Title: SC2 Liaison Report

Date: 2009-5-12
Source: Peter Constable (UTC/SC2 Liaison)
Distribution: UTC

WG2 Meeting 54 was held in Dublin April 20–24, 2009. There were several outcomes and issues raised that are of interest for UTC.

Emoji proposal

The US/UTC proposal to encode Emoji symbols was approved for encoding, with some modifications, and will be included in Amendment 8. It was known beforehand that there were some aspects of the proposal that were considered controversial. During a long ad hoc meeting, consensus was reached on the majority of issues. Some significant accommodations and concessions were made at the request of other participants:

- 66 compatibility Emoji symbols were withdrawn from the proposal by the US.

- No consensus was reached to include 10 regional indicator symbols in an amendment; the majority opinion was to leave these characters out of PDAM 8.

- Five culturally-iconic symbols (MOUNT FUJI, TOKYO TOWER, STATUE OF LIBERTY, SKETCH OF JAPAN, MOYAI) were renamed with abstract names, EMOJI COMPABILITY SYMBOL-1, etc. The glyphs were left as in the US/UTC proposal, however.

- A number of other glyph or name changes were made, 150 additional characters were added, and code points were re-assigned.

Altogether, there was a sense of accomplishment among WG2 members in achieving this result.

The PDAM document will include the private identifiers (e-NNN) used by the original proposers to facilitate their review.

UTC will want to review this outcome and PDAM 8 to consider work that will need to be done in preparation for the August UTC/L2 meeting when we prepare US ballot comments. The repertoire is reflected in L2/09-173.

Status of SC2 projects

At the Dublin meeting, FPDAM6 was approved to progress to FDAM balloting. A small number of technical changes were made. (See the consent docket for details.) Depending on exactly when the ballot is issued, the ballot period may be complete by the August UTC meeting. The repertoire for Amendment 6 is now final and stable.

PDAM 7 was approved to progress to FPDAM balloting. The Tangut script was removed from Amendment 7; there were also a small number of other technical changes. (See the consent docket for details.) Amendment 7 is on track for stabilization this fall.

A new amendment was initiated at this meeting: Amendment 8. This includes a number of significant repertoire additions, including CJK Extension D ("urgently-needed characters"), alchemical symbols, Ethiopc additions, Emoji symbols and others (including one more middle dot!).

The Project Editor has been preparing a Working Draft for a new edition of ISO/IEC 10646. At this meeting, WG2 approved this for a CD ballot within SC2. The plan is to issue an FCD ballot within SC2 at the end of 2009, and an FDIS ballot in JTC1 mid 2010. The draft for the new edition incorporates major changes, including bringing terminology and other text into closer synchronization with Unicode. Therefore, a careful review of the CD should be conducted.

Given the schedule for the new edition and for amendments 7 and 8, consideration is being given to not publishing these amendments but rather rolling their content directly into the new edition. UTC will likely want to synchronize Unicode 6.0 with the new edition of ISO/IEC 10646. Consideration will need to be given as to how to process new proposals from now until the publication of Unicode 6.0 and the new edition of 10646 as stability will be needed for the new edition.

Other synchronization issues

Among FPDAM 6 ballot comments, one national body expressed concern that there was a normative reference to Unicode 5.2 — specifically, Unicode 5.2 versions of UAX#9 and UAX#15 — which is not yet published, and not yet available for review by SC2 member bodies. A temporary workaround for FDAM 6 was provided, though this points to an on-going issue that will need to be considered as new versions of Unicode or new amendments of ISO/IEC 10646 are prepared. In principle, the issue will become even bigger with the publication of the new edition of 10646 as it will reference more parts of Unicode, including the entire UCD. Technically, amendments of 10646 that do not synchronize with a major or minor version of Unicode may be adding new characters with undefined or default-only character properties. This is particularly problematic in relation to normative properties that are supposed to be stable and immutable, such as character decompositions. For instance, since normalization is already normatively defined in 10646, then strictly speaking no amendment should be published with characters that require decomposition unless those decomposition mappings are published with that amendment.

We might assume that such character property issues matter only for implementers, and that implementers are likely to focus on Unicode releases rather than amendments of 10646. Potentially, though, this could lead to concerns being raised by SC2 members.

Korean issues

The Korean national body submitted a document requesting that a reference to KS X 1026-1:2007 be added in ISO/IEC 10646. The wording of the proposed text was problematic in stating that the Korean standard "must be followed". A liaison contribution was prepared and presented expressing concern with this change due to the conflict in relation to normalization between the Korean standard and 10646 (as well as Unicode). In the meeting, use of "must" wording was also mentioned as problematic. After some discussion, consensus was reached to include text of a purely informative nature mentioning guidelines in KS X 1026-1 for text interchange:

> NOTE 3 – Hangul text can be represented in several different ways in this standard. Korean Standard *KS X 10265-1: Information Technology – Universal Multiple-Octet Coded Character Set (UCS) – Hangul – Part 1, Hangul processing guide for information interchange*, provides guidelines on how to ensure interoperability in information interchange.

During WG2 Meeting 52, Korea drew attention to an apparent contradiction between Unicode and ISO/IEC 10646 with regard to the use of jamo filler characters, and requested national bodies to comment with a view to resolving this contradiction. For the Dublin meeting, a Liaison contribution was submitted requesting that 10646 be amended to clarify the use of jamos in a manner that is consistent with Unicode. This proposal was adopted, with slightly modified text, for incorporation into the working draft of the new edition:

> Replace the sentence:
>
> > "An incomplete syllable which starts with a Jungseong or a Jongseong shall be preceded by a CHOSEONG FILLER (0000 115F)."
>
> with the following two sentences:
>
> > "An incomplete syllable which starts with a Jungseong shall be preceded by a CHOSEONG FILLER (0000 115F). An incomplete syllable comprised of a Jongseong alone shall be preceded by a CHOSEONG FILLER (0000 115F) and a JUNGSEONG FILLER (0000 1160)."

The Korean delegation was very pleased with this combination of changes in 10646. For them, this is a welcome change, though concerns remain regarding the impact of normalization on Hangul text representation during data interchange. Therefore, there continues to be a need for UTC and the Editorial committee to evaluate possible changes that may mitigate Korean concerns while not impacting normalization.

CJKV issues

During the Dublin meeting, there was discussion of fonts needed from various national bodies in order to produce multi-column CJKV fonts and, in particular, the fact that delivery of some of the fonts is long overdue. It may be necessary to eliminate some columns from the multi-column charts as a result.

US/UTC contributions regarding the Ideographic Variation Database (IVD) were mentioned during the WG2 meeting, with the suggestion particularly for Japan that they review these documents as they consider compatibility ideographs they wish to have represented in the UCS. Japan indicated that they would review these documents. In private discussion with the Japanese delegation, I learned that they are giving serious consideration to our proposal that they register sequences in the IVD in lieu of requesting encoding of additional compatibility characters. They indicated that the US/UTC documents would aid in their internal discussions. It appears that any technical questions regarding the variation-selector mechanism itself may not be the *primary* concern for Japan at this point.