The Unicode Standard
Version 7.0 – 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 © 1991–2014 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 7.0
Includes bibliographical references and index.
ISBN 978-1-936213-09-2) (http://www.unicode.org/versions/Unicode7.0.0/)
1. Unicode (Computer character set) I. Allen, Julie D. II. Unicode Consortium.
QA268.U545 2014

ISBN 978-1-936213-09-2
Published in Mountain View, CA
October 2014

Contents

Figures xxi
Tables xxv
Preface xxxi
Why Unicode? xxxi
What's New? xxxi
Organization of This Standard xxxiii
Unicode Standard Annexes xxxiv
The Unicode Character Database xxxv
Unicode Code Charts xxxvi
Unicode Technical Standards and Unicode Technical Reports xxxvi
Updates and Errata xxxvi
Acknowledgements xxxvii
1 Introduction 1
1.1 Coverage 3
Standards Coverage 3
New Characters 3
1.2 Design Goals 4
1.3 Text Handling 6
Characters and Glyphs 6
Text Elements 6
2 General Structure 9
2.1 Architectural Context 10
Basic Text Processes 10
Text Elements, Characters, and Text Processes 10
Text Processes and Encoding 11
2.2 Unicode Design Principles 14
Universality 14
Efficiency 15
Characters, Not Glyphs 15
Semantics 18
Plain Text 18
Logical Order 19
Unification 21
Dynamic Composition 22
Stability 23
Convertibility 23

2.3 Compatibility Characters . . . 25
Compatibility Variants . . . 26
Compatibility Decomposable Characters . . . 26
2.4 Code Points and Characters . . . 29
Types of Code Points . . . 30
2.5 Encoding Forms . . . 33
UTF-32 . . . 35
UTF-16 . . . 36
UTF-8 . . . 36
Comparison of the Advantages of UTF-32, UTF-16, and UTF-8 . . . 38
2.6 Encoding Schemes . . . 40
2.7 Unicode Strings . . . 43
2.8 Unicode Allocation . . . 44
Planes . . . 44
Allocation Areas and Character Blocks . . . 45
Assignment of Code Points . . . 46
2.9 Details of Allocation . . . 48
Plane 0 (BMP) . . . 49
Plane 1 (SMP) . . . 51
Plane 2 (SIP) . . . 52
Other Planes . . . 52
2.10 Writing Direction . . . 53
2.11 Combining Characters . . . 55
Sequence of Base Characters and Diacritics . . . 56
Multiple Combining Characters . . . 57
Ligated Multiple Base Characters . . . 59
Exhibiting Nonspacing Marks in Isolation . . . 60
"Characters" and Grapheme Clusters . . . 60
2.12 Equivalent Sequences . . . 62
Normalization . . . 62
Decompositions . . . 63
Non-decomposition of Overlaid Diacritics . . . 64
2.13 Special Characters . . . 67
Special Noncharacter Code Points . . . 67
Byte Order Mark (BOM) . . . 67
Layout and Format Control Characters . . . 68
The Replacement Character . . . 68
Control Codes . . . 68
2.14 Conforming to the Unicode Standard . . . 69
Characteristics of Conformant Implementations . . . 69
Unacceptable Behavior . . . 70
Acceptable Behavior . . . 70
Supported Subsets . . . 71

3 Conformance ... 73

3.1 Versions of the Unicode Standard ... 74

Stability ... 74
Version Numbering ... 75
Errata and Corrigenda ... 76
References to the Unicode Standard ... 76
Precision in Version Citation ... 76
References to Unicode Character Properties ... 77
References to Unicode Algorithms ... 77

3.2 Conformance Requirements ... 79

Code Points Unassigned to Abstract Characters ... 79
Interpretation ... 80
Modification ... 81
Character Encoding Forms ... 82
Character Encoding Schemes ... 83
Bidirectional Text ... 84
Normalization Forms ... 84
Normative References ... 84
Unicode Algorithms ... 84
Default Casing Algorithms ... 85
Unicode Standard Annexes ... 85

