

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1686

Date: 2010-06-18

Doc. Type:	Member body contribution
Title:	Unicode/US Activity Report
Source:	Unicode
Status:	Input to IRG
Action:	None
Distribution:	IRG Members and Ideographic Experts

1. Unicode 6.0.0

The Unicode Technical Committee has announced the availability of a beta version of Unicode 6.0.0 and invited public review. Public review will continue until 2 August 2010. Unicode 6.0.0 is expected to be published in late September 2010. Information on the beta is available at <http://www.unicode.org/versions/beta.html>.

Unicode 6.0.0 will have exactly the same character assignments as the forthcoming Second Edition of 10646, ISO/IEC 10646:2011. It includes the CJK Unified Ideographs Extension D block, as well as several other new blocks.

Unicode 6.0.0 will include an update to the Unihan database. The current version of the Unihan database is described in Unicode Standard Annex 38 at <http://www.unicode.org/reports/tr38/tr38-8.html>, and the data files for the Unihan database are available at <http://www.unicode.org/Public/6.0.0/ucd/Unihan-6.0.0d1.zip>. IRG members are welcome to review the contents of the Unihan database and make corrections or additions.

2. Update to Unicode Technical Report 45

Unicode Technical Report 45, available at <http://www.unicode.org/reports/tr45/>, is the official document describing U-source ideographs used by the IRG in its work. A new version of this document has been published, giving updates as needed for Unicode 6.0.0 and ISO/IEC 10646:2011. Most notably, all compatibility ideographs derived from Unicode now have explicit U-source identifiers.

3. The online Unihan database

The Unicode Technical Committee provides a Web-based front end to the Unihan database. Increased traffic at the Unicode web site has made it necessary to take this front end down for maintenance. It is expected to be available again in the near future.

4. Extension of the scope of Ideographic Variation Sequences

The use of Ideographic Variation Sequences is described in Unicode Technical Standard 37 <http://www.unicode.org/reports/tr37/>. The current rules regarding Ideographic Variation Sequences require that the variant glyphs be acceptable forms for the base glyph. Glyphs which are unifiable under Annex S rules are examples of such acceptable variations.

The Unicode Technical Committee is considering the possibility of extending the scope of Ideographic Variation

Sequences to include glyphs with different abstract shapes. Such glyphs are considered y-variants and non-unifiable under Annex S rules.

The investigation of this proposal is not yet begun. As the investigation proceeds, the UTC will keep the IRG informed of its progress. The UTC also anticipates consulting individual IRG members for advice and guidance.