JTC 1/SC2/WG2/IRG N1150
Date: 2005-11-28
ISO/IEC JTC 1/SC2/WG2/
Ideographic Rapporteur Group

Source / Contribution Identifier: US
Meeting: IRG#25 @ Berkeley, CA
Title: US Activity Report (Nov. 2005)
Status: Informational
Content: This document summarizes US National Body activities

US National Body Activities

(1) US continues to contribute to maintenance of the Unihan database
(<http://www.unicode.org/Public/UNIDATA/Unihan.html>), in conjunction with contributions from Unicode Consortium
members.

(2) US has prepared a detailed review of IRG N1137A (report of CJK Strokes ad-hoc, 2005-05-26), collating comments
against previous contributions (N1096, N1097: "Proposal to add a block of CJK Unified Basic Strokes to the UCS"), and
the currently encoded U+C130..U+C13F. This review document is given below.

(3) US is preparing contributions to proofing of C1 character set, including comments regarding suitability of
proposed (N1153,N1154) usage of IDS in proofing process.

(4) US is preparing contributions for Old Hanzi ad hoc.

Title: Review of N1137A (CJK Strokes)
Status: Informational
Content: This document provides a review of IRG N1137A (report of CJK Strokes ad-hoc, 2005-05-26)

This document is divided into 3 sections, as follows:

I. Unencoded stroke types that are not candidates for unification
II. Stroke-type sets that may or may not be unified
III. Whether to encode all stroke types in the stroke block, including those which are already encoded as CJK
characters and/or radicals?

- I. Unencoded stroke types that are not candidates for unification

The type pg appears to be the only one that clearly belongs in this category. It is the second stroke in U+4E44 乄. It seems clear that pg should be added to the stroke block starting at U+31C0.

- II. Stroke-type sets that may or may not be unified

(1) 丿 p (U+4E3F), 丿 wp, 丿 sp
    N1137A says to unify p and wp, but sp is distinct. sp is unencoded. Should sp then be added to the stroke block starting at U+31C0? (Or maybe U+4E3F is sp, and p is unencoded?)

(2) 乀 n (U+31CF), 乀 dn, 乁 pn, 乀 tn (U+4E40), 乀 tpn
    N1137A says to unify n, dn, and pn. N1137A keeps tn as a distinct type. (tn is already encoded as the character U+4E40.) N1137A says tpn may be unified with n, but it would make more sense to unify tpn with tn.

(3) 一 hg (U+4E5B), フ hp (U+31C7), 乛 hz (U+200CD), hxg
    N1137A makes hg and hp distinct, but says "may" unify hxg with hg or hp. Note that hz is also distinct and may be used in characters like 今 and 敢 (for which hxg was intended).

(4) 乚 pt, ph, pz, 乚 sz (U+200CA)
    N1137A says may unify pt and ph; pz is another name for the same stroke type. sz is distinct and can be used for characters like 牙 and 东. It appears that one new stroke type should be encoded, for the first stroke in 厶. We suggest the name pz, but pt would also be OK.

(5) szz (U+200D1), szxz
    N1137A says may unify szz and szxz. Another name for szxz is szp. Unification with szz seems OK.

(6) hpwg (U+4E59), hzwg (U+31C8), hxg (U+2E84)
    These three are all encoded (in three separate blocks: character, stroke, and radical). N1137A says may unify hzwg with hxg, but since they are already distinctly encoded (though in different blocks), their unification might be a moot point.

- III. Whether to encode all stroke types in the stroke block, including those which are already encoded as CJK characters and/or radicals?

For example, should type h (U+4E00 一) also be encoded in the stroke block that starts at U+31C0? There are some possible advantages to having all the CJK strokes in a single block. One advantage is to distinguish the meaning, for example, of 一 as a character, and 一 as a stroke type. In some contexts they might be displayed or processed differently. This would be analogous to the distinction between 一 (U+4E00) as a character and 一 (U+2F00) as a radical.

(END OF DOCUMENT)