

The Unicode Standard

Version 7.0 – Core Specification

To learn about the latest version of the Unicode Standard, see <http://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.

Copyright © 1991–2014 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 <http://www.unicode.org/reporting.html>. For information about the Unicode terms of use, please see <http://www.unicode.org/copyright.html>.

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 7.0

Includes bibliographical references and index.

ISBN 978-1-936213-09-2 (<http://www.unicode.org/versions/Unicode7.0.0/>)

1. Unicode (Computer character set) I. Allen, Julie D. II. Unicode Consortium.

QA268.U545 2014

ISBN 978-1-936213-09-2

Published in Mountain View, CA

October 2014

Figures

Figure 1-1.	Wide ASCII.	2
Figure 1-2.	Unicode Compared to the 2022 Framework.	5
Figure 2-1.	Text Elements and Characters	11
Figure 2-2.	Characters Versus Glyphs	16
Figure 2-3.	Unicode Character Code to Rendered Glyphs	17
Figure 2-4.	Bidirectional Ordering	20
Figure 2-5.	Writing Direction and Numbers	20
Figure 2-6.	Typeface Variation for the Bone Character.	22
Figure 2-7.	Dynamic Composition	23
Figure 2-8.	Abstract and Encoded Characters	29
Figure 2-9.	Overlap in Legacy Mixed-Width Encodings	33
Figure 2-10.	Boundaries and Interpretation	34
Figure 2-11.	Unicode Encoding Forms	35
Figure 2-12.	Unicode Encoding Schemes	41
Figure 2-13.	Unicode Allocation	48
Figure 2-14.	Allocation on the BMP	49
Figure 2-15.	Allocation on Plane 1.	51
Figure 2-16.	Writing Directions.	53
Figure 2-17.	Combining Enclosing Marks for Symbols.	56
Figure 2-18.	Sequence of Base Characters and Diacritics	56
Figure 2-19.	Reordered Indic Vowel Signs	57
Figure 2-20.	Properties and Combining Character Sequences	57
Figure 2-21.	Stacking Sequences	57
Figure 2-22.	Ligated Multiple Base Characters.	60
Figure 2-23.	Equivalent Sequences	62
Figure 2-24.	Canonical Ordering.	63
Figure 2-25.	Types of Decomposables.	65
Figure 3-1.	Enclosing Marks.	112
Figure 4-1.	Positions of Common Combining Marks	168
Figure 5-1.	Two-Stage Tables.	195
Figure 5-2.	Normalization	204
Figure 5-3.	Consistent Character Boundaries.	215
Figure 5-4.	Dead Keys Versus Handwriting Sequence.	218
Figure 5-5.	Truncating Grapheme Clusters	219
Figure 5-6.	Inside-Out Rule	220
Figure 5-7.	Fallback Rendering	221
Figure 5-8.	Bidirectional Placement	222
Figure 5-9.	Justification.	222
Figure 5-10.	Positioning with Ligatures	224
Figure 5-11.	Positioning with Contextual Forms	225

Figure 5-12.	Positioning with Enhanced Kerning	225
Figure 5-13.	Sublinear Searching	230
Figure 5-14.	Uppercase Mapping for Turkish I	236
Figure 5-15.	Lowercase Mapping for Turkish I	236
Figure 5-16.	Casing of German Sharp S	237
Figure 6-1.	Overriding Inherent Vowels	258
Figure 6-2.	Forms of CJK Punctuation	262
Figure 6-3.	European Quotation Marks	269
Figure 6-4.	Asian Quotation Marks	271
Figure 6-5.	Examples of Ancient Greek Editorial Marks	279
Figure 6-6.	Use of Greek Paragraphos	279
Figure 6-7.	CJK Parentheses	281
Figure 7-1.	Alternative Glyphs in Latin	287
Figure 7-2.	Diacritics on <i>i</i> and <i>j</i>	290
Figure 7-3.	Vietnamese Letters and Tone Marks	290
Figure 7-4.	Variations in Greek Capital Letter Upsilon	302
Figure 7-5.	Coptic Numerals	309
Figure 7-6.	Georgian Scripts and Casing	316
Figure 7-7.	Tone Letters	321
Figure 7-8.	Double Diacritics	325
Figure 7-9.	Positioning of Double Diacritics	325
Figure 7-10.	Use of CGJ with Double Diacritics	325
Figure 7-11.	Interaction of Combining Marks with Ligatures	327
Figure 7-12.	Positioning of Combining Parentheses	328
Figure 7-13.	Use of Vertical Line Overlay for Negation	329
Figure 7-14.	Double Diacritics and Half Marks	330
Figure 8-1.	Distribution of Old Italic	343
Figure 9-1.	Directionality and Cursive Connection	361
Figure 9-2.	Using a Joiner	363
Figure 9-3.	Using a Non-joiner	363
Figure 9-4.	Combinations of Joiners and Non-joiners	364
Figure 9-5.	Placement of Harakat	364
Figure 9-6.	Arabic Year Sign	368
Figure 9-7.	Syriac Abbreviation	385
Figure 9-8.	Use of SAM	385
Figure 11-1.	Interpretation of Hieroglyphic Markup	426
Figure 12-1.	Dead Consonants in Devanagari	437
Figure 12-2.	Conjunct Formations in Devanagari	437
Figure 12-3.	Preventing Conjunct Forms in Devanagari	438
Figure 12-4.	Half-Consonants in Devanagari	439
Figure 12-5.	Independent Half-Forms in Devanagari	439
Figure 12-6.	Half-Consonants in Oriya	439
Figure 12-7.	Consonant Forms in Devanagari and Oriya	440
Figure 12-8.	Rendering Order in Devanagari	445
Figure 12-9.	Marathi Allographs	449

