L2/07-407

Report on IRG meeting 29

John H. Jenkins, Apple Inc.

16 November 2007

The 29th meeting of the IRG (Ideographic Rapporteur Group) was held November 12–15, 2007 in San José, California, hosted by Adobe. I attended as Unicode liaison. Representatives were also present from the US, mainland China, Taiwan Computer Association, Hong Kong SAR, Japan, South Korea, and Vietnam. Members from the US included Ken Lunde and Eric Muller from Adobe, Richard Cook from UC Berkeley, Tom Bishop of Wenlin, Rick McGowan from Unicode (the US head of delegation). Mike Ksar and Dirk Meyer were also present for part of the meeting.

While the bulk of the meeting was given over to continuing editorial work on Extension D, there was significant discussion of other important issues.

One has to do with the creation of set of "urgently needed characters." Members were asked to provide data for this set with the intent that the overall size of the set would be small, ideally no more than a couple of hundred characters. Vietnam withdrew its proposal to help reach this goal, and the PRC removed a large number of characters from theirs. The Korean submission was rejected as being too large.

The UTC set includes 37 characters. Sixteen are characters which we originally submitted for Extension C but were deferred at the last minute to Extension D. The remaining 21 are ideographs from Adobe-Japan1-6. Adobe requires a ruling from the IRG on the unifiability of these ideographs. The intent is that those which are unifiable will be added to Adobe's registry of IVSs and the remainder will be encoded as expeditiously as possible.

The IRG rejected the suggestion that the urgently needed characters be processed before Extension D (with associated renaming of Extension D to Extension E). The current plan is for the two sets to be developed in parallel and to be individually submitted to WG2 when finished. It may therefore be possible for the urgently needed characters to be ready for encoding after IRG meeting 31 (next fall).

Another issue was the report from an ad hoc group on revising Annex S for clarification and coordination with actual current IRG practices. While this consisted mostly of revisiting the lists of unifiable and nonunifiable components contained in Annex S, there was also some rewording to make lucid what was once hard to understand.

Also discussed was the creation of a document outlining IRG principles and procedures. Mike Ksar addressed the IRG to explain the role of WG2's principles and procedures document and the need for a similar document used by the IRG. I also submitted a document outlining some of the concerns of UTC members concerning IRG processes, namely that there is a need to improve IRG record-keeping regarding unification work, and a need for easily accessible documents describing the properties of characters submitted for encoding (such as pronunciation and meaning). Both of these would significantly speed up the IRG's unification work.

As an aside, there was considerable concern expressed by UTC members present at the meeting that the current day-to-day operations of the IRG are too chaotic and need to be tightened up to prevent the recurrence of the sorts of QA issues we saw with Extension C.

In the end, Anan Yasuhiro of Microsoft and I were assigned the responsibility to draft such a document for the IRG, using the current WG2 document as a starting point. This document will be submitted for multiple rounds of review by IRG members before being submitted to WG2 for review at the next WG2 meeting and then submitted for consideration at the next IRG meeting after additional revision based on WG2 input. I will also submit the document to the UTC for comment.

Ken Lunde of Adobe gave his presentation on Ideographic Variation Sequences, Adobe's registered set, and Adobe's extension to the OpenType spec to support them. Richard Cook of UC Berkeley and Tom Bishop of Wenlin gave a demonstration of the current version of Wenlin and summarized the state of their efforts to make the technology and data available to IRG members.