

The Unicode Standard

Version 6.2 – 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 © 1991–2012 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 6.2.

Includes bibliographical references and index.

ISBN 978-1-936213-07-8 (<http://www.unicode.org/versions/Unicode6.2.0/>)

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

QA268.U545 2012

ISBN 978-1-936213-07-8

Published in Mountain View, CA

September 2012

Figures

Figure 1-1.	Wide ASCII	2
Figure 1-2.	Unicode Compared to the 2022 Framework.	4
Figure 2-1.	Text Elements and Characters	9
Figure 2-2.	Characters Versus Glyphs	12
Figure 2-3.	Unicode Character Code to Rendered Glyphs	13
Figure 2-4.	Bidirectional Ordering	15
Figure 2-5.	Writing Direction and Numbers	16
Figure 2-6.	Typeface Variation for the Bone Character.	17
Figure 2-7.	Dynamic Composition	18
Figure 2-8.	Abstract and Encoded Characters	22
Figure 2-9.	Overlap in Legacy Mixed-Width Encodings	25
Figure 2-10.	Boundaries and Interpretation	25
Figure 2-11.	Unicode Encoding Forms	26
Figure 2-12.	Unicode Encoding Schemes	31
Figure 2-13.	Unicode Allocation	36
Figure 2-14.	Allocation on the BMP	37
Figure 2-15.	Allocation on Plane 1.	39
Figure 2-16.	Writing Directions.	40
Figure 2-17.	Combining Enclosing Marks for Symbols.	42
Figure 2-18.	Sequence of Base Characters and Diacritics	42
Figure 2-19.	Reordered Indic Vowel Signs	43
Figure 2-20.	Properties and Combining Character Sequences	43
Figure 2-21.	Stacking Sequences	43
Figure 2-22.	Ligated Multiple Base Characters.	45
Figure 2-23.	Equivalent Sequences	47
Figure 2-24.	Canonical Ordering.	47
Figure 2-25.	Types of Decomposables.	49
Figure 3-1.	Enclosing Marks.	85
Figure 4-1.	Positions of Common Combining Marks	127
Figure 5-1.	Two-Stage Tables	147
Figure 5-2.	Normalization	152
Figure 5-3.	Consistent Character Boundaries.	158
Figure 5-4.	Dead Keys Versus Handwriting Sequence.	161
Figure 5-5.	Truncating Grapheme Clusters	161
Figure 5-6.	Inside-Out Rule	162
Figure 5-7.	Fallback Rendering	163
Figure 5-8.	Bidirectional Placement	163
Figure 5-9.	Justification.	164
Figure 5-10.	Positioning with Ligatures	165
Figure 5-11.	Positioning with Contextual Forms	166
Figure 5-12.	Positioning with Enhanced Kerning	167
Figure 5-13.	Sublinear Searching.	169
Figure 5-14.	Uppercase Mapping for Turkish I	173
Figure 5-15.	Lowercase Mapping for Turkish I	174
Figure 5-16.	Casing of German Sharp S	174
Figure 6-1.	Overriding Inherent Vowels	189
Figure 6-2.	Forms of CJK Punctuation	192
Figure 6-3.	European Quotation Marks	198

Figure 6-4.	Asian Quotation Marks	199
Figure 6-5.	Examples of Ancient Greek Editorial Marks	205
Figure 6-6.	Use of Greek Paragraphos	206
Figure 6-7.	CJK Parentheses	207
Figure 7-1.	Alternative Glyphs in Latin	213
Figure 7-2.	Diacritics on <i>i</i> and <i>j</i>	214
Figure 7-3.	Vietnamese Letters and Tone Marks	214
Figure 7-4.	Variations in Greek Capital Letter Upsilon	224
Figure 7-5.	Coptic Numerals	229
Figure 7-6.	Georgian Scripts and Casing	234
Figure 7-7.	Tone Letters	237
Figure 7-8.	Double Diacritics	240
Figure 7-9.	Positioning of Double Diacritics	240
Figure 7-10.	Use of CGJ with Double Diacritics	241
Figure 7-11.	Interaction of Combining Marks with Ligatures	241
Figure 7-12.	Use of Vertical Line Overlay for Negation	243
Figure 7-13.	Double Diacritics and Half Marks	244
Figure 8-1.	Directionality and Cursive Connection	250
Figure 8-2.	Using a Joiner	252
Figure 8-3.	Using a Non-joiner	252
Figure 8-4.	Combinations of Joiners and Non-joiners	252
Figure 8-5.	Placement of Harakat	253
Figure 8-6.	Arabic Year Sign	255
Figure 8-7.	Syriac Abbreviation	268
Figure 8-8.	Use of SAM	268
Figure 9-1.	Dead Consonants in Devanagari	281
Figure 9-2.	Conjunct Formations in Devanagari	282
Figure 9-3.	Preventing Conjunct Forms in Devanagari	282
Figure 9-4.	Half-Consonants in Devanagari	283
Figure 9-5.	Independent Half-Forms in Devanagari	283
Figure 9-6.	Half-Consonants in Oriya	283
Figure 9-7.	Consonant Forms in Devanagari and Oriya	284
Figure 9-8.	Rendering Order in Devanagari	288
Figure 9-9.	Marathi Allographs	291
Figure 9-10.	Use of Apostrophe in Bodo, Dogri and Maithili	292
Figure 9-11.	Use of Avagraha in Dogri	292
Figure 9-12.	Requesting Bengali Consonant-Vowel Ligature	297
Figure 9-13.	Blocking Bengali Consonant-Vowel Ligature	297
Figure 9-14.	Bengali Syllable tta	298
Figure 9-15.	Kssa Ligature in Tamil	307
Figure 9-16.	Tamil Two-Part Vowels	308
Figure 9-17.	Vowel Reordering Around a Tamil Conjunct	308
Figure 9-18.	Tamil Ligatures with <i>i</i>	309
Figure 9-19.	Spacing Forms of Tamil <i>u</i>	309
Figure 9-20.	Tamil Ligatures with <i>ra</i>	310
Figure 9-21.	Traditional Tamil Ligatures with <i>aa</i>	310
Figure 9-22.	Traditional Tamil Ligatures with <i>o</i>	310
Figure 9-23.	Traditional Tamil Ligatures with <i>ai</i>	311
Figure 9-24.	Vowel <i>ai</i> in Modern Tamil	311
Figure 10-1.	Tibetan Syllable Structure	327
Figure 10-2.	Justifying Tibetan Tseks	334
Figure 10-3.	Phags-pa Syllable Om	338
Figure 10-4.	Phags-pa Reversed Shaping	341
Figure 10-5.	Geographical Extent of the Kharoshthi Script	355

