



Contents

- [Related Unicode](#)
- [Technical Standards](#)
- [Review and Feedback](#)
- [Notable Issues for Beta Reviewers](#)

7.0.0 Unicode Standard Annexes (proposed updates)

- [UAX #9, Unicode Bidirectional Algorithm](#)
- [UAX #11, East Asian Width](#)
- [UAX #14, Unicode Line Breaking Algorithm](#)
- [UAX #15, Unicode Normalization Forms](#)
- [UAX #24, Unicode Script Property](#)
- [UAX #29, Unicode Text Segmentation](#)
- [UAX #31, Unicode Identifier and Pattern Syntax](#)
- [UAX #34, Unicode Named Character Sequences](#)
- [UAX #38, Unicode Han Database \(Unihan\)](#)
- [UAX #41, Common References for Unicode Standard Annexes](#)
- [UAX #42, Unicode Character Database in XML](#)
- [UAX #44, Unicode Character Database](#)
- [UAX #45, U-Source Ideographs](#)

7.0.0 Unicode Technical Standards (proposed updates)

- [UTS #10, Unicode Collation Algorithm](#)

BETA Unicode 7.0.0

The next version of the Unicode Standard will be Version 7.0.0, planned for release in June, 2014. The major feature of this release is the addition of significant new repertoire to the standard: 2,834 new characters are encoded, including characters for 23 new scripts. There are also many additions to existing blocks, including several hundred new pictographs and symbols—many originating from the wingdings and webdings sets. The new currency sign for the ruble has also been encoded in this version.

A beta version of the 7.0.0 Unicode Character Database files is available for public review. We strongly encourage implementers to review the summary description, download the beta 7.0.0 Unicode Character Database files, and test their programs with the new data, well before the end of the beta period. It is especially important to review the [Notable Issues for Beta Reviewers](#).

We encourage users to check the code charts carefully to verify correctness of the new characters added to Unicode 7.0.0 and to ensure that there are no regressions in glyph shapes for previously encoded characters.

Summary description	Unicode 7.0.0
Unicode character database (UCD)	http , ftp
Summary of beta charts	Readme.txt
Single-block charts with yellow highlighting for new characters	delta charts
Single block charts for all of Unicode 7.0.0	http , ftp
Code charts - single download (95MB)	http , ftp
Auxiliary HTML charts for beta review	HTML charts

Related Unicode Technical Standards

In addition to the Unicode Standard proper, two other Unicode Technical Standards have significant text and data file updates that are correlated with the new additions for Unicode 7.0.0. Review of that text and data is also encouraged during the beta

[UTS #46, Unicode IDNA
Compatibility
Processing](#)

Earlier Versions

[Versions of the Unicode
Standard](#)

[Unicode 6.3.0](#)

[Unicode 6.2.0](#)

review period.

Unicode Collation Algorithm	Data files: http , ftp
Unicode IDNA Compatibility Processing	Data files: http , ftp

Review and Feedback

For guidance on how to focus your review, see the section [Notable Issues for Beta Reviewers](#).

Any feedback should be reported using the [contact form](#). Comments on the Unicode Standard Version 7.0.0 or the Unicode Character Database data files, should refer to the beta review [Public Review Issue #271](#). Comments on specific Version 7.0.0 UAXes and UTSeS should refer to the respective [Public Review Issue Numbers](#) for each document, where available.

The comment period ends April 28, 2014. All substantive technical comments must have been received by that date for consideration at the May UTC meeting. Editorial comments (typos, etc.) may be still submitted after that date for consideration in the final editorial work.

Note: *All beta files may be updated, replaced, or superseded by other files at any time. The beta files will be discarded once Unicode 7.0.0 is final. It is inappropriate to cite these files as other than a work in progress. **No products or implementations should be released based on the beta UCD data files -- use only the final, approved Version 7.0.0 data files, expected in [July 2014](#).***

The Unicode Consortium provides early access to updated versions of the data files and text to give reviewers and developers as much time as possible to ensure a problem-free adoption of Version 7.0.0.

The assignment of characters for Unicode 7.0.0 is now stable. There will be no further additions or modifications of code points and no further changes to character names. Please *do not* submit feedback requesting changes to code points or character names for Unicode 7.0.0, as such feedback is not actionable.

