

ISO/IEC JTC 1/SC2/WG2 IRG Ideographic Rapporteur Group

Source/Contributor Identifier: The Unicode Consortium
Meeting: IRG #18 (Tokyo)
Title: Variation Data for Existing and Future Ideographs in
ISO/IEC 10646
Keywords: (none)
Status: For informational purposes
Short description: UTC intentions on how to deal with variations within Unban
Proposed Conclusion/Requested Action:
None yet
Arguments/Text of Contribution:
Follows

The Unicode Technical Committee at its 89th meeting, held in Mountain View, California in November 2001, discussed the issues of variations within the standard's ideographic collections.

Unicode has recently approved the addition of a block of 256 variation selectors to the standard (see <<http://www.unicode.org/unicode/alloc/Pipeline.html>>.) One of the main purposes for this collection is to make it possible to represent catalogued ideographs in plain text without their requiring separate code points.

The UTC will not approve in the future the explicit encoding of variant characters with their own code point unless there is a significant reason why this must be done. The intention is that variant characters should be represented by a combination of a base form and a variation selector, with a database made available indicating how this is to be done.

In particular, simplifications of traditional Chinese characters should be treated in this fashion.

The UTC will be submitting a paper to WG2 on the topic for consideration at the next WG2 meeting. We will ask that WG2 instruct the IRG to continue to catalogue variant characters within its Extension C work. (One will note that a number of the characters in our submission for Extension C, IRG N 849, are variant characters.) We will, however, ask that WG2 not assign these characters their own code points, but that they be represented in plain text via a base form and assigned variation selector.

We will also ask that WG2 instruct the IRG to provide a catalogue of existing variations within the ideographic repertoire of ISO/IEC 10646. Software vendors are finding it increasingly difficult to handle the broad range of variants within East Asian scripts in the absence of explicit directives from the standardization committees on how to do so. The designers of other standards, as well, are having to deal with issues of normalization and transformation, and need guidelines from WG2 and the UTC on how to deal with them. We feel, therefore, that the IRG should be prepared to handle this work if WG2 should approve the UTC's request on the subject.