Title: UTC Liaison Report from WG2

Date: 2011-06-14
Source: Peter Constable, Unicode Liaison to SC2
Action: For review by UTC, Unicode officers

A WG2 meeting was held in Helsinki, Finland from June 6 – 10, 2011. SC2 also met, on June 10. This document reports on select topics arising from those meeting that will be of interest for the Unicode Consortium.

For the complete WG2 resolutions, see L2/11-239. For SC2 resolutions, see L2/11-252.

New JTC1 Procedures and WG2 Working Methods

As has been mentioned in previous liaison reports, revised procedures for JTC1 have been put into effect with the result that certain stages of balloting take longer than before. This leads to a problem with the way that WG2 has operated in the past because it forces WG2 meetings to a less-frequent schedule that could slow down progress of amendments / new editions.

The US presented a document (WG2/N4061 = L2/11-192) suggesting that WG2 adopt more agile working methods—specifically:

- Permit the editor to resolve non-controversial ballot comments and issue multiple PDAM ballots between face-to-face meetings of WG2 (as the JTC1 procedures have always permitted).

- Explore forms of electronic collaboration, such as an e-mail discussion for WG2 experts or teleconferencing, to facilitate establishing consensus on issues.

This proposal was generally well received within WG2. There was some discussion about the kinds of changes the editor might make between PDAM drafts. In particular, there was some reluctance to allowing the editor to introduce new scripts without a WG2 resolution. In the end, there was consensus to allow the editor to introduce certain types of changes on his own, and to adopt use of electronic means of collaboration. The following SC2 resolution was adopted:

> RESOLUTION M17.09: Enhanced method for progressing PDAMs (approved unanimously)
>
> SC 2 endorses Resolution M58.24 of WG 2 to initiate a trial period starting after its meeting 58, encouraging the project editor to make use of a discussion list and teleconferencing facilities to arrive at dispositions to ballot comments, and issuing of any PDAM ballots (within the scope of current SC 2 projects and its subdivisions assigned to WG 2), between WG 2 face to face meetings. The SC 2 secretariat will provide the WG 2 Convener and Project Editor of 10646 with the needed support by establishing a discussion list. National bodies and liaison organizations are invited to nominate experts to participate in the discussions associated with ballot dispositions and related matters. Other subject matter experts may also be invited to participate at the discretion of the project editor.

This will allow work to progress on PDAMs between meetings, which in turn will allow WG2 to meet less frequently without as great a reduction in progress that schedule would otherwise impose. It will also, to some degree, reduce the backlog of agenda items to be covered at WG2 meetings. (There were over 150 documents on the agenda for the Helsinki meeting.)

There may be some consequences for UTC to consider. In particular, the period for a PDAM ballot can be as short as two months. There is a possibility, therefore, that a PDAM ballot will be issued and closed between two consecutive UTC meetings. In other words, UTC may on occasion need to prepare comments to be submitted from the Consortium (or submitted to L2 for consideration as US comments) between meetings.

Related to the new JTC1 process, it was noted that the new procedures allow no more than two amendments / corrigenda before a new edition must be prepared. There was consensus within WG2 and SC2 that this constraint is problematic for ISO/IEC 10646. For instance, it means it would be unlikely for there ever to be a fast-tracked amendment solely for the addition of a new currency symbol like the Rupee. There is, unfortunately, no real expectation that this constraint will be lifted.

Progress on ISO/IEC 10646 amendments and new editions

At this meeting, WG2 processed ballot comments on FCD text for the third edition, and on PDAM text for Amendment 1 to the third edition.

The FCD ballot for the third edition passed in favour of progressing to FDIS. No new characters were added to the third edition, though some characters and some named sequences were removed. Of those, most were not yet approved by UTC, though some were (40 Meroitic Cursive numbers were removed). See the Consent Docket (L2/11-253) for details.

