

This PDF file is an excerpt from The Unicode Standard, Version 5.2, issued and published by the Unicode Consortium. The PDF files have not been modified to reflect the corrections found on the Updates and Errata page (<http://www.unicode.org/errata/>). For information about more recent versions of the Unicode Standard see <http://www.unicode.org/versions/enumeratedversions.html>.

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.

The Unicode® Consortium is a registered trademark, and Unicode™ is a trademark of Unicode, Inc. The Unicode logo is a trademark of Unicode, Inc., and may be registered in some jurisdictions.

The authors and publisher have taken care in the preparation of this book, 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 © 1991–2009 Unicode, Inc.

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. For information regarding permissions, inquire at <http://www.unicode.org/reporting.html>. For terms of use, please see <http://www.unicode.org/copyright.html>.

Visit the Unicode Consortium on the Web: <http://www.unicode.org>

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 5.2.

Includes bibliographical references and index.

ISBN 978-1-936213-00-9 (<http://www.unicode.org/versions/Unicode5.2.0/>)

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

II. Unicode Consortium.

QA268.U545 2009

ISBN 978-1-936213-00-9

Published in Mountain View, CA

December 2009

# 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 . . . . .	14
Figure 2-5.	Writing Direction and Numbers . . . . .	16
Figure 2-4.	Bidirectional Ordering . . . . .	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 2-24.	Canonical Ordering. . . . .	47
Figure 2-23.	Equivalent Sequences . . . . .	47
Figure 2-25.	Types of Decomposables. . . . .	49
Figure 3-1.	Enclosing Marks. . . . .	84
Figure 4-1.	Positions of Common Combining Marks . . . . .	121
Figure 5-1.	Two-Stage Tables. . . . .	139
Figure 5-2.	CJK Ideographic Numbers . . . . .	144
Figure 5-3.	Normalization . . . . .	145
Figure 5-4.	Consistent Character Boundaries. . . . .	151
Figure 5-6.	Truncating Grapheme Clusters . . . . .	154
Figure 5-5.	Dead Keys Versus Handwriting Sequence. . . . .	154
Figure 5-7.	Inside-Out Rule . . . . .	155
Figure 5-8.	Fallback Rendering . . . . .	155
Figure 5-9.	Bidirectional Placement . . . . .	156
Figure 5-10.	Justification. . . . .	157
Figure 5-11.	Positioning with Ligatures . . . . .	158
Figure 5-12.	Positioning with Contextual Forms . . . . .	159
Figure 5-13.	Positioning with Enhanced Kerning . . . . .	160
Figure 5-14.	Sublinear Searching . . . . .	162
Figure 5-15.	Case Mapping for Turkish I . . . . .	166
Figure 5-16.	Casing of German Sharp S . . . . .	167
Figure 6-1.	Overriding Inherent Vowels . . . . .	179
Figure 6-2.	Forms of CJK Punctuation . . . . .	182

Figure 6-3.	European Quotation Marks . . . . .	188
Figure 6-4.	Asian Quotation Marks . . . . .	189
Figure 6-5.	Examples of Ancient Greek Editorial Marks . . . . .	195
Figure 6-6.	Use of Greek Paragraphos . . . . .	195
Figure 6-7.	CJK Parentheses . . . . .	196
Figure 7-1.	Alternative Glyphs in Latin . . . . .	200
Figure 7-3.	Vietnamese Letters and Tone Marks . . . . .	202
Figure 7-2.	Diacritics on <i>i</i> and <i>j</i> . . . . .	202
Figure 7-4.	Variations in Greek Capital Letter Upsilon . . . . .	211
Figure 7-5.	Coptic Numerals . . . . .	217
Figure 7-6.	Georgian Scripts and Casing . . . . .	222
Figure 7-7.	Tone Letters . . . . .	224
Figure 7-8.	Double Diacritics . . . . .	227
Figure 7-9.	Positioning of Double Diacritics . . . . .	228
Figure 7-10.	Use of CGJ with Double Diacritics . . . . .	228
Figure 7-11.	Interaction of Combining Marks with Ligatures . . . . .	229
Figure 7-12.	Use of Vertical Line Overlay for Negation . . . . .	230
Figure 7-13.	Double Diacritics and Half Marks . . . . .	231
Figure 8-2.	Using a Joiner . . . . .	239
Figure 8-3.	Using a Non-joiner . . . . .	239
Figure 8-1.	Directionality and Cursive Connection . . . . .	239
Figure 8-5.	Placement of Harakat . . . . .	240
Figure 8-4.	Combinations of Joiners and Non-joiners . . . . .	240
Figure 8-6.	Arabic Year Sign . . . . .	242
Figure 8-7.	Syriac Abbreviation . . . . .	252
Figure 8-8.	Use of SAM . . . . .	252
Figure 9-1.	Dead Consonants in Devanagari . . . . .	265
Figure 9-3.	Preventing Conjunct Forms in Devanagari . . . . .	266
Figure 9-2.	Conjunct Formations in Devanagari . . . . .	266
Figure 9-5.	Independent Half-Forms in Devanagari . . . . .	267
Figure 9-6.	Half-Consonants in Oriya . . . . .	267
Figure 9-4.	Half-Consonants in Devanagari . . . . .	267
Figure 9-7.	Consonant Forms in Devanagari and Oriya . . . . .	268
Figure 9-8.	Rendering Order in Devanagari . . . . .	272
Figure 9-9.	Marathi Allographs . . . . .	275
Figure 9-10.	Use of Apostrophe in Bodo, Dogri and Maithili . . . . .	276
Figure 9-11.	Use of Avagraha in Dogri . . . . .	276
Figure 9-12.	Requesting Bengali Consonant-Vowel Ligature . . . . .	280
Figure 9-14.	Bengali Syllable tta . . . . .	281
Figure 9-13.	Blocking Bengali Consonant-Vowel Ligature . . . . .	281
Figure 9-15.	Kssa Ligature in Tamil . . . . .	290
Figure 9-16.	Tamil Two-Part Vowels . . . . .	291
Figure 9-17.	Vowel Reordering Around a Tamil Conjunct . . . . .	291
Figure 9-18.	Tamil Ligatures with <i>i</i> . . . . .	292
Figure 9-19.	Spacing Forms of Tamil <i>u</i> . . . . .	292
Figure 9-20.	Tamil Ligatures with <i>ra</i> . . . . .	293
Figure 9-21.	Tamil Ligatures with <i>aa</i> . . . . .	293
Figure 9-22.	Tamil Ligatures with <i>o</i> . . . . .	293
Figure 9-23.	Tamil Ligatures with <i>ai</i> . . . . .	294
Figure 9-24.	Vowel <i>ai</i> in Modern Tamil . . . . .	294
Figure 10-1.	Tibetan Syllable Structure . . . . .	308
Figure 10-2.	Justifying Tibetan Tseks . . . . .	315
Figure 10-3.	Phags-pa Syllable Om . . . . .	320
Figure 10-4.	Phags-pa Reversed Shaping . . . . .	323

