The Unicode® Standard
Version 10.0 – 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.

© 2017 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 the Unicode Consortium. — Version 10.0.

Includes bibliographical references and index.

ISBN 978-1-936213-16-0 (http://www.unicode.org/versions/Unicode10.0.0/)

1. Unicode (Computer character set) I. Unicode Consortium.

QA268.U545 2017

ISBN 978-1-936213-16-0

Published in Mountain View, CA

June 2017

Tables

Table 2-1. The 10 Unicode Design Principles 14
Table 2-2. User-Perceived Characters with Multiple Code Points 16
Table 2-3. Types of Code Points 30
Table 2-4. The Seven Unicode Encoding Schemes 41
Table 2-5. Interaction of Combining Characters 58
Table 2-6. Nondefault Stacking 59
Table 3-1. Named Unicode Algorithms 94
Table 3-2. Normative Character Properties 100
Table 3-3. Informative Character Properties 101
Table 3-4. Examples of Unicode Encoding Forms 124
Table 3-5. UTF-16 Bit Distribution 125
Table 3-6. UTF-8 Bit Distribution 126
Table 3-7. Well-Formed UTF-8 Byte Sequences 126
Table 3-8. Use of U+FFFD in UTF-8 Conversion 129
Table 3-9. Summary of UTF-16BE, UTF-16LE, and UTF-16 132
Table 3-10. Summary of UTF-32BE, UTF-32LE, and UTF-32 133
Table 3-11. Combining Marks and Starter Status 138
Table 3-12. Reorderable Pairs 139
Table 3-13. Hangul Characters Used in Examples 145
Table 3-14. Context Specification for Casing 154
Table 3-15. Case Detection Examples 158
Table 4-1. Relationship of Casing Definitions 167
Table 4-2. Case Function Values for Strings 168
Table 4-3. Sources for Case Mapping Information 168
Table 4-4. General Category 175
Table 4-5. Primary Numeric Ideographs 178
Table 4-6. Ideographs Used as Accounting Numbers 178
Table 4-7. Types of Character Name Aliases 183
Table 4-8. Name Derivation Rule Prefix Strings 186
Table 4-9. Construction of Code Point Labels 188
Table 4-10. Unusual Properties 192
Table 5-1. Hex Values for Acronyms 211
Table 5-2. NLF Platform Correlations 212
Table 5-3. Typing Order Differing from Canonical Order 227
Table 5-4. Permuting Combining Class Weights 227
Table 5-5. Casing and Normalization in Strings 244
Table 6-1. Typology of Scripts in the Unicode Standard 264
Table 6-2. Unicode Space Characters 268
Table 6-3. Unicode Dash Characters 270
Table 6-4. Models of Visual Relationship between Quote Glyphs 273

Table 6-5. East Asian Quotation Marks 274
Table 6-6. Opening and Closing Forms 275
Table 6-7. Names for the @ 280
Table 6-8. Unicode Danda Characters 284
Table 7-1. Preferred Rendering of Cedilla versus Comma Below 294
Table 7-2. Nonspacing Marks Used with Greek 306
Table 7-3. Greek Spacing and Nonspacing Pairs 311
Table 7-4. Typicon Kavyka Symbols 337
Table 8-1. Similar Characters in Linear B and Cypriot 346
Table 8-2. Combining Marks Used in Old Permic 359
Table 9-1. Arabic Digit Names 375
Table 9-2. Glyph Variation in Eastern Arabic-Indic Digits 376
Table 9-3. Primary Arabic Joining Types 379
Table 9-4. Derived Arabic Joining Types 380
Table 9-5. Arabic Glyph Types 381
Table 9-6. Arabic Obligatory Ligature Joining Groups 382
Table 9-7. Arabic Ligature Notation 383
Table 9-8. Dual-Joining Arabic Characters 384
Table 9-9. Right-Joining Arabic Characters 386
Table 9-10. Forms of the Arabic Letter yeh 387
Table 9-11. Arabic Letters With Hamza Above 390
Table 9-12. Miscellaneous Syriac Diacritic Use 398
Table 9-13. Syriac Final Alaph Glyph Types 399
Table 9-14. Dual-Joining Syriac Characters 400
Table 9-15. Right-Joining Syriac Characters 401
Table 9-16. Syriac Alaph Glyph Forms 401
Table 9-17. Syriac Ligatures 402
Table 9-18. Samaritan Performative Punctuation Marks 404
Table 9-19. Dual-Joining Mandaic Characters 406
Table 9-20. Right-Joining Mandaic Characters 407
Table 10-1. Old South Arabian Numeric Characters 413
Table 10-2. Number Formation in Old South Arabian 413
Table 10-3. Number Formation in Aramaic 416
Table 10-4. Dual-Joining Manichaean Letters 419
Table 10-5. Right-Joining Manichaean Letters 419
Table 10-6. Left-Joining Manichaean Letters 420
Table 10-7. Non-Joining Manichaean Letters 420
Table 10-8. Manichaean Ligatures 420
Table 10-9. Inscriptional Parthian Shaping Behavior 423
Table 10-10. Avestan Shaping Behavior 424
Table 11-1. Cuneiform Script Usage 431
Table 11-2. Hieroglyphic Character Sequence 437
Table 12-1. Devanagari Vowel Letters 450
Table 12-2. Sample Devanagari Half-Forms 460
Table 12-3. Sample Devanagari Ligatures 461