The next stage for the third edition is the FDIS ballot. This has a two-month ballot period. This comes at an ambiguous time, however: the new JTC1 procedures start to apply to all JTC1 projects in July, and the new procedures require a five-month DIS ballot prior to the FDIS ballot. WG2 and SC2 have been proceeding on the assumption that the third edition can be completed under the old procedures, without a DIS ballot.

The situation for the third edition is further complicated by the status of review for Extension B multi-column charts: IRG members set a target of July 8 for providing new fonts and were planning on being able to review revised charts until the November IRG meeting. In other words, they were counting on the FDIS ballot not being issued until November or December of this year.

The problem with that plan is that it means the FDIS is submitted for distribution well after the July date when the new JTC 1 procedures take effect. There would be a risk, then, that a DIS ballot could be required, causing a one-year delay in the completion of the third edition and the potential for further technical changes. That would, in turn, present risks for the publication schedule of Unicode 6.1 and the synchronization of content with the third edition.

There was consensus within WG2 that a DIS ballot should be avoided and that a way should be found, if possible, to accommodate Extension B review. The project editor and SC2 secretariat will be working with ITTF to see if an FDIS ballot in November can be sanctioned.

WG2 approved 1565 new characters for addition in Amendment 1, bringing the total count to 1846. This includes 506 Wingding symbols. (Not all of the Webdings were approved at this time.)

Amendment 1 will follow the new JTC1 procedures, meaning that it will progress through PDAM, DAM and FDAM ballots. Because of all the new additions, a second PDAM ballot will be issued. The new WG2 working methods described above will also be put into operation with the PDAM stage for Amendment 1. Thus, it is expected that PDAM1.2 will have a two month ballot, that the editor will dispose of

comments, and that a third (and perhaps even a fourth) PDAM draft will be balloted before the next WG2 meeting.

Future meetings and anticipated timetable for amendments

With the new procedures, the schedule for WG2 meetings is likely to be driven by the schedule for DAM / DIS ballots. The following is a tentative schedule for future meetings and potential timetable for progress on amendments:

WG2 Meeting	Amd 1	Amd 2	4th Edition ??
June 2011 (Helsinki)	Continue work on PDAM	n/a	n/a
Feb 2012 (Mountain View)	Dispose of PDAM comments; authorize 5-month DAM ballot	Initiate Amd 2; authorize PDAM ballot(s)	n/a
Oct 2012 (Chiang Mai)	Ballot disposition for DAM; authorize FDAM (or proceed directly to publication)	Dispose of PDAM comments; authorize 5-month DAM ballot	Initiate 4th edition & authorize CD ballot(s)??
June? 2013 (Berlin)	n/a	Ballot disposition for DAM; authorize FDAM (or proceed directly to publication)	Dispose of CD comments & authorize DIS ballot??

This is, of course, speculative: there is no current plan of record for either Amendment 2 or for a 4th edition, and the rate at which either of those could progress could be other than suggested here. This is just one possible course of events.

Update to UTR #45 *U-Source Ideographs* requested

The US submitted a contribution (WG2/N4058 = L2/11-199) commenting on the need for source references and requesting that U sources be added whenever there is no other source reference. There was discussion of this request within WG2, with a variety of comments including the following (not verbatim):

- The need to have source references isn't explained in the US contribution.

- "UTC" source references would combine these "mistake" ideographs from legitimate ideographs documented in UTR #45.

The first point was addressed when it was pointed out that Clause 23.1 requires every unified ideograph to have a unique source reference, and that the 2nd edition (in which orphaned ideographs all share common source reference tags) is, therefore, not consistent. Regarding the second point, it was agreed that a distinct label "UCI" be created for unified ideographs that have lost all other source references.

In support of that consensus, the US delegation prepared a new document (WG2/N4111 = L2/11-243) that cites Clause 23.1 as a rationale and that proposes an update to UTR #45 to define the new "UCI" source reference label.