Figure 10-6.	Kharoshthi Number 1996 . . . . .	333
Figure 10-5.	Geographical Extent of the Kharoshthi Script . . . . .	333
Figure 10-7.	Kharoshthi Rendering Example . . . . .	334
Figure 11-1.	Common Ligatures in Khmer. . . . .	354
Figure 11-2.	Common Multiple Forms in Khmer . . . . .	354
Figure 11-4.	Ligation in <i>Muul</i> Style in Khmer . . . . .	356
Figure 11-3.	Examples of Syllabic Order in Khmer . . . . .	356
Figure 11-5.	Buginese Ligature. . . . .	368
Figure 11-6.	Writing dharma in Balinese . . . . .	371
Figure 12-1.	Han Spelling . . . . .	386
Figure 12-2.	Semantic Context for Han Characters. . . . .	386
Figure 12-3.	Three-Dimensional Conceptual Model. . . . .	388
Figure 12-4.	CJK Source Separation . . . . .	389
Figure 12-5.	Not Cognates, Not Unified. . . . .	390
Figure 12-6.	Ideographic Component Structure . . . . .	390
Figure 12-7.	The Most Superior Node of an Ideographic Component . . . . .	390
Figure 12-8.	Using the Ideographic Description Characters. . . . .	398
Figure 13-1.	Mongolian Glyph Convergence . . . . .	415
Figure 13-2.	Mongolian Consonant Ligation . . . . .	415
Figure 13-3.	Mongolian Positional Forms . . . . .	415
Figure 13-4.	Mongolian Free Variation Selector . . . . .	416
Figure 13-5.	Mongolian Gender Forms. . . . .	417
Figure 13-6.	Mongolian Vowel Separator . . . . .	418
Figure 13-7.	Tifinagh Contextual Shaping . . . . .	421
Figure 13-8.	Examples of N’Ko Ordinals . . . . .	423
Figure 13-9.	Short Words Equivalent to Deseret Letter Names . . . . .	430
Figure 14-1.	Distribution of Old Italic. . . . .	439
Figure 14-2.	Interpretation of Hieroglyphic Markup . . . . .	457
Figure 15-1.	Alternative Glyphs for Dollar Sign . . . . .	460
Figure 15-2.	Alternative Glyphs for Numero Sign . . . . .	462
Figure 15-3.	Wide Mathematical Accents . . . . .	464
Figure 15-4.	Style Variants and Semantic Distinctions in Mathematics . . . . .	465
Figure 15-5.	Easily Confused Shapes for Mathematical Glyphs . . . . .	466
Figure 15-6.	Alternate Forms of Vulgar Fractions . . . . .	467
Figure 15-7.	Usage of Crops and Quine Corners . . . . .	477
Figure 15-8.	Usage of the Decimal Exponent Symbol . . . . .	479
Figure 15-9.	Examples of Specialized Music Layout . . . . .	493
Figure 15-10.	Precomposed Note Characters. . . . .	494
Figure 15-11.	Alternative Noteheads . . . . .	494
Figure 15-12.	Augmentation Dots and Articulation Symbols. . . . .	495
Figure 16-1.	Prevention of Joining. . . . .	505
Figure 16-2.	Exhibition of Joining Glyphs in Isolation . . . . .	505
Figure 16-3.	Effect of Intervening Joiners . . . . .	506
Figure 16-4.	Annotation Characters . . . . .	517
Figure 16-5.	Tag Characters . . . . .	520
Figure 17-1.	CJK Chart Format for the URO . . . . .	532
Figure 17-2.	CJK Chart Format for Extension A . . . . .	532
Figure 17-3.	CJK Chart Format for Extension B . . . . .	533
Figure A-1.	Example of Rendering. . . . .	538

