Universal Multiple-Octet Coded Character Set
International Organization for Standardization

Doc Type: Working Group Document

Title: IRG #48 Liaison Report

Source: Ken Lunde, Adobe Systems Incorporated

Date: 2017-06-25

IRG #48 took place in Bundang-gu, Seongnam-si, Gyeonggi-do, South Korea from 2017-06-19 through 2017-06-22, was hosted by the National Radio Research Agency (RRA) and the Korean Agency for Technology and Standards (KATS or 국가기술표준원 in Korean), and was held in Grand Conference Hall A on the ninth floor of the Telecommunications Technologies Association (TTA or 한국정보통신기술협회 in Korean) building. The main discussion topic was **IRG Working Set 2015 Version 3.0** (aka Extension G), in terms of dealing with comments and member body responses from the latest review cycle. The IRG #48 Recommendations (IRG N2220) are now available. Also be sure to read IRG N2219 for more details.

Including myself as the sole US/Unicode representative, in attendance were **29 representatives and experts** from China (11), Hong Kong SAR (1), TCA (2), ROK (7), Japan (3), and SAT (2), along with one invited expert from Hong Kong SAR, and one from China. Below is the group photo that was taken immediately after the meeting ended on 2017-06-22:

As stated in Recommendation IRG M48.1, the following is the future IRG meeting schedule:

IRG #49   San José, California, USA, **2017-10-16 through 2017-10-20**

IRG #50   China, 2018-05-21 through 2018-05-25 (to be confirmed)

IRG #51   Hong Kong SAR, 2018-10-22 through 2016-10-26 (to be confirmed)

IRG #52   Seeking host

IRG #53   Seeking host

IRG Working Set 2015 (aka Extension G) Status

Most of the second day of the meeting, along with the morning of the third day, was spent performing IRG Working Set 2015 editorial work, which entailed going through review comments for Version 3.0 (IRG N2179), along with the responses from the submitters.

The following **seven UTC-submitted characters** needed to be discussed in greater detail, and their final dispositions are provided below (IRG Working Set 2015 serial numbers are also provided):

UTC-01024   01859: **Unified with U+6BC0 or U+6BC1**. Justification: two SAT-submitted characters in Extension F were unified with U+22758 for the same reason, and a new UCV may be added.

UTC-01161   03123: **Unified with U+7BC9**. Justification: a China-submitted character in Extension F, G_Z0511201/00393, was unified with U+20385 for the same reason.

UTC-01164   03444: **Disunified**.

UTC-01276   02284: **Pending**. There is a desire to unify with UTC-01272/02262 based on the evidence, which strongly suggests that the two different sources refer to the same name/person, and that the glyph in one of the sources is incorrect.

UTC-01301   04429: **Pending**. There is a desire to unify with U+28AB5 for the same reason as UTC-01276/UTC-01272 above.

UTC-01304   04435: **Pending**. There is a desire to unify with U+28B02 for the same reason as UTC-01276/UTC-01272 above.

UTC-01307   04448: **Withdrawn due to unclear evidence**.

All other UTC-submitted characters were handled according to the UTC's response to the review comments. Note that for the three characters that are marked as **pending**, the UTC needs to either 1) provide additional evidence as part of the next review cycle in order to keep them in the working set; 2) withdraw them; or 3) accept unification.

The UTC also needs to modify the glyphs for the following ten characters according to the review comments, and resubmit the bitmaps and TrueType font by **2017-07-21**:

UTC-00992   00828

UTC-01006   01208

UTC-01013   01245

UTC-01018	01399: Same adjustment as UTC-01013/01245 above.
UTC-01168	04134
UTC-01169	04721
UTC-01219	00777
UTC-01220	00771
UTC-01226	00983
UTC-01243	02041

For UK's benefit, I am listing **eleven UK-submitted characters**, with UTC source references, that needed to be discussed in greater detail, and their final dispositions are provided below:

