

Title: UTC Liaison Report from WG2

Date: 2010-10-30

Source: Peter Constable, Unicode Liaison to SC2

Action: For review by UTC, Unicode officers

A WG2 meeting was held in Busan, South Korea from October 4 – 8, 2010. This document reports on select topics that will be of interest for the Unicode Consortium.

Progress on ISO/IEC 10646 amendments and new editions

At this meeting, WG2 processed ballot comments on CD text for the third edition, and work on Amendment 1 to the third edition was started. There was also a last-minute change made in the FDIS of the second edition; this is discussed below in the section on [CJK Chart Issues](#).

The CD ballot for the third edition passed in favour of progressing to FCD. WG2 also approved 61 characters for addition to the third edition; see the WG2 Consent Docket ([L2/10-384](#)) for details. Target dates for FCD and FDIS ballots are:

- FCD: 2010-12
- FDIS: 2011-09

IRG will continue reviewing source references and glyphs for CJK Extension B and will also be reviewing G and T columns for Extension C. Some revisions resulting from this review may be incorporated into the FCD.

WG2 has approved 281 characters for addition in Amendment 1. Note that some new characters requested by the US that we would have expected to go into this amendment were approved. See the WG2 Consent Docket ([L2/10-384](#)) for details. A PDAM ballot document is in preparation.

Amendment 1 will need to follow the new JTC1 procedures; this means that it will progress through PDAM, DAM, FDAM ballots rather than PDAM, FPDAM, FDAM. Whereas a FPDAM ballot required four months and was circulated only among SC2 members, a DAM ballot requires five months and is circulated among all JTC1 members.

(DIS approval requires a minimum of $\frac{1}{2}$ of the JTC1 P-members casting a vote, a favourable vote by $\frac{2}{3}$ of voting P-members, and no more than $\frac{1}{4}$ of the total votes cast being negative. Abstentions and negative votes not accompanied by technical reasons are excluded from the count for purposes of the $\frac{2}{3}$ / $\frac{1}{4}$ thresholds.)

There was some discussion of timetable for progress of Amendment 1. Note that under JTC1 rules, a PDAM ballot can be circulated for as little as three months,¹ and if there are no contentious comments requiring a ballot-resolution meeting, a second PDAM ballot or even a DIS ballot can be circulated without waiting to have a WG2 meeting first. Thus, there may be ways to manage the schedule for progress of amendments and new editions in JTC1 to facilitate alignment with releases of future editions of Unicode. (Synchronization with ISO/IEC 14651 and with UTS #10 may also be considerations—see

¹ See JTC 1 Supplement, JA.4.

further discussion below.) However, creative options are contingent on absence of contentious issues that would require ballot-resolution meetings.

In any case, UTC should start considering plans for the next release of Unicode and anticipated synchronization with amendments to the third edition. Note that, under revised JTC1 procedures, no more than two amendments to a given edition of a standard are permitted: after two amendments have been published, subsequent changes require preparation of a new edition.² Given the characters that have been approved by UTC but that were not approved by WG2 at this meeting, UTC might prefer to see a second PDAM ballot.

Progress on ISO/IEC 14651 and version synchronization for 14651 and 10646

Meeting 12 of OWG-Sort was also held in Busan during the week of WG2 #57. FPDAM ballot comments on amendment 2 were discussed and resolved with the ballot passing. In an exception to normal process, it was decided to incorporate INDIAN RUPEE SIGN at this stage. Rather than publishing an FDAM for Amendment 2, it was decided to incorporate amended text into an FDIS for a second edition. The second edition would be synchronized with Unicode 6.0 and Version 6.0 of UTS #10. An FDIS ballot is expected this fall.

There was some discussion of planning for future amendments or editions to 14651. It was considered a desirable goal to get the editions of 14651 and 10646 into sync at some point. To achieve this, some factors that would need to be taken into consideration include Unicode release cycles and also the stipulation in the ISO/IEC procedures that now apply to JTC1 that limits the number of amendments to a given edition of a standard to two. Alain Labonté will be preparing a draft recommendation to present at the next SC2 plenary, in Helsinki.

CJK chart issues

As mentioned in the last liaison report ([L2/10-199](#)) and referred to above, a review of Extension B characters is being conducted by IRG with a goal of having corrected multi-column charts for all CJK ideographs in the third edition.

To facilitate IRG review, Michel Suignard prepared working charts for Extension B that included a column for the single glyph that appeared in second edition in addition to the various region-specific columns. This was considered so helpful as to merit incorporation into the published standard. This change was approved by WG2 for the third edition.