Related to this, WG2 adopted the following resolution:

> RESOLUTION M58.05 (New source identifiers for orphaned CJK ideographs): (Unanimously accepted)
>
> WG2 accepts the proposal for new source references of the form UCI-xxxxx, per document N4111, and requests the Unicode Consortium to update Unicode Technical Report #45, assigning UCI source labels for all the ideographs which have no other source reference in ISO/IEC 10646, coordinating with the project editor. Three CJK Unified Ideograph Extension C characters --- 2AD12, 2B089, and 2B08F (per item 5 in document N4021 from IRG) are also to be given UCI source labels. The project editor is instructed to add appropriate text to the standard.

Also, the new US contribution (N4111) states,

> In accordance with this proposal, the US will request that the Unicode Consortium revise UTR #45 "U-Source Ideographs" at the earliest possible opportunity to accommodate source references using the "UCI" label for this purpose.

WG2 is counting on these changes for the FDIS of the 3rd edition, which will go out for balloting by November at the latest. Even if the ballot isn't issued until November, though, the editor will need to refer to changes in UTR #45 sooner. Hence, it will be helpful if this work can be initiated sooner than the August UTC meeting.

WG2 concerns regarding updates to UTRs

ISO/IEC 10646 currently has normative references to certain UTRs (or UAXs, UTSs). Concerns were raised by the Japanese national body that the Unicode Consortium could update these specs outside of the ISO process, the implication being that normative content of ISO/IEC 10646 can be changed outside of the control of and without input from ISO member bodies.

In fact, the ISO directives (Part 2, clause 6.2.2) state the following requirements regarding normatively referenced documents published by other bodies:

> Documents published by other bodies may be referred to in a normative manner provided that

> a) the referenced document is recognized by the ISO and/or IEC committee concerned as having wide acceptance and authoritative status as well as being publicly available,

> b) the ISO and/or IEC committee concerned has obtained the agreement of the authors or publishers (where known) of the referenced document to its inclusion and to its being made available as required — the authors or publishers will be expected to make available such documents on request,

> c) the authors or publishers (where known) have also agreed to inform the ISO and/or IEC committee concerned of their intention to revise the referenced document and of the points the revision will concern, and

d) the ISO and/or IEC committee concerned undertakes to review the situation in the light of any changes in the referenced document.

(See also commentary on this clause in "Policy Concerning Normative References in ISO Publications".)

Thus, Unicode is not constrained from updating its specification provided we inform WG2 and SC2 of changes so they have opportunity to consider how the changes impact ISO/IEC 10646.

Now, the NB contribution in which concerns were raised did not reference this portion of the ISO procedures, nor were they mentioned during the discussion in WG2. Even so, the discussion was conducted in the spirit of cooperation suggested by the ISO procedures and that has been the norm in interactions between Unicode and SC2 for years.

In order to capture SC2's desire for cooperation from Unicode in this matter, the following resolution was adopted:

RESOLUTION M17.04: Coordination regarding Normatively Referenced Unicode Technical Reports (Unanimously accepted)

Recognizing the concerns expressed in the document WG 2 N 4092, SC 2 requests the Unicode Consortium to accept the following:

a. When proposing updates to any Unicode Technical report that is normatively referenced by ISO/IEC 10646, the Unicode Consortium will submit a notification of the planned update and/or the draft update text to SC 2, before the Consortium officially approves the update;

b. The Unicode Consortium will consider all feedback on such a planned update received as liaison contributions from SC 2, and expressing the opinions of SC 2 and/or its participating national body members regarding a planned update; and

c. In the event that the Unicode Consortium and SC 2 have different opinions on details of such an update, the Consortium will give full consideration to the SC 2 opinions and make its best effort to collaborate with SC 2 to reach a consensus

(I will note that the Japanese NB contacted me before their document was submitted to WG2 and accepted suggested changes in wording.)

I think it would be appropriate for Unicode to send a response to WG2 that acknowledges the concern, that references the relevant clause from the ISO Procedures, and that expresses our commitment to keep SC2 and WG2 informed of changes to the documents in question in accordance with the terms of the ISO Procedures.

UTC should also consider making it standard practice to submit draft updates of the UTRs in question to WG2.

