

The Unicode® Standard

Version 14.0 – Core Specification

To learn about the latest version of the Unicode Standard, see <https://www.unicode.org/versions/latest/>.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed with initial capital letters or in all capitals.

Unicode and the Unicode Logo are registered trademarks of Unicode, Inc., in the United States and other countries.

The authors and publisher have taken care in the preparation of this specification, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

The *Unicode Character Database* and other files are provided as-is by Unicode, Inc. No claims are made as to fitness for any particular purpose. No warranties of any kind are expressed or implied. The recipient agrees to determine applicability of information provided.

© 2021 Unicode, Inc.

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction. For information regarding permissions, inquire at <https://www.unicode.org/reporting.html>. For information about the Unicode terms of use, please see <https://www.unicode.org/copyright.html>.

The Unicode Standard / the Unicode Consortium; edited by the Unicode Consortium. — Version 14.0.

Includes index.

ISBN 978-1-936213-29-0 (<https://www.unicode.org/versions/Unicode14.0.0/>)

1. Unicode (Computer character set) I. Unicode Consortium.

QA268.U545 2021

ISBN 978-1-936213-29-0

Published in Mountain View, CA

September 2021

Appendix B

Unicode Publications and Resources

This appendix provides information about the Unicode Consortium and its activities, particularly regarding publications other than the Unicode Standard. The Unicode Consortium publishes a number of technical standards and technical reports. *Appendix B.2, Unicode Publications* describes the kinds of reports in more detail.

The Unicode website also has many useful online resources. *Appendix B.3, Other Unicode Online Resources*, provides a guide to the kinds of information available.

B.1 The Unicode Consortium

The Unicode Consortium was incorporated in January 1991, under the name Unicode, Inc., to promote the Unicode Standard as an international encoding system for information interchange, to aid in its implementation, and to maintain quality control over future revisions.

To further these goals, the Unicode Consortium cooperates with the Joint Technical Committee 1 of the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC JTC1). It holds a Class C liaison membership with ISO/IEC JTC1/SC2; it participates in the work of both JTC1/SC2/WG2 (the technical working group for the subcommittee within JTC1 responsible for character set encoding) and the Ideographic Research Group (IRG) of WG2. The Consortium is a member company of the InterNational Committee for Information Technology Standards, Technical Committee L2 (INCITS/L2), an accredited U.S. standards organization. Many members of the Unicode Consortium have representatives in many countries who also work with other national standards bodies. In addition, a number of organizations are Liaison Members of the Consortium. For a list, see the Unicode website.

Membership in the Unicode Consortium is open to organizations and individuals anywhere in the world who support the Unicode Standard and who would like to assist in its extension and widespread implementation. Full, Institutional, Supporting, and Associate Members represent a broad spectrum of corporations and organizations in the computer and information processing industry. For a list, see the Unicode website. The Consortium is supported financially solely through membership dues and donations.

The Unicode Technical Committee

The Unicode Technical Committee (UTC) is the working group within the Consortium responsible for the creation, maintenance, and quality of the Unicode Standard. The UTC follows an open process in developing the Unicode Standard and its other technical publications. It coordinates and reviews all technical input to these documents and decides their contents. For more information on the UTC and the process by which the Unicode Standard and the other technical publications are developed, see:

<https://www.unicode.org/consortium/utc.html>

Other Activities

Going beyond developing technical standards, the Unicode Consortium acts as registration authority for the registration of script identifiers under ISO 15924, and it has a technical committee dedicated to the maintenance of the Unicode Common Locale Data Repository (CLDR). The repository contains a large and rapidly growing body of data used in the locale definition for software internationalization. For further information about these and other activities of the Unicode Consortium, visit:

<https://www.unicode.org>

B.2 Unicode Publications

In addition to the Unicode Standard, the Unicode Consortium publishes Unicode Technical Standards and Unicode Technical Reports. These materials are published as electronic documents only and, unlike Unicode Standard Annexes, do not form part of the Unicode Standard.

A *Unicode Standard Annex* (UAX) forms an integral part of the Unicode Standard, but is published online as a separate document. The Unicode Standard may require conformance to normative content in a Unicode Standard Annex, if so specified in the Conformance chapter of that version of the Unicode Standard. The version number of a UAX document is always the same as the version of the Unicode Standard of which it forms a part.

A *Unicode Technical Standard* (UTS) is an independent specification. Conformance to the Unicode Standard does not imply conformance to any UTS.

A *Unicode Technical Report* (UTR) contains informative material. Conformance to the Unicode Standard does not imply conformance to any UTR. Other specifications, however, are free to make normative references to a UTR.