Table 12-4. RA + Vocalic Letter Ligature Forms 462
Table 12-5. Sample Devanagari Half-Ligature Forms 462
Table 12-6. Marathi and Nepali Allographs 463
Table 12-7. Devanagari Vowels Used in Bihari Languages 465
Table 12-8. Prishthamatra Orthography 466
Table 12-9. Bengali Vowel Letters 469
Table 12-10. Diphthong Vowel Letters in Kokborok 470
Table 12-11. Assamese Consonant-Vowel Combinations 470
Table 12-12. Bengali Consonant-Vowel Combinations 471
Table 12-13. Use of Apostrophe in Bangla 474
Table 12-14. Gurmukhi Vowel Letters 476
Table 12-15. Gurmukhi Conjuncts 477
Table 12-16. Additional Pairin and Addha Forms in Gurmukhi 478
Table 12-17. Use of Joiners in Gurmukhi 478
Table 12-18. Gujarati Vowel Letters 480
Table 12-19. Gujarati Conjuncts 481
Table 12-20. Oriya Vowel Letters 482
Table 12-21. Oriya Conjuncts 483
Table 12-22. Oriya Vowel Placement 483
Table 12-23. Ligation for the Syllable om 484
Table 12-24. Tamil Ligatures with u 488
Table 12-25. Tamil Vowels, Consonants, and Syllables 492
Table 12-26. Telugu Vowel Letters 494
Table 12-27. Rendering of Telugu na + virama 495
Table 12-28. Kannada Vowel Letters 497
Table 12-29. Rendering of Kannada na + virama 500
Table 12-30. Malayalam Vowel Letters 501
Table 12-31. Malayalam Orthographic Reform 503
Table 12-32. Malayalam Conjuncts 504
Table 12-33. Candrakkala Examples 504
Table 12-34. Use of Joiners in Malayalam 505
Table 12-35. Malayalam /rara/ and /uua/ 507
Table 12-36. Malayalam /nr/ and /nt/ 507
Table 12-37. Atomic Encoding of Malayalam Chillus 508
Table 13-1. Thaana Glyph Placement 511
Table 13-2. Sinhala Vowel Letters 513
Table 13-3. Murmured Resonants in Nepal Bhasa 516
Table 13-4. Positions of Limbu Combining Characters 539
Table 13-5. Lepcha Syllabic Structure 549
Table 13-6. Various Signs in Masaram Gondi 553
Table 14-1. Brahmi Vowel Letters 556
Table 14-2. Brahmi Positional Digits 559
Table 14-3. Kharoshthi Vowel Signs 562
Table 14-4. Kharoshthi Vowel Modifiers 564
Table 14-5. Kharoshthi Consonant Modifiers 564