It would also be appropriate to review the UTRs that are normatively referenced in ISO/IEC 10646 and to evaluate if normative references are really required, or whether stabilized references to particular versions would be appropriate.

IVD issues

The IRG Rapporteur gave a report on IRG activity (WG2/N4021 = L2/11-249) in which concerns were raised regarding the possibility that UTC might submit an IVD registration. The concern was that a registration coming from UTC might be seen to have a more *de jure* status than other registrations. The recently-prepared proposed update to UTS #37 has changes aimed at addressing such concerns. (To facilitate review by WG2 members, WG2 requested that a copy of PUUTS #37 be submitted to the WG2 register. That has been submitted along with an introductory cover page—WG2/N4109.) There still appears to be some gaps in mutual understanding with regard to UTS #37. Our IRG representatives should continue to dialog on these matters with other IRG members.

WG2 also considered a UK contribution (WG2/N4075 = L2/11-250) that added to the IRG comments in N4021. That document raised concerns regarding the "proposed Ideographic Variation Database (IVD) registration of 21 simplified CJK characters by the UTC" (L2/11-109). Unlike the IRG concerns voiced in N4021, however, the concern here appears to be with proposed change in approach to CJK in terms of the treatment of y-variants. That issue was not discussed at length in WG2, but was remanded to the IRG for consideration.

WG2 also discussed the Unicode liaison contribution (WG2/N4084 = L2/11-205) requesting WG2 to instruct the IRG to investigate the use of IVD registrations as a means of tracking CJK glyph variants and their metadata. This raised some questions as to purpose—both for using IVD in this way and for WG2 taking a resolution on the issue. In the end, no WG2 action was taken, but the IRG Rapporteur agreed to have the document considered within IRG.

File format of NUSI.txt and NameSequences.txt

Named UCS Sequence Identifiers were introduced into ISO/IEC 10646 in the 2nd edition. Initially, the drafts had a file using the same format as NamedSequences.txt. However, the text describing NUSI.txt (Clause 25) described the format as using USIs, defined in Clause 6.6 using a different format. Japan had pointed out the inconsistency, and so the editor was forced to change the format (in spite of US ballot comments on the CD of the 3rd edition to change revert the format).

For the Helsinki meeting, the US had a contribution (WG2/N4063 = L2/11-188) proposing changes in the definition of USIs in Clause 6.6 that would allow the format of NUSI.txt to be consistent with that definition while matching the format of NamedSequences.txt. There was some push back from other NBs on making the proposed changes to the definition of USIs, but a compromise was reached allowing the file format to be changed. Accordingly, the US delegation prepared a revised proposal document at the meeting (WG2/N4114 = L2/11-246), making changes in Clause 25 so that the file format of NUSI.txt would match that of NameSequences.txt.

On this issue, WG2 adopted the following resolution:

> RESOLUTION M58.09 (Format of NUSI.txt file): (Unanimously accepted)
>
> WG2 accepts the proposed change to clause 25 describing the format used for NUSI.txt data file, detailed in document N4114, and to change the data in NUSI.txt file to follow this new format. This change will allow the NUSI.txt files in ISO/IEC 10646 and the Unicode Standard to be identical.

Updates to Principles and Procedures

WG2 took up the proposed updates to the Principles and Procedures in WG2/N4049 (= L2/11-093R—c.f. agenda item B.1.7 from UTC #127). Most of the proposed changes were accepted, but one was not: the proposal that a question be added to the proposal summary form to collect information on stroke count for Han ideographs, which originated in UTC.

That proposal invoked several comments. On the one hand, the editor noted that radical / stroke-count information is needed since it is included in the CJK charts. On the other hand, various concerns with the proposal were raised, including that the proposal specified "stroke count" rather than "radical plus stroke count", and that the notion of a single stroke count for unified ideographs is problematic since variants can different in stroke count. Eventually, the discussion was cut off, and no action on this item was taken.

CJK experts within UTC may want to revisit this issue and consider a revised proposal that can avoid the kinds of objections raised in WG2.