UTC-01430	03677: Unified with U+2C7CF (Extension E).
UTC-01677	00142: Disunified.
UTC-01941	03798: Unified with U+891D.
UTC-01942	03800: Unified with U+2B304 (Extension C).
UTC-02614	05541: Disunified.
UTC-02810	00597: Unified with U+2D227 (Extension F).
UTC-02930	03399: Withdrawn.
UTC-02932	00112: Disunified with UTC-01647/00044, but **UK needs to modify the glyph for UTC-01647 to fill the em-box**.
UTC-02935	02001: **Pending**. There is a strong desire to unify with U+23D21. Need more information about U+23D21. *Sufficient information was received during the meeting from Ming Fan, but there wasn't time to present it, so it will be presented as part of the next review cycle.*
UTC-02651	02179: **Pending**. There is a strong desire to unify with U+241C6. Need more/better evidence. The referenced dictionary is not approved.
UTC-02960	04752: **Pending**. Need more evidence, preferably from a dictionary. This is the 84-stroke ideograph.

Per Recommendation IRG M48.2, the following is the **IRG Working Set 2015 Version 4.0** schedule between now and IRG #49:

2017-07-21	The IRG Chief Editor distributes IRG Stroke Count Guidelines (IRG N2221) and the list of new examples in IWDS
2017-07-21	The appropriate IRG member bodies submit updated bitmaps to the IRG Chief Editor
2017-08-04	The IRG Chief Editor distributes IRG Working Set 2015 Version 4.0 (IRG N2223)
2017-09-01	The IRG member bodies' chief editors submit review comments to the IRG Chief Editor
2017-09-15	The IRG Chief Editor distributes consolidated comments

2017-10-06   The IRG member bodies' chief editors submit to the IRG Rapporteur responses to the consolidated comments

On a related note, the changes that I proposed in IRG N2182, specifically to remove the FS (*First Stroke*) data from the IRG Working Set code charts, and to add the IDS (*Ideographic Description Sequence*) and TS (*kTotalStrokes*) data, were accepted. This is important, because it means that these two bits of metadata will be subject to more review as they will be put into the reviewers' faces.

IRG's P&P: Revised Unification Rules, Stroke Counting Rules & IVD

Per WG2 #64 Recommendation M64.11 (*Review of CJK Unification Rules*) in WG2 N4701, which stated "WG2 recommends that IRG reviews its CJK unification rules to minimize the number of glyph variants that are coded as separate characters," the IRG has been considering ways to limit the number of encoded variants starting from the next IRG Working set, and received two proposals that were discussed during this meeting:

- Japan submitted IRG N2216 that recommended ways to revise the unification rules to reduce the number of variants, but the conclusion was that the current rules and principles, which are effectively Annex S and the UCV/NUC lists, already cover this sufficiently, in terms of what Japan recommended.
- Henry Chan submitted IRG N2211 that went a bit further in that it recommended that particular classes of variants be rejected, such as one-off corruptions and so-called unorthodox transcriptions.

Per Recommendation IRG M48.3, a new version of the IRG P&P, Version 10, will include verbiage about this, though it will still need to be reviewed by IRG member bodies, and subsequently approved by WG2. Related to all of this, the UTC will need to consider expanding the scope of what is considered a unifiable variant in UTS #37 (*Unicode Ideographic Variation Database*), though its current wording can be construed to be sufficient.

IRG N2171, which proposed updated stroke-counting guidelines, was discussed in detail, including feedback. Per Recommendation IRG M48.5, a new document, which will incorporate the feedback, along with consolidating the contents of IRG N954AR, will be submitted as IRG N2221 by 2017-07-21.

My recommendations in IRG N2183 related to encouraging IVD use by member bodies and strongly discouraging the encoding of new CJK Compatibility Ideographs faced no objections and were therefore accepted by the IRG, and will be reflected in the next version of the IRG's P&P (Version 10).

In related news, Japan is planning to submit a second set of sequences to be registered in the Moji_Joho IVD Collection, and most of them apparently use Extension F base characters. I told Japan that they can submit the new sequences at their earliest convenience, because Extension F is now officially encoded in Unicode Version 10.0, which was released on the 2017-06-20, and announced during the IRG meeting on the morning of 2017-06-21.

