

Universal Multiple-Octet Coded Character Set International Organization for Standardization

Doc Type: Working Group Document

Title: IRG 38 Liason Report

Author: Dr. Ken Lunde, Unicode Liason & L2 Delegation Head

Date: 2012-06-21

IRG 38 took place in Gyeongju, Republic of Korea from June 18th through 21st, and was hosted by KATS (*Korean Agency for Technology and Standards*). The two main topics were to finish Extension E, and to begin Extension F. The IRG 38 Resolutions are now available.*

In addition to me, who was the sole L2/Unicode delegate, there were delegates from China (6), Hong Kong SAR (3), Taiwan (5), South Korea (5), and Japan (4), along with two individuals who happened to come from Japan.

The following is the future IRG meeting schedule:

IRG 39: Hanoi, Vietnam, 11/12/2012 through 11/16/2012 (approved by WG2)

IRG 40: Hong Kong SAR, 05/20/2013 through 05/24/2013 (approved by WG2)

IRG 41: Japan, 11/18/2013 through 11/22/2013 (tentative)

New K Source

Korea proposed a horizontal extensions for 154 characters (IRG N1857), which will introduce a new Korean source, K6. This new Korean source will correspond to KS X 1027-5, which is expected to be published by the end of this year. At some point, this will need to be reflected in UAX #38.

The kMandarin & kTotalStrokes Fields

I discussed John Jenkins' document about the procedures for changes to the *kMandarin* and *kTotalStrokes* fields, which mainly involves keeping the UTC aware of any changes to these important fields, due to their use in the CLDR for sorting purposes (IRG N1861).

Urgently-Needed Characters

The UNC (*Urgently Needed Character*) proposal that we originally submitted to WG2 (IRG N1843) was discussed, along with somewhat critical comments from Japan. The conclusion was that Annex C of the IRG's P&P (*Principles and Procedures*) would be modified to refer to small UNC repertoires, with a specified soft limit of 30, and a higher unspecified hard limit, along with "urgently-needed" being more clearly defined. Two proposals in Japan's comments were accepted, specifically 1) that each national body's UNC submission would be parallel to those of other national bodies, and thus processed separately (this is their "isolated" principle), and 2) that no national body can have two concurrent UNC submissions, meaning that as soon as the one being processed is handed off to WG2, meaning out of the IRG's hands, they could then initiate another. I was tasked with rewriting Annex C of the IRG's P&P (IRG N1823, Version 5, Draft 2) to reflect these changes.

Extension E

As part of the final Extension E review, the properties for the following U-Source ideographs were changed:

00315 (UTC00497) Radical: 人 → 儿

03193 (UTC00042) KangXi Index: 0427.151 → 0427.161

* <http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg38/IRGN1870Resolutions.doc>

05882 (UTC00361) KangXi Index: 0861.051 → 0861.241
01166 (UTC00562) Stroke Count: 11 → 12
01463 (UTC00606) Stroke Count: 6 → 7
04779 (UTC00766) Stroke Count: 12 → 13
06884 (UTC00439) Stroke Count: 6 → 7
08078 (UTC00109) Stroke Count: 7 → 8
00535 (UTC00506) First Stroke: 2 → 5
01189 (UTC00560) First Stroke: 4 → 3
01293 (UTC00436) First Stroke: 3 → 4
01339 (UTC00603) First Stroke: 3 → 4
01371 (UTC00611) First Stroke: 3 → 4
04324 (UTC00683) First Stroke: 1 → 2
06151 (UTC00722) First Stroke: 2 → 5
06851 (UTC00762) First Stroke: 4 → 3

There was also a discussion of specific Extension E unification cases, some of which affected U-Source ideographs. The following were postponed, meaning that they will not be included in Extension E, but no unification decision was made:

UTC00092 (possible unification with U+8C81)
UTC00154 (possible unification with U+8821)

The following U-Source ideograph was unified as follows:

UTC00278 (unified with UTC00424, which is in Extension E as 02377)

I also reported the following genuine duplicate ideograph (same U-Source) that I previously reported against Extension E Version 6.2 (IRG N1779), but appropriate action was not taken at that time:

UTC00839 (duplicate of U+9FC0, which has the same U-Source)

All so-called “D Set” characters were not discussed, and will thus not be included in Extension E. Postponed characters can be submitted for Extension F.