3.3 Semantics ... 87

Definitions ... 87
Character Identity and Semantics ... 87

3.4 Characters and Encoding ... 90

3.5 Properties ... 94

Types of Properties ... 94
Property Values ... 95
Default Property Values ... 96
Classification of Properties by Their Values ... 97
Property Status ... 98
Context Dependence ... 101
Stability of Properties ... 101
Simple and Derived Properties ... 103
Property Aliases ... 103
Private Use ... 104

3.6 Combination ... 105

Combining Character Sequences ... 105
Grapheme Clusters ... 107
Application of Combining Marks ... 109

3.7 Decomposition ... 115

Compatibility Decomposition ... 115
Canonical Decomposition ... 116

3.8 Surrogates ... 118
3.9 Unicode Encoding Forms ... 119
UTF-32 ... 123
UTF-16 ... 124
UTF-8 ... 124
Encoding Form Conversion ... 126
Constraints on Conversion Processes ... 126
3.10 Unicode Encoding Schemes ... 130
3.11 Normalization Forms ... 134
Normalization Stability ... 134
Combining Classes ... 135
Specification of Unicode Normalization Forms ... 136
Starters ... 136
Canonical Ordering Algorithm ... 137
Canonical Composition Algorithm ... 138
Definition of Normalization Forms ... 140
3.12 Conjoining Jamo Behavior ... 142
Definitions ... 142
Hangul Syllable Decomposition ... 144
Hangul Syllable Composition ... 146
Hangul Syllable Name Generation ... 147
Sample Code for Hangul Algorithms ... 148
3.13 Default Case Algorithms ... 152
Definitions ... 152
Default Case Conversion ... 154
Default Case Folding ... 155
Default Case Detection ... 156
Default Caseless Matching ... 157
4 Character Properties ... 159
4.1 Unicode Character Database ... 161
4.2 Case ... 164
Definitions of Case and Casing ... 164
Case Mapping ... 166
4.3 Combining Classes ... 168
Reordrant, Split, and Subjoined Combining Marks ... 168
4.4 Directionality ... 173
4.5 General Category ... 174
4.6 Numeric Value ... 177
Ideographic Numeric Values ... 178
4.7 Bidi Mirrored ... 180

4.8 Name ... 181
Unicode Name Property ... 184
Code Point Labels ... 185
Use of Character Names in APIs and User Interfaces ... 186
4.9 Unicode 1.0 Names ... 187
4.10 Letters, Alphabetic, and Ideographic ... 188
4.11 Properties Related to Text Boundaries ... 189
4.12 Characters with Unusual Properties ... 190
5 Implementation Guidelines ... 193
5.1 Data Structures for Character Conversion ... 194
Issues ... 194
Multistage Tables ... 194
5.2 Programming Languages and Data Types ... 197
Unicode Data Types for C ... 197
5.3 Unknown and Missing Characters ... 199
5.4 Handling Surrogate Pairs in UTF-16 ... 201
5.5 Handling Numbers ... 203
5.6 Normalization ... 204
5.7 Compression ... 206
5.8 Newline Guidelines ... 207
Definitions ... 207
Line Separator and Paragraph Separator ... 208
Recommendations ... 209
5.9 Regular Expressions ... 212
5.10 Language Information in Plain Text ... 213
Requirements for Language Tagging ... 213
Language Tags and Han Unification ... 213
5.11 Editing and Selection ... 215
Consistent Text Elements ... 215
5.12 Strategies for Handling Nonspacing Marks ... 217
Keyboard Input ... 218
Truncation ... 218
5.13 Rendering Nonspacing Marks ... 220
Canonical Equivalence ... 223
Positioning Methods ... 224
5.14 Locating Text Element Boundaries ... 226
5.15 Identifiers ... 227