Glyph Normalization

An updated version of Korea's glyph normalization rules (IRG N2154) was discussed, and per Recommendation IRG M48.8, a new version will be submitted as IRG N2226. At my suggestion, IRG N2226 will also be added to IWDS (*IRG Working Document Series*) for easy reference.

Unihan Database Topics

Henry Chan proposed in IRG N2206 new annotations for CJK Unified Ideographs, comparable to what is provided in the code charts for other characters. I pointed out that the Unihan Database is the right home for such information, and after some email exchanges with John Jenkins, we concluded that Henry should submit a proposal to the UTC. This will be in the form of new or updated Unihan Database properties, with the latter being preferred.

My proposal to update and add Unihan Database properties, IRG N2200, was discussed, and there were no strong objections.

CJK Supplementary Components

In an effort to reduce PUA and other non-standard component use in the IDS database, Henry Chan presented a proposal (IRG N2204) to encode approximately 50 ideograph-like components. After some discussion, and per Recommendation IRG M48.10, Henry was tasked to submit by 2017-07-14 a more comprehensive proposal (IRG N2225) that lists "CJK Supplementary Components" and statistics that show how many characters benefit from them.

Various Error Reports

Various error reports were submitted and discussed, and many of the ones that affect representative glyphs seemed to have been introduced in the Unicode Version 6.0 code charts due to font changes, mainly from Korea and TCA. When the Version 5.2 code charts were referenced, most of the errors were not present.

Unless otherwise noted, written responses from the appropriate member body are due prior to IRG #49 per Recommendation IRG M48.6 and Recommendation IRG M48.7:

IRG N2131: TCA needs further study of the 16 possible T-Source representative glyph errors. At least for U+8FBC, U+93BF, and U+22E3F, the representative glyphs are correct in the Unicode Version 5.2 code charts (and presumably earlier), but incorrect in CNS 11643-2007 and the Unicode Version 7.0 (and later) code charts.

IRG N2181: U+2C0C4 (Extension E; T-Source) and U+2DAA2 (Extension F; USAT-Source) appear to be duplicates, and SAT needs further study of U+2DAA2.

IRG N2197. I was tasked to ask Vietnam (Lee Collins is their representative) to check and confirm the V-Source representative glyphs for U+2017B, U+204AF, U+21846, U+21BC0, U+235F1, and U+244C7. U+24CB2 is a non-issue because it affects the UCS2003 representative glyph, but it can be added to ISO/ISO 10646 Annex P. Macao needs to check and confirm the M-Source representative glyph for U+2CC43. China will fix the G-Source representative glyph for U+2278B. TCA needs to check and confirm the T-Source representative glyphs for U+22E3F,

U+237E2, U+2386E, U+23A26, U+24120, U+24662, U+26C16, and U+281AF. There is no issue with U+2459B.

IRG N2199: Korea needs further study to determine whether to add K-Source references to U+665A, U+7D55, and U+2E569 (Extension F) as a **horizontal extension**. I am working with Korea offline about this.

IRG N2201: China will not change the G-Source representative glyph for U+3B05, but will change that of U+3B9D.

IRG N2202: China will add G-Source references to 13 TGH-2013 (通用规范汉字表 *Tōngyòng Guīfàn Hànzìbiǎo*) characters as a **horizontal extension**, but needs further study to determine whether to change the G-Source representative glyphs for U+2A8FB and U+2CE88.

IRG N2203: China agreed to update the G-Source representative glyph for U+4A76, but those for U+809E, U+9AAA, and U+4D1F do not need to be changed. The G-Source representative glyph for U+45A0 needs further study by both China and TCA.

IRG N2205: Japan will not change the J-Source representative glyph for U+3B27. Three IVSes in the Moji_Joho IVD collection share U+3B27 as their base character.

IRG N2207: TCA will propose to move the source reference of U+2F8FD (not cognate with U+6CCD), T6-2C51, to U+2DC09 (Extension F). This will orphan U+2F8FD, which means that the UTC needs to assign a UCI source reference.

IRG N2208: TCA needs further study to determine whether to move the T-Source of U+2F89B to U+38E3.

