

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2/IRG N1936A-R

Date: 2013-05-20

Source:	UTC
Title:	UTC/US Urgently-needed Character Submission
Meeting:	IRG #40
Status :	
Actions required	
Distribution:	IRG
Medium :	Electronic
Page:	
References:	

This UTC/US urgently-needed character submission consists of the following documents:

- IRG N1936A: This document
- IRG N1936B: Proposal summary form from the IRG P&P
- IRG N1936C: A table of the proposed characters and key attributes
- IRG N1936D: A Excel spreadsheet with character attributes
- IRG N1936E: UTC document L2/12-333
- IRG N1936F: A zip archive of bitmaps
- UTCHan.ttf: A font containing glyphs for all the characters

The table includes the following attributes:

- A. The U-source identifier
- B. The character's glyph
- C. The character's IDS (if any)
- D. The KangXi radical (number form)
- E. The KangXi radical (character form)
- F. The number of additional strokes
- G. The stroke type of the first additional stroke
- H. An indication of whether the character is for traditional Chinese (TC) or simplified Chinese (SC)
- I. A character of which the proposed character is a variant. In the case of SC characters, the variant is the traditional form(s) (if encoded)
- J. The code point of the variant character

The Excel spreadsheet includes the following attributes:

- A. The U-source identifier
- B. The PUA code point for the character's glyph in UTCHan.ttf
- C. The character's IDS (if any)
- D. The KangXi radical (number form)
- E. The number of additional strokes
- F. The stroke type of the first additional stroke
- G. An indication of whether the character is for traditional Chinese (TC) or simplified Chinese (SC)
- H. The code point of a character of which the proposed character is a variant. In the case of SC characters, the variant is the traditional form(s) (if encoded)
- I. The index of the character in L2/12-333 (IRG N1936E), if any

UTC document L2/12-333 (IRG N1936E) is a contribution from sinologist Andrew West. Characters UTC-00953 onward all have indices into L2/12-333 (IRG N1936E), which itself provides the detailed evidence for those characters.

The remaining character, UTC-00791, is well-known (see http://en.wikipedia.org/wiki/Biangbiang_noodles). It exists in numerous variants. We recommend that only one form be actually encoded, and that a future Ideographic Variation Sequence registration be used for the others.

Of the characters included here, six (UTC-00953 [IRG Ext. F v 1.0 07659], UTC-00960 [01075], UTC-00961 [01162], UTC-00967 [01142], UTC-01171 [07656], and UTC-01173 [03339]) are included in Extension F submissions from other IRG members. Per the IRG P&P, should this proposal be accepted by the IRG, they would be removed from consideration for Extension F.

Rational for “Urgent Need”

In general, the characters included here have sufficiently established current use that we feel it inappropriate to wait for encoding in Extension G, which would otherwise be the earliest opportunity.

In some cases, these characters are currently represented in Wikipedia using graphics. As a rule, presence on the Internet is not sufficient evidence for encoding. In these cases, however, other considerations are involved.

UTC-00791 is so well known that it actually has its own article in the *English* Wikipedia. Its extreme complexity has made it something of a curiosity. Although a curiosity, it is nonetheless in actual use and therefore a legitimate candidate for encoding.

UTC-00953 and UTC-01171 are both recent coinages for newly-synthesized chemical elements. As such, they are used in the periodic table of the elements and other common reference works

for chemistry and physics.

UTC-00954, UTC-00955, UTC-00957, and UTC-00958 are comparatively common fish and plant names, found in standard reference materials in both simplified and traditional Chinese.

UTC-00959 is required by the catalogue of the Morrison collection, a major Western collection of old Chinese books and manuscripts at the SOAS at London University.

The remainder are used in Chinese-language Tangut studies. The fact that they are not encoded is actively creating problems for Tangut scholars.