One of the main purposes of the beta review period is to verify and correct the preliminary character property assignments in the Unicode Character Database. Reviewers should check for

property changes to existing Unicode 6.3.0 characters, as well as the property values for the new Unicode 7.0.0 character additions. The Auxiliary HTML charts include the new characters highlighted in yellow, with names appearing when hovering over a cell. These charts may be useful for reviewing information such as the default collation order, Script property assignments, and so forth during beta review.

To facilitate verification of the property changes and additions, diffable XML versions of the Unicode Character Database are available. These XML files are dated, so that people can check the details of changes that occurred during the beta review period. The XML files are in the <http://www.unicode.org/Public/7.0.0/diffs/> directory. For more information, see the [diffs.readme.txt](#) file.

The beta review period is a good opportunity to add support for the new Unicode 7.0.0 characters in internal versions of software, so that software can be tested to verify that the new characters and property assignments do not cause problems when upgraded to Version 7.0.0 of Unicode.

Notable Issues for Beta Reviewers

Changes to Unicode Standard Annexes

Some of the Unicode Standard Annexes have modifications for Unicode 7.0.0, often in coordination with changes to character properties. Most notably for Unicode 7.0.0:

- Many new scripts were added to Table 4, Candidate Characters for Exclusion from Identifiers in [UAX #31, Unicode Identifier and Pattern Syntax](#).
- Definitions were added for *Unicode namespace* and the *Unicode namespace for character names* in [UAX #34, Unicode Named Character Sequences](#), and there was a major rewrite of Section 4, Names.
- There were numerous updates to [UAX #44, Unicode Character Database](#).

Core Specification Update

The core specification is undergoing extensive review, with reorganization and many additions for Version 7.0.0. Although the draft text for Version 7.0.0 is not yet available, specific reports of any technical or editorial issues in the [currently published core specification](#) are also welcome during the beta

review period. Such reports will be taken into consideration for corrections to the Version 7.0.0 draft. (Note: The Unicode Consortium has ongoing opportunities for subject-matter volunteers: experts interested in contributing to or editing relevant parts of the core specification or other Unicode specifications.)

Addition of a System for Writing Shorthand

Version 7.0.0 adds, for the first time, the encoding of a shorthand notational system in the standard. ("Shorthand notational system" in this sense refers to a system for writing shorthand, as used historically for taking dictation, or other fast manual handwriting.) See the Duployan block, U+1BC0..U+1BC9F, and the Shorthand Format Controls block, U+1BCA0..U+1BCAF. Shorthand notations introduce new classes of layout issues, and implementation of rendering will be difficult. The UTC is interested, in particular, in receiving any implementers' feedback on the correctness of character property assignments for Duployan and for shorthand format controls.

Other Issues

Please also check the following specific items carefully:

- This version of the Unicode Standard adds many new scripts, so implementations that process script data should be checked very carefully.
- There have been significant additions to script extensions. Implementations of script extensions should also be checked carefully.
- The new character repertoire is diverse and complex. It includes new case pairs and new uppercase letters which form case pairs with previously encoded lowercase letters. It also includes punctuation marks whose line-breaking and terminal-punctuation properties should be examined closely.
- Note that some of the newly added scripts, and in particular, Manichaean and Psalter Pahlavi, have complex shaping behavior. New `Joining_Group` property values have been defined for Manichaean. Two Manichaean letters have received the `Joining_Type` property value `L`, which previously had been assigned to only one Phags-pa character.
- Parsers of `@missing` directives should be aware of the directives added in `PropertyValueAliases.txt` for the default

values of the `General_Category`, `Lowercase_Mapping`, `Titlecase_Mapping`, and `Uppercase_Mapping` properties.

