

The Unicode Standard

Version 6.0 – 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 © 1991–2011 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 6.0.

Includes bibliographical references and index.

ISBN 978-1-936213-01-6 (<http://www.unicode.org/versions/Unicode6.0.0/>)

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

QA268.U545 2011

ISBN 978-1-936213-01-6

Published in Mountain View, CA

February 2011

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.	84
Figure 4-1.	Positions of Common Combining Marks	123
Figure 5-1.	Two-Stage Tables	141
Figure 5-2.	CJK Ideographic Numbers	146
Figure 5-3.	Normalization	147
Figure 5-4.	Consistent Character Boundaries.	153
Figure 5-5.	Dead Keys Versus Handwriting Sequence.	156
Figure 5-6.	Truncating Grapheme Clusters	156
Figure 5-7.	Inside-Out Rule	157
Figure 5-8.	Fallback Rendering	158
Figure 5-9.	Bidirectional Placement	158
Figure 5-10.	Justification.	159
Figure 5-11.	Positioning with Ligatures	160
Figure 5-12.	Positioning with Contextual Forms	161
Figure 5-13.	Positioning with Enhanced Kerning	161
Figure 5-14.	Sublinear Searching.	164
Figure 5-15.	Uppercase Mapping for Turkish I	168
Figure 5-16.	Lowercase Mapping for Turkish I	168
Figure 5-17.	Casing of German Sharp S	169
Figure 6-1.	Overriding Inherent Vowels	181
Figure 6-2.	Forms of CJK Punctuation	184

Figure 6-3.	European Quotation Marks	190
Figure 6-4.	Asian Quotation Marks	191
Figure 6-5.	Examples of Ancient Greek Editorial Marks	197
Figure 6-6.	Use of Greek Paragraphos.	197
Figure 6-7.	CJK Parentheses	199
Figure 7-1.	Alternative Glyphs in Latin	204
Figure 7-2.	Diacritics on <i>i</i> and <i>j</i>	206
Figure 7-3.	Vietnamese Letters and Tone Marks	206
Figure 7-4.	Variations in Greek Capital Letter Upsilon	215
Figure 7-5.	Coptic Numerals	221
Figure 7-6.	Georgian Scripts and Casing.	226
Figure 7-7.	Tone Letters	229
Figure 7-8.	Double Diacritics	232
Figure 7-9.	Positioning of Double Diacritics	232
Figure 7-10.	Use of CGJ with Double Diacritics.	233
Figure 7-11.	Interaction of Combining Marks with Ligatures	233
Figure 7-12.	Use of Vertical Line Overlay for Negation.	235
Figure 7-13.	Double Diacritics and Half Marks	236
Figure 8-1.	Directionality and Cursive Connection	242
Figure 8-2.	Using a Joiner.	243
Figure 8-3.	Using a Non-joiner	243
Figure 8-4.	Combinations of Joiners and Non-joiners	244
Figure 8-5.	Placement of Harakat	244
Figure 8-6.	Arabic Year Sign.	247
Figure 8-7.	Syriac Abbreviation	258
Figure 8-8.	Use of SAM.	258
Figure 9-1.	Dead Consonants in Devanagari	271
Figure 9-2.	Conjunct Formations in Devanagari	272
Figure 9-3.	Preventing Conjunct Forms in Devanagari.	272
Figure 9-4.	Half-Consonants in Devanagari.	273
Figure 9-5.	Independent Half-Forms in Devanagari	273
Figure 9-6.	Half-Consonants in Oriya.	273
Figure 9-7.	Consonant Forms in Devanagari and Oriya	274
Figure 9-8.	Rendering Order in Devanagari	278
Figure 9-9.	Marathi Allographs	281
Figure 9-10.	Use of Apostrophe in Bodo, Dogri and Maithili.	282
Figure 9-11.	Use of Avagraha in Dogri	282
Figure 9-12.	Requesting Bengali Consonant-Vowel Ligature	287
Figure 9-13.	Blocking Bengali Consonant-Vowel Ligature	287
Figure 9-14.	Bengali Syllable tta	287
Figure 9-15.	Kssa Ligature in Tamil.	297
Figure 9-16.	Tamil Two-Part Vowels	297
Figure 9-17.	Vowel Reordering Around a Tamil Conjunct.	298
Figure 9-18.	Tamil Ligatures with <i>i</i>	299
Figure 9-19.	Spacing Forms of Tamil <i>u</i>	299
Figure 9-20.	Tamil Ligatures with <i>ra</i>	300
Figure 9-21.	Tamil Ligatures with <i>aa</i>	300
Figure 9-22.	Tamil Ligatures with <i>o</i>	300
Figure 9-23.	Tamil Ligatures with <i>ai</i>	301
Figure 9-24.	Vowel <i>ai</i> in Modern Tamil	301
Figure 10-1.	Tibetan Syllable Structure.	316
Figure 10-2.	Justifying Tibetan Tseks	323
Figure 10-3.	Phags-pa Syllable Om	328
Figure 10-4.	Phags-pa Reversed Shaping.	331