Table 14-6. Examples of Kharoshthi Virama 565
Table 14-7. Phags-pa Positional Forms of I, U, E, and O 571
Table 14-8. Contextual Glyph Mirroring in Phags-pa 572
Table 14-9. Phags-pa Standardized Variants 573
Table 15-1. Takri Vowel Letters 593
Table 15-2. Siddham Punctuation Characters 595
Table 15-3. Khudawadi Vowel Letters 600
Table 15-4. Representation of Arabic Sounds in Khudawadi 601
Table 15-5. Tirhuta Vowel Letters 604
Table 15-6. Modi Vowel Letters 607
Table 15-7. Rendering of Explicit Virama Forms in Grantha 610
Table 15-8. Additional Svara Marks used in Grantha 611
Table 16-1. Glyph Positions in Thai Syllables 618
Table 16-2. Glyph Positions in Lao Syllables 621
Table 16-3. Modern Burmese Syllabic Structure 626
Table 16-4. Khamti Shan Tone Marks 629
Table 16-5. Independent Khmer Vowel Characters 632
Table 16-6. Two Registers of Khmer Consonants 633
Table 16-7. Khmer Subscript Consonant Signs 634
Table 16-8. Khmer Composite Dependent Vowel Signs with Nikahit 636
Table 16-9. Khmer Subscript Independent Vowel Signs 637
Table 16-10. Tai Le Tone Marks 642
Table 16-11. Myanmar Digits in Tai Le 643
Table 16-12. New Tai Lue Vowel Placement 645
Table 16-13. New Tai Lue Registers and Tones 646
Table 16-14. Tai Viet Symbols and Punctuation 651
Table 16-15. Cham Syllabic Structure 655
Table 17-1. Hanunóo and Buhid Vowel Sign Combinations 661
Table 17-2. Balinese Base Consonants and Conjunct Forms 665
Table 17-3. Sasak Extensions for Balinese 667
Table 17-4. Balinese Consonant Clusters with u and u: 669
Table 17-5. Modern Sundanese Syllabic Structure 678
Table 18-1. Blocks Containing Han Ideographs 683
Table 18-2. Small Extensions to the URO 683
Table 18-3. Common Han Characters 686
Table 18-4. Source Encoding for Sword Variants 691
Table 18-5. Ideographs Not Unified 693
Table 18-6. Ideographs Unified 694
Table 18-7. Han Ideograph Arrangement 694
Table 18-8. Mandarin Tone Marks 705
Table 18-9. Minnan and Hakka Tone Marks 706
Table 18-10. Separating Jamo Characters 714
Table 18-11. Line-Based Placement of Jungseong 716
Table 18-12. Lisu Tone Letters 722
Table 18-13. Punctuation Adopted in Lisu Orthography 723

Table 19-1. Labialized Forms in Ethiopic -WAA 731
Table 19-2. Labialized Forms in Ethiopic -WE 731
Table 19-3. N'Ko Diacritic Usage 739
Table 19-4. N'Ko Tone Diacritics on Vowels 740
Table 19-5. N'Ko Letter Shaping 741
Table 19-6. Number Formation in Mende Kikakui 748
Table 20-1. Combining Marks used in Osage 758
Table 20-2. IPA Transcription of Deseret 761
Table 21-1. Examples of Ornamentation 770
Table 21-2. Representation of Ancient Greek Vocal and Instrumental Notation 772
Table 22-1. Currency Symbols Encoded in Other Blocks 782
Table 22-2. Mathematical Alphanumeric Symbols 789
Table 22-3. Script-Specific Decimal Digits 793
Table 22-4. Compatibility Digits 795
Table 22-5. Mathematical Operators Disunified from Punctuation 807
Table 22-6. Use of Mathematical Symbol Pieces 816
Table 22-7. Geometric Shape Collections 821
Table 22-8. Japanese Era Names 836
Table 23-1. Control Codes Specified in the Unicode Standard 841
Table 23-2. Letter Spacing 844
Table 23-3. Bidirectional Ordering Controls 851
Table 23-4. Paired Stateful Controls 852
Table 23-5. Paired Stateful Controls (Deprecated) 853
Table 23-6. Unicode Encoding Scheme Signatures 866
Table 23-7. U+FEFF Signature in Other Charsets 867
Table 24-1. IRG Sources 889
Table A-1. Extended BNF 895
Table A-2. Character Class Examples 897
Table A-3. Operators 897
Table C-1. Timeline 907
Table C-2. Zero Extending 913
Table D-1. Versions of Unicode and ISO/IEC 10646 919
Table D-2. Allocation of Code Points by Type (Versions 1.0.0 to 3.0) 920
Table D-3. Allocation of Code Points by Type (Versions 3.1 to 5.1) 921
Table D-4. Allocation of Code Points by Type (Versions 5.2 to 7.0) 922
Table D-5. Allocation of Code Points by Type (Versions 8.0 to 10.0) 923
Table F-1. CJK Strokes 932