- In `PropertyValueAliases.txt`, the `@missing` directives for the default values of `Bidi_Paired_Bracket_Type` and `Jamo_Short_Name` were moved to the end of the enumerations of value aliases for those properties. The newly-added `@missing` directive for `General_Category` was similarly placed after the `General_Category` property value aliases. This placement should enable single-pass parsing.
- Parsers of `NamesList.txt` should take note of the fact that beginning with this version, the repertoire of characters found in subheads in the file extends beyond ASCII values. This may break certain assumptions about handling that data that parsers may be making. (Note that `NamesList.txt` has been posted in UTF-8 since Version 6.2, but for reasons having to do with tooling restrictions, the repertoire associated with various elements of the file is still limited to the Latin-1 range: U+0020..U+00FF.)
- According to the resolution of [PRI #251](#), three ranges of enclosed capital Latin alphabetic symbols, U+1F130..U+1F149, U+1F150..U+1F169, and U+1F170..U+1F189, were assigned the contributory binary property `Other_Uppercase` and, by derivation, the properties `Uppercase`, `Alphabetic`, and corresponding values of text segmentation properties.
- The `General_Category` and `Line_Break` property values of the two ornate parentheses, U+FD3E..U+FD3F, which do not mirror and which are input visually, have been swapped (`General_Category`: Ps ↔ Pe; `Line_Break`: OP ↔ CL) to correctly reflect their usage in RTL contexts.
- The `Script` property value of U+061C ARABIC LETTER MARK (ALM) was changed from `Arabic` to `Common` for consistent treatment with the similar bidirectional controls LRM and RLM.
- There are additional property changes listed in [UAX #44](#), Unicode Character Database that may affect some implementations.

The following blocks are new in Unicode 7.0.0. Check implementations carefully for any range or property value assumptions regarding these new blocks.

Range	Block Name
1AB0..1AFF	Combining Diacritical Marks Extended
A9E0..A9FF	Myanmar Extended-B

AB30..AB6F	Latin Extended-E
102E0..102FF	Coptic Epact Numbers
10350..1037F	Old Permic
10500..1052F	Elbasan
10530..1056F	Caucasian Albanian
10600..1077F	Linear A
10860..1087F	Palmyrene
10880..108AF	Nabataean
10A80..10A9F	Old North Arabian
10AC0..10AFF	Manichaean
10B80..10BAF	Psalter Pahlavi
11150..1117F	Mahajani
111E0..111FF	Sinhala Archaic Numbers
11200..1124F	Khojki
112B0..112FF	Khudawadi
11300..1137F	Grantha
11480..114DF	Tirhuta
11580..115FF	Siddham
11600..1165F	Modi
118A0..118FF	Warang Citi
11AC0..11AFF	Pau Cin Hau
16A40..16A6F	Mro
16AD0..16AFF	Bassa Vah
16B00..16B8F	Pahawh Hmong
1BC00..1BC9F	Duployan
1BCA0..1BCAF	Shorthand Format Controls
1E800..1E8DF	Mende Kikakui
1F650..1F67F	Ornamental Dingbats
1F780..1F7FF	Geometric Shapes Extended
1F800..1F8FF	Supplemental Arrows-C

General Issues

For current proposed updates to the particular UAXes, see [Proposed Updates for Standard Annexes](#) or use the links in the navigation bar on this page. Particular issues in the UAXes may also be the focus of specific [Public Review Issues](#). Each proposed textual change in a UAX is highlighted, so that you can focus your review on those sections if you have limited time. The changes are also listed in detail in the Modifications sections (linked from the table of contents of each document), and are summarized in [UAX changes](#), so you can check on those areas that might be of most interest.

Some links between beta documents and the proposed updates for UAXes will not work correctly during the beta review period. This is a known problem which does not need to be reported, as such links are links to the eventual final names or revision numbers for the released versions.

Stability

Certain character properties for newly assigned characters *cannot* be changed after the formal release of each version of the standard, because of the [Character Encoding Stability Policy](#). Such character property values need special attention during the beta review process, as they cannot be corrected after publication. These include:

- Any property affecting Unicode Normalization, including Decomposition_Mapping, Canonical_Combining_Class, and Composition_Exclusion.
- The determination of whether a character is included in identifiers (XID_Start, XID_Continue).
- Case mappings and case foldings.

Copyright © 1991–2014 Unicode, Inc.
All Rights Reserved
[Terms of Use](#)

Last updated: - Wednesday, April 09, 2014 4:36:08 PM - [Contact Us](#)