Figure 10-6.	Kharoshthi Number 1996	356
Figure 10-7.	Kharoshthi Rendering Example	356
Figure 10-8.	Consonant Ligatures in Brahmi	360
Figure 11-1.	Common Ligatures in Khmer	380
Figure 11-2.	Common Multiple Forms in Khmer	380
Figure 11-3.	Examples of Syllabic Order in Khmer	382
Figure 11-4.	Ligation in <i>Muul</i> Style in Khmer	382
Figure 11-5.	Buginese Ligature.	394
Figure 11-6.	Writing dharma in Balinese	397
Figure 11-7.	Representation of Javanese Two-Part Vowels.	400
Figure 12-1.	Han Spelling	412
Figure 12-2.	Semantic Context for Han Characters.	413
Figure 12-3.	Three-Dimensional Conceptual Model.	415
Figure 12-4.	CJK Source Separation	415
Figure 12-5.	Not Cognates, Not Unified.	416
Figure 12-6.	Ideographic Component Structure	417
Figure 12-7.	The Most Superior Node of an Ideographic Component	417
Figure 12-8.	Using the Ideographic Description Characters.	425
Figure 12-9.	Japanese Historic Kana for e and ye.	429
Figure 13-1.	Mongolian Glyph Convergence	443
Figure 13-2.	Mongolian Consonant Ligation	443
Figure 13-3.	Mongolian Positional Forms	443
Figure 13-4.	Mongolian Free Variation Selector	444
Figure 13-5.	Mongolian Gender Forms.	446
Figure 13-6.	Mongolian Vowel Separator	446
Figure 13-7.	Tifinagh Contextual Shaping	449
Figure 13-8.	Tifinagh Consonant Joiner and Bi-consonants.	450
Figure 13-9.	Examples of N’Ko Ordinals	452
Figure 13-10.	Short Words Equivalent to Deseret Letter Names	460
Figure 14-1.	Distribution of Old Italic.	469
Figure 14-2.	Interpretion of Hieroglyphic Markup	489
Figure 15-1.	Alternative Glyphs for Dollar Sign	494
Figure 15-2.	Alternative Glyphs for Numero Sign	497
Figure 15-3.	Wide Mathematical Accents	499
Figure 15-4.	Style Variants and Semantic Distinctions in Mathematics	499
Figure 15-5.	Easily Confused Shapes for Mathematical Glyphs	501
Figure 15-6.	CJK Ideographic Numbers	504
Figure 15-7.	Regular and Old Style Digits.	505
Figure 15-8.	Alternate Forms of Vulgar Fractions	509
Figure 15-9.	Usage of Crops and Quine Corners	518
Figure 15-10.	Usage of the Decimal Exponent Symbol	520
Figure 15-11.	Examples of Specialized Music Layout	538
Figure 15-12.	Precomposed Note Characters.	538
Figure 15-13.	Alternative Noteheads	539
Figure 15-14.	Augmentation Dots and Articulation Symbols.	539
Figure 16-1.	Prevention of Joining.	549
Figure 16-2.	Exhibition of Joining Glyphs in Isolation	550
Figure 16-3.	Effect of Intervening Joiners	550
Figure 16-4.	Annotation Characters	563
Figure 16-5.	Tag Characters	566
Figure 17-1.	CJK Chart Format for the Main CJK Block.	578
Figure 17-2.	CJK Chart Format for CJK Extension A	579
Figure 17-3.	CJK Chart Format for CJK Extension B	579
Figure 17-4.	CJK Chart Format for Compatibility Ideographs.	579

Figure A-1. Example of Rendering 582