

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

Tables

Table 2-1.	The 10 Unicode Design Principles	11
Table 2-2.	User-Perceived Characters with Multiple Code Points	12
Table 2-3.	Types of Code Points	23
Table 2-4.	The Seven Unicode Encoding Schemes	30
Table 2-5.	Interaction of Combining Characters	44
Table 2-6.	Nondefault Stacking	45
Table 3-1.	Named Unicode Algorithms	69
Table 3-2.	Normative Character Properties	74
Table 3-3.	Informative Character Properties	74
Table 3-4.	Examples of Unicode Encoding Forms	92
Table 3-5.	UTF-16 Bit Distribution	93
Table 3-6.	UTF-8 Bit Distribution	94
Table 3-7.	Well-Formed UTF-8 Byte Sequences	94
Table 3-8.	Use of U+FFFD in UTF-8 Conversion	96
Table 3-9.	Summary of UTF-16BE, UTF-16LE, and UTF-16	98
Table 3-10.	Summary of UTF-32BE, UTF-32LE, and UTF-32	99
Table 3-11.	Combining Marks and Starter Status	103
Table 3-12.	Reorderable Pairs	103
Table 3-13.	Hangul Syllable No-Break Rules	108
Table 3-14.	Korean Syllable Break Examples	109
Table 3-15.	Context Specification for Casing	113
Table 3-16.	Case Detection Examples	116
Table 4-1.	Relationship of Casing Definitions	121
Table 4-2.	Case Function Values for Strings	121
Table 4-3.	Sources for Case Mapping Information	122
Table 4-4.	Class Zero Combining Marks—Reordrant	124
Table 4-5.	Thai, Lao, and Tai Viet Logical Order Exceptions	124
Table 4-6.	Class Zero Combining Marks—Split	125
Table 4-7.	Class Zero Combining Marks—Subjoined	125
Table 4-8.	Class Zero Combining Marks—Strikethrough	126
Table 4-9.	General Category	127
Table 4-10.	Primary Numeric Ideographs	129
Table 4-11.	Ideographs Used as Accounting Numbers	129
Table 4-12.	Construction of Code Point Labels	133
Table 4-13.	Unusual Properties	136
Table 5-1.	Hex Values for Acronyms	148
Table 5-2.	NLF Platform Correlations	149
Table 5-3.	Typing Order Differing from Canonical Order	159
Table 5-4.	Permuting Combining Class Weights	160
Table 5-5.	Casing and Normalization in Strings	171
Table 5-6.	Paired Stateful Controls	176
Table 5-7.	Paired Stateful Controls (Deprecated)	177
Table 6-1.	Typology of Scripts in the Unicode Standard	183
Table 6-2.	Unicode Space Characters	186
Table 6-3.	Unicode Dash Characters	188
Table 6-4.	East Asian Quotation Marks	190
Table 6-5.	Opening and Closing Forms	191
Table 6-6.	Names for the @	195

Table 6-7.	Unicode Danda Characters	198
Table 7-1.	Nonspacing Marks Used with Greek	214
Table 7-2.	Greek Spacing and Nonspacing Pairs	218
Table 8-1.	Arabic Digit Names	245
Table 8-2.	Glyph Variation in Eastern Arabic-Indic Digits	245
Table 8-3.	Primary Arabic Joining Types	248
Table 8-4.	Derived Arabic Joining Types	248
Table 8-5.	Arabic Glyph Types	248
Table 8-6.	Arabic Obligatory Ligature Joining Groups	249
Table 8-7.	Arabic Ligature Notation	250
Table 8-8.	Dual-Joining Arabic Characters	251
Table 8-9.	Right-Joining Arabic Characters	252
Table 8-10.	Forms of the Arabic Letter yeh	253
Table 8-11.	Miscellaneous Syriac Diacritic Use	259
Table 8-12.	Syriac Final Alaph Glyph Types	260
Table 8-13.	Dual-Joining Syriac Characters	261
Table 8-14.	Right-Joining Syriac Characters	261
Table 8-15.	Syriac Alaph Glyph Forms	262
Table 8-16.	Syriac Ligatures	262
Table 8-17.	Samaritan Performative Punctuation Marks	264
Table 8-18.	Thaana Glyph Placement	265
Table 9-1.	Devanagari Vowel Letters	271
Table 9-2.	Sample Devanagari Half-Forms	279
Table 9-3.	Sample Devanagari Ligatures	280
Table 9-4.	Sample Devanagari Half-Ligature Forms	281
Table 9-5.	Prishthamatra Orthography	283
Table 9-6.	Bengali Vowel Letters	285
Table 9-7.	Bengali Consonant-Vowel Combinations	286
Table 9-8.	Use of Apostrophe in Bangla	289
Table 9-9.	Gurmukhi Vowel Letters	291
Table 9-10.	Gurmukhi Conjuncts	292
Table 9-11.	Additional Pairin and Addha Forms in Gurmukhi	292
Table 9-12.	Use of Joiners in Gurmukhi	292
Table 9-13.	Gujarati Vowel Letters	293
Table 9-14.	Gujarati Conjuncts	294
Table 9-15.	Oriya Vowel Letters	294
Table 9-16.	Oriya Conjuncts	295
Table 9-17.	Oriya Vowel Placement	295
Table 9-18.	Tamil Vowel Reordering	297
Table 9-19.	Tamil Vowel Splitting and Reordering	298
Table 9-20.	Tamil Ligatures with u	299
Table 9-21.	Tamil Vowels, Consonants, and Syllables	302
Table 9-22.	Telugu Vowel Letters	303
Table 9-23.	Kannada Vowel Letters	305
Table 9-24.	Malayalam Vowel Letters	307
Table 9-25.	Malayalam Orthographic Reform	307
Table 9-26.	Malayalam Conjuncts	308
Table 9-27.	Candrakala Examples	308
Table 9-28.	Atomic Encoding of Malayalam Chillus	309
Table 9-29.	Malayalam /rr/ and /tt/	310
Table 9-30.	Malayalam /nr/ and /nt/	310
Table 10-1.	Sinhala Vowel Letters	315
Table 10-2.	Lepcha Syllabic Structure	326
Table 10-3.	Phags-pa Positional Forms of I, U, E, and O	329