Figure 10-5.	Geographical Extent of the Kharoshthi Script	341
Figure 10-6.	Kharoshthi Number 1996	341
Figure 10-7.	Kharoshthi Rendering Example	342
Figure 10-8.	Consonant Ligatures in Brahmi	345
Figure 11-1.	Common Ligatures in Khmer.	366
Figure 11-2.	Common Multiple Forms in Khmer	366
Figure 11-3.	Examples of Syllabic Order in Khmer	368
Figure 11-4.	Ligation in <i>Muul</i> Style in Khmer	368
Figure 11-5.	Buginese Ligature.	380
Figure 11-6.	Writing dharma in Balinese	383
Figure 11-7.	Representation of Javanese Two-Part Vowels.	386
Figure 12-1.	Han Spelling	398
Figure 12-2.	Semantic Context for Han Characters.	399
Figure 12-3.	Three-Dimensional Conceptual Model.	401
Figure 12-4.	CJK Source Separation	401
Figure 12-5.	Not Cognates, Not Unified.	402
Figure 12-6.	Ideographic Component Structure	403
Figure 12-7.	The Most Superior Node of an Ideographic Component	403
Figure 12-8.	Using the Ideographic Description Characters.	411
Figure 12-9.	Japanese Historic Kana for e and ye.	416
Figure 13-1.	Mongolian Glyph Convergence	429
Figure 13-2.	Mongolian Consonant Ligation	429
Figure 13-3.	Mongolian Positional Forms	429
Figure 13-4.	Mongolian Free Variation Selector	430
Figure 13-5.	Mongolian Gender Forms.	432
Figure 13-6.	Mongolian Vowel Separator	432
Figure 13-7.	Tifinagh Contextual Shaping	435
Figure 13-8.	Tifinagh Consonant Joiner and Bi-consonants.	436
Figure 13-9.	Examples of N’Ko Ordinals	438
Figure 13-10.	Short Words Equivalent to Deseret Letter Names	446
Figure 14-1.	Distribution of Old Italic.	455
Figure 14-2.	Interpretion of Hieroglyphic Markup	475
Figure 15-1.	Alternative Glyphs for Dollar Sign	478
Figure 15-2.	Alternative Glyphs for Numero Sign	480
Figure 15-3.	Wide Mathematical Accents	482
Figure 15-4.	Style Variants and Semantic Distinctions in Mathematics	483
Figure 15-5.	Easily Confused Shapes for Mathematical Glyphs	484
Figure 15-6.	Alternate Forms of Vulgar Fractions	485
Figure 15-7.	Usage of Crops and Quine Corners	496
Figure 15-8.	Usage of the Decimal Exponent Symbol	497
Figure 15-9.	Examples of Specialized Music Layout	515
Figure 15-10.	Precomposed Note Characters.	515
Figure 15-11.	Alternative Noteheads	516
Figure 15-12.	Augmentation Dots and Articulation Symbols.	516
Figure 16-1.	Prevention of Joining.	527
Figure 16-2.	Exhibition of Joining Glyphs in Isolation	527
Figure 16-3.	Effect of Intervening Joiners	528
Figure 16-4.	Annotation Characters	539
Figure 16-5.	Tag Characters	542
Figure 17-1.	CJK Chart Format for the Main CJK Block.	554
Figure 17-2.	CJK Chart Format for CJK Extension A	554
Figure 17-3.	CJK Chart Format for CJK Extension B	555
Figure 17-4.	CJK Chart Format for Compatibility Ideographs.	555
Figure A-1.	Example of Rendering.	558