5.16 Sorting and Searching . . . 228
Culturally Expected Sorting and Searching . . . 228
Language-Insensitive Sorting . . . 228
Searching . . . 229
Sublinear Searching . . . 229
5.17 Binary Order . . . 231
UTF-8 in UTF-16 Order . . . 231
UTF-16 in UTF-8 Order . . . 232
5.18 Case Mappings . . . 234
Titlecasing . . . 234
Complications for Case Mapping . . . 235
Reversibility . . . 237
Caseless Matching . . . 238
Normalization and Casing . . . 239
5.19 Mapping Compatibility Variants . . . 241
5.20 Unicode Security . . . 243
5.21 Ignoring Characters in Processing . . . 246
Characters Ignored in Text Segmentation . . . 246
Characters Ignored in Line Breaking . . . 247
Characters Ignored in Cursive Joining . . . 247
Characters Ignored in Identifiers . . . 247
Characters Ignored in Searching and Sorting . . . 248
Characters Ignored for Display . . . 248
5.22 Best Practice for U+FFFD Substitution . . . 252
6 Writing Systems and Punctuation . . . 255
6.1 Writing Systems . . . 256
6.2 General Punctuation . . . 261
Blocks Devoted to Punctuation . . . 262
Format Control Characters . . . 263
Space Characters . . . 264
Dashes and Hyphens . . . 265
Paired Punctuation . . . 267
Language-Based Usage of Quotation Marks . . . 268
Apostrophes . . . 272
Other Punctuation . . . 272
Archaic Punctuation and Editorial Marks . . . 277
Indic Punctuation . . . 279
CJK Punctuation . . . 281
Unknown or Unavailable Ideographs . . . 283
CJK Compatibility Forms . . . 283

7 Europe-I 285
Modern and Liturgical Scripts
7.1 Latin 287
Letters of Basic Latin: U+0041–U+007A 291
Letters of the Latin-1 Supplement: U+00C0–U+00FF 291
Latin Extended-A: U+0100–U+017F 291
Latin Extended-B: U+0180–U+024F 292
IPA Extensions: U+0250–U+02AF 294
Phonetic Extensions: U+1D00–U+1DBF 296
Latin Extended Additional: U+1E00–U+1EFF 297
Latin Extended-C: U+2C60–U+2C7F 297
Latin Extended-D: U+A720–U+A7FF 298
Latin Extended-E: U+AB30–U+AB6F 299
Latin Ligatures: U+FB00–U+FB06 299
7.2 Greek 300
Greek: U+0370–U+03FF 300
Greek Extended: U+1F00–U+1FFF 304
Ancient Greek Numbers: U+10140–U+1018F 305
7.3 Coptic 307
7.4 Cyrillic 310
Cyrillic: U+0400–U+04FF 310
Cyrillic Supplement: U+0500–U+052F 311
Cyrillic Extended-A: U+2DE0–U+2DFF 311
Cyrillic Extended-B: U+A640–U+A69F 312
7.5 Glagolitic 313
7.6 Armenian 314
7.7 Georgian 316
7.8 Modifier Letters 318
Spacing Modifier Letters: U+02B0–U+02FF 319
Modifier Tone Letters: U+A700–U+A71F 321
7.9 Combining Marks 322
Combining Diacritical Marks: U+0300–U+036F 327
Combining Diacritical Marks Extended: U+1AB0–U+1AFF 328
Combining Diacritical Marks Supplement: U+1DC0–U+1DFF 328
Combining Marks for Symbols: U+20D0–U+20FF 329
Combining Half Marks: U+FE20–U+FE2F 330
Combining Marks in Other Blocks 330
8 Europe-II 333
Ancient and Other Scripts
8.1 Linear A 335