Figure 12-10.	Use of Apostrophe in Bodo, Dogri and Maithili.	450
Figure 12-11.	Use of Avagraha in Dogri	451
Figure 12-12.	Requesting Bengali Consonant-Vowel Ligature	457
Figure 12-13.	Blocking Bengali Consonant-Vowel Ligature	458
Figure 12-14.	Bengali Syllable tta	459
Figure 12-15.	Kssa Ligature in Tamil.	471
Figure 12-16.	Tamil Two-Part Vowels	472
Figure 12-17.	Vowel Reordering Around a Tamil Conjunct.	473
Figure 12-18.	Tamil Ligatures with i	474
Figure 12-19.	Spacing Forms of Tamil u.	475
Figure 12-20.	Tamil Ligatures with ra	475
Figure 12-21.	Traditional Tamil Ligatures with aa.	475
Figure 12-22.	Traditional Tamil Ligatures with o	476
Figure 12-23.	Traditional Tamil Ligatures with ai	476
Figure 12-24.	Vowel ai in Modern Tamil	476
Figure 13-1.	Tibetan Syllable Structure.	500
Figure 13-2.	Justifying Tibetan Tseks	509
Figure 13-3.	Mongolian Glyph Convergence	513
Figure 13-4.	Mongolian Consonant Ligation	514
Figure 13-5.	Mongolian Positional Forms	514
Figure 13-6.	Mongolian Free Variation Selector	515
Figure 13-7.	Mongolian Gender Forms.	517
Figure 13-8.	Mongolian Vowel Separator	518
Figure 14-1.	Consonant Ligatures in Brahmi	539
Figure 14-2.	Geographical Extent of the Kharoshthi Script	542
Figure 14-3.	Kharoshthi Number 1996	543
Figure 14-4.	Kharoshthi Rendering Example	544
Figure 14-5.	Phags-pa Syllable Om	550
Figure 14-6.	Phags-pa Reversed Shaping.	553
Figure 15-1.	Siddham Consonant Cluster.	568
Figure 15-2.	Splitting Large Conjunct Stacks in Grantha	583
Figure 16-1.	Common Ligatures in Khmer.	610
Figure 16-2.	Common Multiple Forms in Khmer	610
Figure 16-3.	Examples of Syllabic Order in Khmer	612
Figure 16-4.	Ligation in <i>Muul</i> Style in Khmer	613
Figure 16-5.	Pahawh Hmong Syllable Structure.	627
Figure 17-1.	Buginese Ligature.	635
Figure 17-2.	Writing dharma in Balinese	640
Figure 17-3.	Representation of Javanese Two-Part Vowels.	644
Figure 18-1.	Han Spelling	657
Figure 18-2.	Semantic Context for Han Characters	658
Figure 18-3.	Three-Dimensional Conceptual Model.	660
Figure 18-4.	CJK Source Separation	661
Figure 18-5.	Not Cognates, Not Unified.	662
Figure 18-6.	Ideographic Component Structure	662

Figure 18-7.	The Most Superior Node of an Ideographic Component	663
Figure 18-8.	Using the Ideographic Description Characters	673
Figure 18-9.	Japanese Historic Kana for e and ye	679
Figure 19-1.	Tifinagh Contextual Shaping	700
Figure 19-2.	Tifinagh Consonant Joiner and Bi-consonants	701
Figure 19-3.	Examples of N’Ko Ordinals	704
Figure 20-1.	Short Words Equivalent to Deseret Letter Names	720
Figure 21-1.	Examples of Specialized Music Layout	728
Figure 21-2.	Precomposed Note Characters	729
Figure 21-3.	Alternative Noteheads	729
Figure 21-4.	Augmentation Dots and Articulation Symbols	729
Figure 22-1.	Alternative Glyphs for Dollar Sign	739
Figure 22-2.	Alternative Glyphs for Numero Sign	742
Figure 22-3.	Wide Mathematical Accents	745
Figure 22-4.	Style Variants and Semantic Distinctions in Mathematics	745
Figure 22-5.	Easily Confused Shapes for Mathematical Glyphs	747
Figure 22-6.	CJK Ideographic Numbers	751
Figure 22-7.	Regular and Old Style Digits	753
Figure 22-8.	Alternate Forms of Vulgar Fractions	758
Figure 22-9.	Usage of Crops and Quine Corners	772
Figure 22-10.	Usage of the Decimal Exponent Symbol	774
Figure 23-1.	Prevention of Joining	803
Figure 23-2.	Exhibition of Joining Glyphs in Isolation	803
Figure 23-3.	Effect of Intervening Joiners	804
Figure 23-4.	Annotation Characters	824
Figure 23-5.	Tag Characters	828
Figure 24-1.	CJK Chart Format for the Main CJK Block	842
Figure 24-2.	CJK Chart Format for CJK Extension A	842
Figure 24-3.	CJK Chart Format for CJK Extension B	842
Figure 24-4.	CJK Chart Format for Compatibility Ideographs	843
Figure A-1.	Example of Rendering	846