

Electronic Edition

This file is part of the electronic edition of *The Unicode Standard, Version 5.0*, provided for online access, content searching, and accessibility. It may not be printed. Bookmarks linking to specific chapters or sections of the whole Unicode Standard are available at

<http://www.unicode.org/versions/Unicode5.0.0/bookmarks.html>

Purchasing the Book

For convenient access to the full text of the standard as a useful reference book, we recommend purchasing the printed version. The book is available from the Unicode Consortium, the publisher, and booksellers. Purchase of the standard in book format contributes to the ongoing work of the Unicode Consortium. Details about the book publication and ordering information may be found at

<http://www.unicode.org/book/aboutbook.html>

Joining Unicode

You or your organization may benefit by joining the Unicode Consortium: for more information, see [Joining the Unicode Consortium](http://www.unicode.org/consortium/join.html) at

<http://www.unicode.org/consortium/join.html>

This PDF file is an excerpt from *The Unicode Standard, Version 5.0*, issued by the Unicode Consortium and published by Addison-Wesley. The material has been modified slightly for this electronic edition, however, 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 on more recent versions of the 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. *Dai Kan-Wa Jiten*, used as the source of reference Kanji codes, was written by Tetsuji Morohashi and published by Taishukan Shoten.

Cover and CD-ROM label design: Steve Mehallo, www.mehallo.com

The publisher offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales, which may include electronic versions and/or custom covers and content particular to your business, training goals, marketing focus, and branding interests. For more information, please contact U.S. Corporate and Government Sales, (800) 382-3419, corpsales@pearsoned.com. For sales outside the United States please contact International Sales, international@pearsoned.com

Visit us on the Web: www.awprofessional.com

Library of Congress Cataloging-in-Publication Data

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 5.0.
p. cm.

Includes bibliographical references and index.

ISBN 0-321-48091-0 (hardcover : alk. paper)

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

II. Unicode Consortium.

QA268.U545 2007

005.7'22—dc22

2006023526

Copyright © 1991–2007 Unicode, Inc.

All rights reserved. Printed in the United States of America. 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, write to Pearson Education, Inc., Rights and Contracts Department, 75 Arlington Street, Suite 300, Boston, MA 02116. Fax: (617) 848-7047

ISBN 0-321-48091-0

Text printed in the United States on recycled paper at Courier in Westford, Massachusetts.

First printing, October 2006

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	15
Figure 2-3.	Unicode Character Code to Rendered Glyphs	17
Figure 2-4.	Bidirectional Ordering	19
Figure 2-5.	Writing Direction and Numbers	20
Figure 2-6.	Typeface Variation for the Bone Character.	22
Figure 2-7.	Dynamic Composition	22
Figure 2-8.	Codespace and Encoded Characters	26
Figure 2-9.	Overlap in Legacy Mixed-Width Encodings	29
Figure 2-10.	Boundaries and Interpretation	29
Figure 2-11.	Unicode Encoding Forms	30
Figure 2-12.	Unicode Encoding Schemes	37
Figure 2-13.	Unicode Allocation	42
Figure 2-14.	Allocation on the BMP	44
Figure 2-15.	Allocation on Plane 1.	45
Figure 2-16.	Writing Directions.	47
Figure 2-17.	Combining Enclosing Marks for Symbols.	49
Figure 2-18.	Sequence of Base Characters and Diacritics	49
Figure 2-19.	Properties and Combining Sequences	49
Figure 2-20.	Reordered Indic Vowel Signs	50
Figure 2-21.	Stacking Sequences	50
Figure 2-22.	Ligated Multiple Base Characters.	53
Figure 2-23.	Equivalent Sequences	54
Figure 2-24.	Canonical Ordering.	55
Figure 2-25.	Types of Decomposables.	57
Figure 3-1.	Enclosing Marks.	112
Figure 4-1.	Positions of Common Combining Marks	134
Figure 5-1.	Two-Stage Tables.	153
Figure 5-2.	CJK Ideographic Numbers	159
Figure 5-3.	Normalization	160
Figure 5-4.	Consistent Character Boundaries.	168
Figure 5-5.	Dead Keys Versus Handwriting Sequence.	171
Figure 5-6.	Truncating Grapheme Clusters	172
Figure 5-7.	Inside-Out Rule	173
Figure 5-8.	Fallback Rendering	173
Figure 5-9.	Bidirectional Placement	174