8.2 Linear B . . . 336
Linear B Syllabary: U+10000–U+1007F . . . 336
Linear B Ideograms: U+10080–U+100FF . . . 336
Aegean Numbers: U+10100–U+1013F . . . 336
8.3 Cypriot Syllabary . . . 338
8.4 Ancient Anatolian Alphabets . . . 339
Lycian: U+10280–U+1029F . . . 339
Carian: U+102A0–U+102DF . . . 339
Lydian: U+10920–U+1093F . . . 339
8.5 Old Italic . . . 341
8.6 Runic . . . 344
8.7 Gothic . . . 347
8.8 Elbasan . . . 348
8.9 Caucasian Albanian . . . 349
8.10 Old Permic . . . 350
8.11 Ogham . . . 351
8.12 Shavian . . . 352
9 Middle East-I . . . 353
Modern and Liturgical Scripts
9.1 Hebrew . . . 355
Hebrew: U+0590–U+05FF . . . 355
Alphabetic Presentation Forms: U+FB1D–U+FB4F . . . 359
9.2 Arabic . . . 361
Arabic: U+0600–U+06FF . . . 361
Arabic Cursive Joining . . . 368
Arabic Ligatures . . . 371
Arabic Joining Groups . . . 373
Arabic Supplement: U+0750–U+077F . . . 380
Arabic Extended-A: U+08A0–U+08FF . . . 380
Arabic Presentation Forms-A: U+FB50–U+FDFF . . . 380
Arabic Presentation Forms-B: U+FE70–U+FEFF . . . 382
9.3 Syriac . . . 383
Syriac: U+0700–U+074F . . . 383
Syriac Shaping . . . 387
9.4 Samaritan . . . 392
9.5 Mandaic . . . 394
10 Middle East-II . . . 397
Ancient Scripts
10.1 Old North Arabian . . . 399
10.2 Old South Arabian . . . 400
10.3 Phoenician . . . 402

10.4 Imperial Aramaic 404
10.5 Manichaean 406
10.6 Pahlavi and Parthian 410
Inscriptional Parthian: U+10B40–U+10B5F 410
Inscriptional Pahlavi: U+10B60–U+10B7F 410
Psalter Pahlavi: U+10B80–U+10BAF 411
10.7 Avestan 412
10.8 Nabataean 414
10.9 Palmyrene 415
11 Cuneiform and Hieroglyphs 417
11.1 Sumero-Akkadian 418
Cuneiform: U+12000–U+123FF 418
Cuneiform Numbers and Punctuation: U+12400–U+1247F 421
11.2 Ugaritic 422
11.3 Old Persian 423
11.4 Egyptian Hieroglyphs 424
11.5 Meroitic 428
12 South and Central Asia-I 431
Official Scripts of India
12.1 Devanagari 433
Devanagari: U+0900–U+097F 433
Principles of the Devanagari Script 434
Rendering Devanagari 440
Devanagari Digits, Punctuation, and Symbols 449
Extensions in the Main Devanagari Block 450
Devanagari Extended: U+A8E0–U+A8FF 452
Vedic Extensions: U+1CD0–U+1CFF 453
12.2 Bengali (Bangla) 455
12.3 Gurmukhi 461
12.4 Gujarati 466
12.5 Oriya (Odia) 468
12.6 Tamil 471
Tamil: U+0B80–U+0BFF 471
Tamil Vowels 472
Tamil Ligatures 473
Tamil Named Character Sequences 477
12.7 Telugu 480
12.8 Kannada 483
Kannada: U+0C80–U+0CFF 483
Principles of the Kannada Script 483
Rendering Kannada 484