IRG N2209: TCA needs further study to determine whether the T-Source representative glyphs for U+5284 and U+5DD5 need to be changed. Hong Kong SAR does not need to change anything, mainly because there is a forthcoming overhaul of the H-Source representative glyphs to align them with the HKSCS-2016 standard that was published last month.

IRG N2210: Pending further discussion, U+2F879 (not cognate with U+5CC0) and U+2F936 (not cognate with U+753E) might be disunified, which will involve orphaning one or two CJK Compatiblity Ideographs that will require UCI-Source source references. This would also mean that the UTC can submit UTC-01159, which was withdrawn from IRG Working Set 2015, to IRG Working Set 2017.

IRG N2212: SAT needs further study to determine whether the USAT-Source representative glyph of U+2E30B (Extension F) is incorrect in that it is missing a dot.

IRG N2213: Korea needs further study of the K-Source representative glyphs for U+8C6C (same as U+27CEF), U+8641 (same as U+27144), and U+56CD (same as U+21155). (This is a Unicode Version 5.2 versus Version 6.0 and later issue.) The K-Source representative glyph of U+7A3D is the same as U+25874, and this is covered by UCV #341. I am working with Korea offline about this.

I privately asked China whether they would be willing to change the G-Source representative glyph of U+6FF9 to match PRC conventions, and they agreed. The background is that this is a GBK fill-in character for the G-column, and currently matches the original J-Source representative glyph that comes from JIS X 0212, but the J-Source representative glyph changed when

JIS X 0213 became preferred in the Unicode 7.0 code charts, and is now in a form that matches PRC conventions.

UNC (Urgently Needed Character) Proposals from China, TCA & Japan

China, TCA, and Japan each submitted UNC proposals that include three, two, and four characters, respectively.

China's three-character UNC proposal, IRG N2198, was discussed during UTC #151 last month, and as expected, it was not controversial and therefore accepted by the IRG.

TCA's two-character UNC proposal, IRG N2224, is a subset of China's UNC proposal, and was therefore accepted. I did point out that the same thing could have been accomplished via a **horizontal extension** to two of the three characters in China's UNC proposal, but a separate and completely overlapping UNC proposal was preferred, mainly for political reasons.

Japan's four-character UNC proposal, IRG N2214, was a bit more controversial and was eventually whittled down to only two characters, per the following actions:

JMJ-013582	Japan needs to discuss with their experts. The IRG wants to unify with U+67A9. Add to the Moji_Joho IVD collection if unified.
JMJ-057449	Accepted. Variant of U+5E2D.
JMJ-057586	Unified with U+6556 per UCV #278. Add to the Moji_Joho IVD collection.
JMJ-060040	Accepted.

As long as WG2 accepts these UNC proposals, the UTC will need to consider adding these characters to Unicode Version 11.0 by appending them to the URO, starting from U+9FEB.

IRG Working Set 2017 (aka Extension H)

Per Recommendation IRG M48.11, in order to limit the size of IRG Working Set 2017, **each submission shall not exceed 1,000 characters**. The estimates that were given during the meeting were as follows: UTC (approximately 100), UK (1,500 to 2,000), Korea (approximately 1,000), China (1,000 to 3,000), SAT (approximately 500), Japan (none), and TCA (approximately 1,000). In other words, China and UK will need to roll back their expectations. The UTC's submission will be IRG N2233. The following is the tentative schedule:

2017-09-15	IRG members' chief editors send their submissions to the IRG chief editor
2017-10-12	The IRG chief editor produces the review code chart for preliminary review during IRG #49

In other words, the UTC needs to approve a submission during the UTC #152 meeting.

Related to IRG Working Set 2017, I was very pleased to see explicit mention of the SJ/T 11239-2001 standard in China's Activity Report (IRG N2187). This standard includes 2,501 ideographs, and approximately half of them are already encoded. I also learned from Mr. Chen that while "GB" translates to *National Standard*, "SJ" means *Industrial Standard* (or *IT Standard*, based on 电子行业标准 *diànzǐ hángyè biāozhǔn*).

That is all.