



Universal Multiple Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Unicode Liaison Report
Source: Asmus Freytag, Unicode Liaison to WG2
Status: Liaison contribution
Action: For consideration by JTC1/SC2/WG2
Related:

For review at WG2 #45 meeting in Markham, Ontario

Loose Matching for Character names

To account for the requirement to have some hyphens be distinctive in character names, the UTC has adopted the following rule for loose character name matching. The UTC requests that WG2 notify the character name guidelines (Annex L, Rule 4) on the addition of non-colliding names to the following:

Ensure that character names are distinct even when the following are ignored: case, SPACE, LOW LINE, and all medial hyphens (except the hyphen in U+1180). (Medial hyphens are those between letters).

Character encoding proposals

The Unicode Consortium has received and reviewed many character encoding proposals since the last WG2 meeting in Mountain View, CA. Character proposals approved by the UTC have been submitted under separate cover.

Ballot comments

The Unicode Consortium supports the PDAM Ballot comments submitted by the US National Member Body, whose experts meet in joint session with the Unicode Technical Committee.

For Information

Font submissions policy

The Unicode Consortium has recently adopted a policy for receiving fonts and font data that have been submitted for use in the production of the Unicode Standard. These fonts are also used by the Unicode Consortium for related documents including its editorial support to WG2 in publishing editions of ISO/IEC 10646. The new font policy specifies the terms under which the Unicode Consortium may use the submitted fonts and establishes the requirement that submitters warrant that their submission do not infringe on the intellectual property rights of third parties.

In addition, and like WG2, The Unicode Consortium has adopted a policy that character or script proposals, for which a suitable font is not available, will not be published.

Unicode Public Review Issues

The Unicode Consortium has instituted a process whereby significant technical issues and drafts of technical documents are made available for public review. National Bodies are encouraged to provide input during this public review process. Public Review Issues are posted at irregular intervals at <http://www.unicode.org/review/>. A notification of pending issues is sent to the Unicode mailing list (see <http://www.unicode.org/> on how to subscribe) or to member representatives and liaisons.

Version 4.0.1 of the Unicode Standard Released

The Unicode® Consortium announced on March 31, 2004 a new update of the Unicode Standard, Version 4.0.1. This update represents a significant revision of its Unicode Character Database, widely used in software products. No new characters were added to the standard at this time—the total number of characters still stands at 96,382, which matches ISO/IEC 10646:2003. However, the information in the Character Database has been refined to improve the quality of text processing in all languages of the world.

This version of the Unicode Character Database includes the first major update of the CJK database (Unihan) in two years. The Unihan Database provides character properties, definitions, pronunciations, mappings, and other information for the CJK characters in the standard—the characters used in particular for Chinese, Japanese, and Korean. This update includes thousands of additions and corrections, including major new correlations with traditional Chinese and Japanese dictionary sources

This version of the Unicode Standard significantly improves the ability to interchange languages such as Arabic, Hebrew, Urdu, and Pashto, by limiting the overrides when using the Unicode Bidirectional Algorithm. This version also clarifies the implementation of such languages as Bengali and the relationship between base form letters and accent marks.

Full technical details regarding the Unicode Standard, Version 4.0.1 are published online at <http://www.unicode.org/versions/Unicode4.0.1/>.

Version 4.0.1 amends the book version of the Unicode Standard, Version 4.0, which was published by Addison-Wesley in September of 2003 (ISBN 0-321-18578-1)

Unicode Sponsors Locale Data Project–CLDR

The Unicode Consortium announced on April 21, 2004 that it will be hosting the Common Locale Data Repository project, formerly hosted by the Open i18n Group.

To support users in different languages, programs must not only use translated text, but must also be adapted to local conventions. These conventions differ by language or region and include the formatting of numbers, dates, times, and currency values, as well as support for differences in measurement units or text sorting order. Most operating systems and many application programs currently maintain their own repositories of locale data to support these conventions. But such data are often incomplete, idiosyncratic, or gratuitously different from program to program. In the age of the internet, software components must work together seamlessly, without the problems caused by these discrepancies.

The Common Locale Data Repository (CLDR) provides a general XML format (LDML) for the exchange of locale information for use in application and system software development, combined with a public repository for a common set of locale data generated in that format. CLDR will be managed by a dedicated technical committee of the Unicode Consortium.

The Unicode Consortium encourages all National Bodies to establish a liaison relationship with its new technical committee, LTC, and to review the locale data available in the repository. For more information about the project, see <http://www.unicode.org/cldr/>.

Unicode becomes ISO 15924/RA

ISO has appointed the Unicode Consortium as the Registration Authority for International Standard, Codes for the representation of names of scripts. Michael Everson of Everson Typography has been appointed Registrar by the Registration Authority.

The ISO 15924/RA receives and reviews applications for requesting new script codes and for the change of existing ones according to criteria indicated in the standard. It maintains an accurate list of information associated with registered script codes, processes updates of registered script codes, and distributes them on a regular basis to subscribers and other parties. Additions to the ISO 15924 codes for scripts will be announced on the Unicode discussion list. Discussion about ISO 15924 and script codes is welcome on this list.

Proposals for additions and changes can be made with the request form. If you have questions concerning ISO 15924 please contact the Registrar. For more information, see <http://www.unicode.org/iso15924/>