Thus, the Extension B charts in the FCD for the third edition will include this additional column showing the unified glyph used in the second edition. The glyphs will be given pseudo-source references of the form “UCS11-2xxxx” where 2xxxx is the code point in the SIP. (See section 1.b of the WG2 Consent Docket, [L2/10-384](#), for discussion of other decisions related to pseudo-source references taken at WG2 #57.) UTC should consider whether to adopt the same format for Unicode code charts for Extension B.

At this WG2 meeting, concerns of a geo-political nature were raised regarding the column headings that appear in the URO charts in the FDIS for the second edition. Specifically, the headings can be perceived as references to countries, and the distinction presented between “C” and “H” could hence be seen as

² See clause 2.10.3 of the JTC1 Supplement, and clause 2.10.4 of the ISO/IEC Procedures.

implying that Hong Kong is a distinct nation from China. To accommodate these national body concerns, WG2 has decided to remove the headers C, J, K, V and H from the charts in the FDIS for the second edition. Since each glyph has a source reference displayed underneath it, the headers were considered redundant information and hence editorial. The charts will not be otherwise impacted. (In particular, the editor indicated that this would be in a manner that did not impact pagination.)

It is anticipated that there will be national body comments on the FCD for the third edition that may lead to revisions to headers or even a re-ordering of columns.

In view of the geopolitical nature of the issue, UTC may want to consider the possible implications for Unicode code charts, including those published for Unicode 6.0.

Glyph and annotation changes for keyboard symbols

ISO/IEC JTC1/SC35/WG1 submitted a document to WG1 (WG2/N3897 = [L2/10-351](#)) requesting a variety of changes related to keyboard symbols used in ISO/IEC 9995-7:2009 and in Amendment 1 of that standard. Changes include new characters, additional named character sequences, and glyph and annotation changes for some already-encoded characters.

New character additions were left to be considered at a subsequent meeting. (I will assume that UTC will have an agenda item to discuss new characters proposed in L2/10-351; hence, these are not discussed further here.) WG2 agreed to add the named character sequences in the FCD for the third edition; these are covered in the WG2 Consent Docket ([L2/10-384](#)).

WG2 also agreed to incorporate glyph and name changes into the FCD for the third edition. [L2/10-391](#) (= WG2/N3943) captures the specific details related to the FCD. (This doc references and depends on L2/10-351.)

[L2/10-432](#) (= WG2/N3948) is a liaison report back to SC35 summarizing the WG2 actions taken.

UTC should consider the glyph and annotation changes being made in the FCD for the third edition and determine what to do in relation to Unicode charts and names lists.

Principles and Procedures

WG2 discussed the proposals in L2/10-397 (= WG2/N3944) to incorporate certain changes in the Principles and Procedures document. The proposed changes relate to three issues:

- Contiguous encoding of decimal digits
- Gathering information about common characters used with a script (intended for use in ScriptExtensions.txt)
- Gathering information about confusable character pairs (intended for use in UTS #39, confusables.txt)

WG2 members had no particular concern with the first issue. On the other points, however, concerns were raised from some WG2 members:

- Common characters used with a script: The issues motivating ScriptExtensions.txt were new to WG2 members. Clarification on the information requested and its intended use was requested.
- Confusables: There was a concern of a burden being placed on proposal submitters to provide detailed information related to security topics with which they are not familiar. In discussion, attempts were made to explain that the intent was not to impose onerous requirements, though this did not sufficiently assuage concerns. Clarification was requested to make clear that the expectations can be easily met by proposal submitters.

In view of these requests, a revised document should be prepared providing the requested clarifications. V.S. Umamaheswaran took an action item for this.

Disunification of “Eng”

In ballot comments on the CD for the third edition, Norway objected to an annotation on U+014A LATIN CAPITAL LETTER ENG, and proposed the addition of a new case pair. Their concern is that, for purposes of Sami usage, there is only one acceptable glyph for this character, whereas the annotation refers to possible use of a glyph that is not considered acceptable for use in Sami. Their comment proposed a new case pair, allowing 014A to have unambiguous presentation using the Sami glyph. This comment was disposed as being out of scope for CD ballot comments, with an instruction to use normal procedures for proposing new characters. (See [L2/10-385](#) for discussion of Norway’s comments and their disposition.)

In subsequent email discussion outside of WG2, it appears that Norway may yet be unsatisfied with the current status, and it is certainly clear that the representative from Ireland would like to see characters disunified. UTC should anticipate that a proposal to this effect may come to WG2 for discussion at the Helsinki meeting.