The Extension E report (IRG N1864) indicates that Extension E will be declared final on 09/21/2012 (one month prior to WG2 60), at which point it will be submitted to WG2.

Extension F

National body Extension F submission plans were discussed and recorded (IRG N1865), which are summarized as follows:

- Japan plans to submit approximately 3,000 ideographs, which was originally presented during IRG 37 (IRG N1813).
- The SAT Project, which is based in Japan, is planning to make an expert contribution of 6,440 ideographs from the SAT Taishō Daizōkyō (大正大藏經) database (IRG N1858).
- I estimated the US/Unicode submission to be no more than 100 ideographs, mainly from UTR #45 (IRG N1872). Among these are likely to be unencoded Adobe-CNS1 glyphs that have been in UTR #45 for quite some time, though many of them are likely to be unified, which is okay.
- Korea is expecting to submit anywhere from 100 to 1,000 ideographs (IRG N1855).
- China plans to submit approximately 7,000 ideographs, approximately 1,000 of which are from the Square Zhuang Script (IRG N1874).
- Taiwan is expecting to submit approximately 4,000 ideographs, which includes 83 from CNS 11643-2007 (IRG N1873).

National bodies were strongly encouraged to keep the size of their submissions to 4,000 or less. The Extension F submission deadline was set to 10/20/2012, which is one month prior to IRG 39.

CNS 11643 References

Japan raised an issue about incorrect or stale CNS 11643 references in the IRG's P&P, which also plague the text of ISO/IEC 10646. One issue is that all planes, meaning 1 through 7 and 11 through 15, are associated with the year 1992, which only the first seven planes existed at that time. The history of Plane 15 is particularly complex, and I helped by providing accurate details.

Old Hanzi

The final Old Hanzi Group Report was read, and the group was officially dissolved. There was unanimous appreciation for the work that had gone into this effort. This activity will now be handled directly by WG2.

Strokes

Total strokes and first strokes were discussed.

In response to WG2's request for total stroke information for submissions, the discussion concluded that total-stroke values are locale-dependent. The total number of strokes can vary by region, and a single value is guaranteed to be inaccurate for a large number of ideographs. In other words, the IRG will not require this information for submissions, and that indexing radical and number of remaining strokes is sufficient.

First-stroke information, which is currently required for submissions, was also discussed, and the conclusion was that this information is still deemed necessary, because it helps in the ordering process, which is according to Kangxi principles, but that the submitter does not need to include it. Instead, the IRG Technical Editor will assign first-stroke information according to Kangxi principles, which will result in better uniformity. In other words, first-stroke information will be handled internal to the IRG, with no discussion, and will be an Informative property. As part of this discussion, additional variant shapes will be added to the "Variant Shapes" column of the relevant first-stroke specification document (IRG N954AR).

IVD

Four IVD-related documents were discussed (IRG N1757, IRG N1788, IRG N1840, and IRG N1844R), along with critical comments from Japan and Korea. There is still much resistance to using the IVD as part of the IRG's work. To some extent, I agree. But, the IVD provides national bodies with an avenue for handling characters that would otherwise be unified.

Other Tidbits

I suggested to Chuck Muller that he give a short presentation and demo of his online CJKV-English dictionary[†], which he did. He was also asked to prepare an accompanying document that provides IRG attendees with more information about this resource (IRG N1868).

Taichi KAWABATA (Japan), who was not present at IRG 38, submitted a new radical/stroke resource (IRG N1869). It is useful because it lists many ideographs multiple times, if the stroke count of the radical can vary, such as by locale. It makes finding ideographs an easier task.

That is all.

[†] <http://www.buddhism-dict.net/dealt/>