 N4621 as below.

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 榦
木 75.15 TE-353F

Details:

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

2AD12 榦
TE-353F

(Note the amendment specific format)

But in the following 2nd edition, it was shown as follows:

2AD12 榦
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 榦
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.

Solution:

Remove this request.

5. UCI-00940 for U+2B08F

C	UCI-00940	籟	籟	U+2B08F
---	-----------	---	---	---------

There is no need to do the horizontal extension for a UCI character. Please see the details in Section 1 of WG2 N4621 as below.

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

2B08F
竹 118.9
𪗇 TD-6162 𪗈 UCI-00940

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).

2B08F
竹 118.9
𪗈 TD-6162

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.

Solution:

Remove this request.

6. More HE requests for the UK-Source characters

There are 4 UK-Source characters in IRGN2369, but there are more UK-Source characters which needs to do the horizontal extensions according to the discussion result of WS2015. I list them as below.

A	UK-02787	𪗉	𪗊	U+38C7
A	UK-02918	𪗋	𪗌	U+46E9
B	UK-01678	𪗍	𪗎	U+20219
B	UK-01344	𪗏	𪗐	U+21249
B	UK-01470	𪗑	𪗒	U+21827
B	UK-01586	𪗓	𪗔	U+2327B
B	UK-02922	𪗕	𪗖	U+2363B

B	UK-02923	欖	欖	U+23839
B	UK-01627	漚	漚	U+23FD5
B	UK-01462	𪗇	𪗇	U+24261
B	UK-02656	碇	碇	U+2548E
B	UK-01967	蔘	蔘	U+26C9E
E	UK-01327	𠄎	𠄎	U+2C176
F	UK-02896	街	街	U+2DE4A

7. A special case

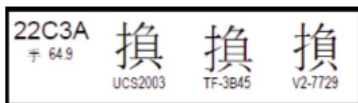
UK-02790 is under *U+3050C in WG2 N5032.



But, the Irish NB has pointed out UK-02790 (WS2015-01436) should be unified to U+22C3A in the ballot comment, WG2 N5021R.

T7. Page 2671, Row 3050: CJK Unified Ideographs Extension G

Ireland requests the removal of U+3050C (UK-02790) from CJK Unified Ideographs Extension G as it is unifiable with U+22C3A. See attached image. (The T glyph form is anomalous, but as it has the reading *fū* it should be the same character as UK-02790.).



Not accepted

It is true that U+3050C 損 looks a lot like V2-7729. But the alternative solution may be to move V2-7729 to Ext G at the position U+3050C. This should be discussed by IRG experts at IRG #52 before deciding.

If the comment from Irish NB were accepted in IRG #52, UK-02790 (WS2015-01436) should be removed from WS2015, and it's better to do the horizontal extension to U+22C3A as below.

B	UK-02790	損	損	U+22C3A
---	----------	---	---	---------

(End of Document)