In the past, some normative material was published as Unicode Technical Reports. Currently, however, such material is published either as a Unicode Technical Standard or a Unicode Standard Annex.

The Unicode website is the source for the most current version of all three categories of technical reports:

<https://www.unicode.org/reports/>

B.3 Other Unicode Online Resources

The Unicode Consortium provides a number of online resources for obtaining information and data about the Unicode Standard as well as updates and corrigenda.

Unicode Online Resources

Unicode Website

<https://www.unicode.org>

Charts. The charts section of the website provides code charts for all of the Unicode characters, plus specialized charts for normalization, collation, case mapping, script names, and CJK unified ideographs.

<https://www.unicode.org/charts/>

Character Index. Online index by character name, to look up Unicode code points. This index also makes it easy to look up the location of scripts in the standard, and indexes common alternative names for characters as well.

<https://www.unicode.org/charts/charindex.html>

Character Statistics. Access to tables containing raw character counts for every version of the Unicode Standard. This page also provides access to counts of emoji for various versions, and to the history of addition of scripts to the standard.

<https://www.unicode.org/versions/stats/>

Conferences. The Internationalization and Unicode Conferences are of particular value to anyone implementing the Unicode Standard or working on internationalization. A variety of tutorials and conference sessions cover current topics related to the Unicode Standard, the World Wide Web, software, internationalization, and localization.

<https://www.unicode.org/conference/>

E-mail Discussion List. Subscription instructions for the public e-mail discussion list are posted on the Unicode website.

Emoji. This page contains a collection of information about emoji resources and the Unicode Emoji Subcommittee.

<https://www.unicode.org/emoji/>

Emoji Charts. This section of the Unicode website collects together a number of charts that illustrate various features of the emoji characters in the standard. Some of these charts can be considered as adjuncts to the regular code charts. They show emoji with colorful glyphs, as they are displayed on many platforms.

FAQ (Frequently Asked Questions). The FAQ pages provide an invaluable resource for understanding the Unicode Standard and its implications for users and implementers.

<https://www.unicode.org/faq/>

Glossary. Online listing of definitions for technical terms used in the Unicode Standard and other publications of the Unicode Consortium.

<https://www.unicode.org/glossary/>

Online Unicode Character Database. This page supplies information about the online Unicode Character Database (UCD), including links to documentation files and the most up-to-date version of the data files, as well as instructions on how to access any particular version of the UCD.

<https://www.unicode.org/ucd/>

Online Unihan Database. The online Unihan Database provides interactive access to all of the property information associated with CJK ideographs in the Unicode Standard.

<https://www.unicode.org/charts/unihan.html>

Pipeline. This page lists characters, standardized variation sequences, and named character sequences which have reached some level of approval and/or are in international balloting, but which have not yet been published in a version of the Unicode Standard. The pipeline provides some visibility about what characters will soon be in the standard.

<https://www.unicode.org/alloc/Pipeline.html>

Policies. These pages describe Unicode Consortium policies on stability, patents, and Unicode website privacy. The stability policies are particularly important for implementers, documenting invariants for the Unicode Standard that allow implementations to be compatible with future and past versions.

<https://www.unicode.org/policies/>

References. This online page lists sources and up-to-date references for the Unicode Standard, as well as resources by script.

<https://www.unicode.org/references/>

Roadmap. This section of the Unicode website provides a roadmap for planning future allocation of scripts and major blocks of symbols. The roadmap is organized by plane, and provides information about the locations of published, approved, and proposed blocks, often with links to current proposals. The roadmap provides the long term perspective on future work by the encoding committees.

<https://www.unicode.org/roadmaps/>

Unicode Common Locale Data Repository (CLDR). Machine-readable repository, in XML format, of locale information for use in application and system development.

<https://www.unicode.org/cldr/>

Updates and Errata. This page lists periodic updates with corrections of typographic errors and new clarifications of the text.

<https://www.unicode.org/errata/>

Versions. This page describes the version numbering used in the Unicode Standard, the nature of the Unicode character repertoire, and ways to cite and reference the Unicode Standard, the Unicode Character Database, and Unicode Technical Reports. It also specifies the exact contents of each and every version of the Unicode Standard, back to Unicode 1.0.0.

<https://www.unicode.org/versions/>

Where Is My Character? This page provides basic guidance to finding Unicode characters, especially those whose glyphs do not appear in the charts, or that are represented by sequences of Unicode characters.

<https://www.unicode.org/standard/where/>

How to Contact the Unicode Consortium

The best way to contact the Unicode Consortium to obtain membership information is via the website:

<https://www.unicode.org/contacts.html>

The website also lists the current telephone number. The Consortium's postal address is:

P.O. Box 391476
Mountain View, CA 94039-1476
USA