Figure 5-10.	Justification.	175
Figure 5-11.	Positioning with Ligatures	177
Figure 5-12.	Positioning with Contextual Forms.	177
Figure 5-13.	Positioning with Enhanced Kerning	178
Figure 5-14.	Sublinear Searching.	181
Figure 5-15.	Case Mapping for Turkish I	186
Figure 6-1.	Overriding Inherent Vowels	200
Figure 6-2.	Forms of CJK Punctuation	203
Figure 6-3.	European Quotation Marks	210
Figure 6-4.	Asian Quotation Marks.	211
Figure 6-5.	Examples of Ancient Greek Editorial Marks.	217
Figure 6-6.	Use of Greek Paragraphos.	218
Figure 6-7.	CJK Parentheses.	219
Figure 7-1.	Alternative Glyphs in Latin.	227
Figure 7-2.	Diacritics on <i>i</i> and <i>j</i>	228
Figure 7-3.	Vietnamese Letters and Tone Marks	229
Figure 7-4.	Variations in Greek Capital Letter Upsilon.	239
Figure 7-5.	Coptic Numerals	245
Figure 7-6.	Georgian Scripts and Casing.	250
Figure 7-7.	Tone Letters	252
Figure 7-8.	Double Diacritics	255
Figure 7-9.	Positioning of Double Diacritics	255
Figure 7-10.	Use of CGJ with Double Diacritics.	256
Figure 7-11.	Interaction of Combining Marks with Ligatures	256
Figure 7-12.	Use of Vertical Line Overlay for Negation.	258
Figure 7-13.	Double Diacritics and Half Marks	259
Figure 8-1.	Directionality and Cursive Connection.	270
Figure 8-2.	Using a Joiner.	271
Figure 8-3.	Using a Non-joiner	271
Figure 8-4.	Combinations of Joiners and Non-joiners	271
Figure 8-5.	Placement of Harakat	272
Figure 8-6.	Arabic Year Sign.	274
Figure 8-7.	Syriac Abbreviation	285
Figure 8-8.	Use of SAM.	285
Figure 9-1.	Dead Consonants in Devanagari	300
Figure 9-2.	Conjunct Formations in Devanagari	300
Figure 9-3.	Preventing Conjunct Forms in Devanagari.	301
Figure 9-4.	Half-Consonants in Devanagari.	302
Figure 9-5.	Independent Half-Forms in Devanagari	302
Figure 9-6.	Half-Consonants in Oriya.	302
Figure 9-7.	Consonant Forms in Devanagari and Oriya	303
Figure 9-8.	Rendering Order in Devanagari	308
Figure 9-9.	Marathi Allographs	311
Figure 9-10.	Requesting Bengali Consonant-Vowel Ligature	314
Figure 9-11.	Blocking Bengali Consonant-Vowel Ligature	314

Figure 9-12.	Bengali Syllable tta	315
Figure 9-13.	Kssa Ligature in Tamil.	325
Figure 9-14.	Tamil Two-Part Vowels	326
Figure 9-15.	Vowel Reordering Around a Tamil Conjunct.	327
Figure 9-16.	Tamil Ligatures with i	327
Figure 9-17.	Spacing Forms of Tamil u.	328
Figure 9-18.	Tamil Ligatures with ra	328
Figure 9-19.	Tamil Ligatures with aa	329
Figure 9-20.	Tamil Ligatures with o.	329
Figure 9-21.	Tamil Ligatures with ai	330
Figure 9-22.	Vowel ai in Modern Tamil	330
Figure 10-1.	Tibetan Syllable Structure.	344
Figure 10-2.	Justifying Tibetan Tseks	352
Figure 10-3.	Phags-pa Syllable Om	356
Figure 10-4.	Phags-pa Reversed Shaping.	359
Figure 10-5.	Geographical Extent of the Kharoshthi Script	365
Figure 10-6.	Kharoshthi Number 1996	365
Figure 10-7.	Kharoshthi Rendering Example	366
Figure 11-1.	Common Ligatures in Khmer.	389
Figure 11-2.	Common Multiple Forms in Khmer	389
Figure 11-3.	Examples of Syllabic Order in Khmer	391
Figure 11-4.	Ligation in <i>Muul</i> Style in Khmer	392
Figure 11-5.	Buginese Ligature.	398
Figure 11-6.	Writing <i>dharma</i> in Balinese	402
Figure 12-1.	Han Spelling	414
Figure 12-2.	Semantic Context for Han Characters.	415
Figure 12-3.	Three-Dimensional Conceptual Model.	417
Figure 12-4.	CJK Source Separation	418
Figure 12-5.	Not Cognates, Not Unified.	419
Figure 12-6.	Ideographic Component Structure	420
Figure 12-7.	The Most Superior Node of an Ideographic Component	420
Figure 12-8.	Using the Ideographic Description Characters.	429
Figure 13-1.	Mongolian Glyph Convergence	451
Figure 13-2.	Mongolian Consonant Ligation	451
Figure 13-3.	Mongolian Positional Forms	452
Figure 13-4.	Mongolian Free Variation Selector	453
Figure 13-5.	Mongolian Gender Forms.	455
Figure 13-6.	Mongolian Vowel Separator	455
Figure 13-7.	Examples of N'Ko Ordinals	461
Figure 13-8.	Short Words Equivalent to Deseret Letter Names	466
Figure 14-1.	Distribution of Old Italic.	475
Figure 15-1.	Alternative Glyphs for Dollar Sign	491
Figure 15-2.	Alternative Glyphs for Numero Sign	493
Figure 15-3.	Wide Mathematical Accents	495
Figure 15-4.	Style Variants and Semantic Distinctions in Mathematics	496

Figure 15-5.	Easily Confused Shapes for Mathematical Glyphs	498
Figure 15-6.	Alternate Forms of Vulgar Fractions	499
Figure 15-7.	Usage of Crops and Quine Corners	509
Figure 15-8.	Examples of Specialized Music Layout	522
Figure 15-9.	Precomposed Note Characters	523
Figure 15-10.	Alternative Noteheads	524
Figure 15-11.	Augmentation Dots and Articulation Symbols	524
Figure 16-1.	Prevention of Joining	538
Figure 16-2.	Exhibition of Joining Glyphs in Isolation	539
Figure 16-3.	Effect of Intervening Joiners	539
Figure 16-4.	Annotation Characters	553
Figure 16-5.	Tag Characters	556
Figure A-1.	Example of Rendering	1078