

Title: Unicode Liaison Report

L2/09-382

Date: 2009-10-26

Source: Unicode Consortium

Status: Liaison contribution

Action: For review by WG2 experts

Distribution: WG2

The Unicode Consortium is pleased to report on-going progress in development of the Universal Character Set resulting from collaboration with SC2, as well as progress on the Unicode Standard and related standards and technologies.

Release of Unicode 5.2

[Unicode 5.2](#) was released in September, with content synchronized with ISO/IEC 10646:2003 plus amendments 1 through 6. A summary of the delta from Unicode 5.1, complete [code charts](#) and Unicode Character Database (UCD) [data files](#) are available online now. Also, the complete revised text will be made available online; chapters 1 to 8 are ready now; remaining chapters are in preparation and will be posted when ready.

Amendment 5 introduced the use of the multi-column chart format for CJK Extension C, and the multi-column format is used for all unified ideograph blocks in the draft for the new edition; similarly, code charts for Unicode 5.2 utilize a multi-column format for all CJK unified ideograph blocks.

Certain changes to the Unicode conformance clauses were made. Noteworthy is a change to clause C7, which was tightened to eliminate security risks resulting from deletion of non-characters from uninterpreted text strings.

Various additions were made in the UCD data files. These include some new character properties, two new joining groups for Arabic shaping behaviour, and the introduction of the file *CJKRadicals.txt* which provides a mapping from kRSUnicode property values to characters in the CJK Radicals block and to radical-only ideographs in the CJK Unified Ideographs block. Also new is the *BidiTest.txt* file which can be used to test conformance to the Unicode Bidi Algorithm.

Release of Unicode Collation Algorithm 5.2

[Version 5.2 of the Unicode Collation Algorithm](#) (UCA) has just been released. This update synchronizes with version 5.2 of the Unicode Standard and, correspondingly, with ISO/IEC 10646 Amendment 6.

Note: Amendment 1 of ISO/IEC 14651 synchronized with Amendment 4 of ISO/IEC 10646, and the plan for the next amendment of 14651 is to synchronize with Amendment 8 of 10646. Thus, there is no amendment of 14651 with which UCA 5.2 is synchronized.

Synchronization

At Unicode Technical Committee (UTC) meetings since WG2 meeting 54, UTC experts have reviewed progress in on-going SC2 work to incorporate appropriate changes into the Unicode Standard. UTC experts also continue to work on processing new US contributions for submission to WG2 as well as

contributions coming from other countries with a view to maintaining close synchronization between Unicode and ISO/IEC 10646.

The Unicode Consortium re-affirms the importance of synchronization between the Unicode Standard and ISO/IEC 10646. Businesses and organizations throughout the world make reference to one or the other standard. Accordingly, implementers throughout the world face requirements of interoperability and conformance with both of these standards. Synchronization of these two standards, therefore, has a significant impact and value for implementers worldwide.

In this regard, then, the Unicode Consortium welcomes and applauds the work of WG2 in preparation of a new edition of ISO/IEC 10646 that aligns concepts and terminology more closely with that of the Unicode Standard, thereby making it easier for consumers of the two standards to compare and coordinate specifications. Likewise, the references in ISO/IEC CD/10646 to the Unicode Character Database, the Unicode Bidirectional Algorithm and other portions of the Unicode Standard, and references in the Unicode Standard to ISO/IEC 10646, also are a great benefit to implementers that need to conform to both standards. It is recognized that a high degree of synchronization and cross-references presents challenges in preparation of new amendments or editions since it may be necessary for each standard to reference work on the other standard that is still in development. The benefits to users of the standards are well worth the difficulties that our technical committees face, however. Therefore, we encourage SC2 and WG2 to persevere in the direction being taken in the CD in spite of the challenges.

CJK Fonts

A critical issue impacting the work of editors of the Unicode Standard is access to fonts used for preparation of code charts. In particular, as we move to multi-column CJK formats, it is necessary to have access to fonts supplied by each of the CJK-source countries. The Unicode Consortium has rendered tools and services to SC2 for use in code chart preparation, yet Unicode editors have not been granted licenses to all relevant CJK fonts. This is a serious problem impacting production of the Unicode Standard and, as a result, puts sustained synchronization of Unicode and ISO/IEC 10646 at risk.

Therefore, the Unicode Consortium most earnestly requests that SC2 and WG2 endeavour to secure licensing terms for CJK fonts from *all* CJK-source countries that will permit use by the editors of Unicode as well as ISO/IEC 10646 for preparation and publication of the two standards in a synchronized manner.

We would also like to raise a general concern regarding potential risks for regression as new versions of fonts for CJK repertoires are introduced. We would encourage WG2 to consider ways in which quality might be ensured in order to avoid unintended problems with glyphs and character definition as the CJK repertoire evolves in future amendments and editions.

Planning for Unicode 6.0

The Unicode Consortium has been preparing for a major-version release, Unicode 6.0, in late 2010 with content synchronized with Amendment 8 of ISO/IEC 10646. As plans for preparation of a new edition of ISO/IEC 10646 are firmed up, the Unicode Consortium will incorporate those plans into their planning for Unicode 6.0. We request that WG2 continue work on Amendments 7 and 8 on a timeline for both to be completed and published by fall 2010.

CJK Ideograph Variations and Compatibility Characters

The Unicode Consortium notes the recent contribution from Japan regarding compatibility ideographs and the use of IVS sequences (N3706) and welcomes their decision to adopt the use of IVS sequences to represent CJK glyph variations. The Unicode Consortium would be glad to offer whatever assistance may be needed to Japan or other SC2 members in preparing registration submissions for the Ideographic Variation Database.

Common Locale Data Repository (CLDR) Version 1.7.1

An update release of the CLDR, Version 1.5.1, was released on June 29, 2009. This was a minor release with no new locales or new data elements. Work is in progress on version 1.8 of CLDR, with a target release date of March 2010.

The Unicode Consortium feels confident that National Bodies and experts represented in WG2 will find the CLDR offers useful benefits in enabling support in software products for languages and cultures from across the world. As always, experts in WG2 are invited to participate in the on-going development of CLDR. Current information on CLDR can be found on the Unicode Web site at <http://unicode.org/cldr/>.