Table 10-4.	Contextual Glyph Mirroring in Phags-pa	330
Table 10-5.	Phags-pa Standardized Variants	331
Table 10-6.	Positions of Limbu Combining Characters	333
Table 10-7.	Kharoshthi Vowel Signs	342
Table 10-8.	Kharoshthi Vowel Modifiers	343
Table 10-9.	Kharoshthi Consonant Modifiers	344
Table 10-10.	Examples of Kharoshthi Virama	344
Table 10-11.	Brahmi Vowel Letters	345
Table 10-12.	Brahmi Positional Digits	347
Table 11-1.	Glyph Positions in Thai Syllables	351
Table 11-2.	Glyph Positions in Lao Syllables	353
Table 11-3.	Myanmar Syllabic Structure	357
Table 11-4.	Khamti Shan Tone Marks	358
Table 11-5.	Independent Khmer Vowel Characters	360
Table 11-6.	Two Registers of Khmer Consonants	362
Table 11-7.	Khmer Subscript Consonant Signs	363
Table 11-8.	Khmer Composite Dependent Vowel Signs with Nikahit	364
Table 11-9.	Khmer Subscript Independent Vowel Signs	365
Table 11-10.	Tai Le Tone Marks	369
Table 11-11.	Myanmar Digits	370
Table 11-12.	New Tai Lue Vowel Placement	371
Table 11-13.	New Tai Lue Registers and Tones	371
Table 11-14.	Tai Viet Symbols and Punctuation	375
Table 11-15.	Cham Syllabic Structure	377
Table 11-16.	Hanunóo and Buhid Vowel Sign Combinations	379
Table 11-17.	Balinese Base Consonants and Conjunct Forms	381
Table 11-18.	Sasak Extensions for Balinese	382
Table 11-19.	Balinese Consonant Clusters with u and u:	383
Table 11-20.	Sundanese Syllabic Structure	390
Table 12-1.	Sources for Unified Han	393
Table 12-2.	Blocks Containing Han Ideographs	395
Table 12-3.	Small Extensions to the URO	396
Table 12-4.	Common Han Characters	397
Table 12-5.	Source Encoding for Sword Variants	402
Table 12-6.	Ideographs Not Unified	403
Table 12-7.	Ideographs Unified	404
Table 12-8.	Han Ideograph Arrangement	404
Table 12-9.	Mandarin Tone Marks	413
Table 12-10.	Minnan and Hakka Tone Marks	414
Table 12-11.	Separating Jamo Characters	418
Table 12-12.	Line-Based Placement of Jungseong	419
Table 13-1.	Labialized Forms in Ethiopic -WAA	424
Table 13-2.	Labialized Forms in Ethiopic -WE	425
Table 13-3.	N’Ko Tone Diacritics on Vowels	437
Table 13-4.	Other N’Ko Diacritic Usage	438
Table 13-5.	N’Ko Letter Shaping	439
Table 13-6.	IPA Transcription of Deseret	446
Table 13-7.	Lisu Tone Letters	448
Table 13-8.	Punctuation Adopted in Lisu Orthography	449
Table 14-1.	Similar Characters in Linear B and Cypriot	460
Table 14-2.	Old South Arabian Numeric Characters	462
Table 14-3.	Number Formation in Old South Arabian	462
Table 14-4.	Number Formation in Aramaic	465
Table 14-5.	Dual-Joining Mandaic Characters	466

Table 14-6.	Right-Joining Mandaic Characters	467
Table 14-7.	Inscriptional Parthian Shaping Behavior	468
Table 14-8.	Avestan Shaping Behavior	469
Table 14-9.	Cuneiform Script Usage	472
Table 14-10.	Hieroglyphic Character Sequence	475
Table 15-1.	Currency Symbols Encoded in Other Blocks	479
Table 15-2.	Mathematical Alphanumeric Symbols	483
Table 15-3.	Use of Mathematical Symbol Pieces	497
Table 15-4.	Japanese Era Names	510
Table 15-5.	Examples of Ornamentation	516
Table 15-6.	Representation of Ancient Greek Vocal and Instrumental Notation	518
Table 16-1.	Control Codes Specified in the Unicode Standard	523
Table 16-2.	Letter Spacing	524
Table 16-3.	Bidirectional Ordering Controls	530
Table 16-4.	Unicode Encoding Scheme Signatures	538
Table 16-5.	U+FEFF Signature in Other Charsets	539
Table 17-1.	IRG Sources	553
Table A-1.	Extended BNF	559
Table A-2.	Character Class Examples	560
Table A-3.	Operators	560
Table C-1.	Timeline	570
Table C-2.	Zero Extending	573
Table D-1.	Versions of Unicode and ISO/IEC 10646-1	577
Table D-2.	Allocation of Code Points by Type	578
Table D-3.	Allocation of Code Points by Type (Early Versions)	578
Table D-4.	Version 5.1 Clause and Definition Updates	579
Table D-5.	Version 5.2 Clause and Definition Updates	579
Table D-6.	Version 6.0 Clause and Definition Updates	580
Table F-1.	CJK Strokes	590