12.9 Malayalam 487
Malayalam Orthographic Reform 487
Rendering Malayalam 488
Malayalam Numbers and Punctuation 492
13 South and Central Asia-II 493
Other Modern Scripts
13.1 Thaana 495
13.2 Sinhala 497
Sinhala Archaic Numbers: U+111E0–U+111FF 498
13.3 Tibetan 499
13.4 Mongolian 511
13.5 Limbu 520
13.6 Meetei Mayek 524
13.7 Mro 526
13.8 Warang Citi 527
13.9 Ol Chiki 528
13.10 Chakma 530
13.11 Lepcha 532
13.12 Saurashtra 535
14 South and Central Asia-III 537
Ancient Scripts
14.1 Brahmi 538
14.2 Kharoshthi 542
Kharoshthi: U+10A00–U+10A5F 542
Rendering Kharoshthi 544
14.3 Phags-pa 548
14.4 Old Turkic 555
15 South and Central Asia-IV 557
Other Historic Scripts
15.1 Syloti Nagri 559
15.2 Kaithi 561
15.3 Sharada 564
15.4 Takri 566
15.5 Siddham 568
15.6 Mahajani 570
15.7 Khojki 572
15.8 Khudawadi 574
15.9 Tirhuta 576
15.10 Modi 579

15.11 Grantha 582
Rendering Behavior 582
15.12 Sora Sompeng 585
16 Southeast Asia 587
16.1 Thai 589
16.2 Lao 593
16.3 Myanmar 596
Myanmar: U+1000–U+109F 596
Myanmar Extended-A: U+AA60–U+AA7F 600
Khamti Shan 600
Aiton and Phake 602
Myanmar Extended-B: U+A9E0–U+A9FF 602
16.4 Khmer 603
Khmer: U+1780–U+17FF 603
Principles of the Khmer Script 603
Khmer Symbols: U+19E0–U+19FF 613
16.5 Tai Le 614
16.6 New Tai Lue 616
16.7 Tai Tham 618
16.8 Tai Viet 621
16.9 Kayah Li 624
16.10 Cham 625
16.11 Pahawh Hmong 627
16.12 Pau Cin Hau 629
17 Indonesia and Oceania 631
17.1 Philippine Scripts 632
Tagalog: U+1700–U+171F 632
Hanunóo: U+1720–U+173F 632
Buhid: U+1740–U+175F 632
Tagbanwa: U+1760–U+177F 632
Principles of the Philippine Scripts 632
17.2 Buginese 635
17.3 Balinese 637
17.4 Javanese 643
17.5 Rejang 647
17.6 Batak 648
17.7 Sundanese 649
Sundanese: U+1B80–U+1BBF 649
Sundanese Supplement: U+1CC0–U+1CCF 650

18 East Asia 651
18.1 Han 653
CJK Unified Ideographs 653
Blocks Containing Han Ideographs 654
General Characteristics of Han Ideographs 656
Principles of Han Unification 660
Unification Rules 661
Abstract Shape 662
Han Ideograph Arrangement 664
Radical-Stroke Indices 665
Mappings for Han Ideographs 666
CJK Unified Ideographs Extension B: U+20000–U+2A6D6 666
CJK Unified Ideographs Extension C: U+2A700–U+2B734 667
CJK Unified Ideographs Extension D: U+2B740–U+2B81D 667
CJK Compatibility Ideographs: U+F900–U+FAFF 667
CJK Compatibility Supplement: U+2F800–U+2FA1D 668
Kanbun: U+3190–U+319F 668
Symbols Derived from Han Ideographs 668
CJK and KangXi Radicals: U+2E80–U+2FD5 668
CJK Additions from HKSCS and GB 18030 669
CJK Strokes: U+31C0–U+31EF 670
18.2 Ideographic Description Characters 671
18.3 Bopomofo 675
18.4 Hiragana and Katakana 678
Hiragana: U+3040–U+309F 678
Katakana: U+30A0–U+30FF 678
Katakana Phonetic Extensions: U+31F0–U+31FF 679
Kana Supplement U+1B000–U+1B0FF 679
18.5 Halfwidth and Fullwidth Forms 680
18.6 Hangul 681
Hangul Jamo: U+1100–U+11FF 681
Hangul Jamo Extended-A: U+A960–U+A97F 682
Hangul Jamo Extended-B: U+D7B0–U+D7FF 682
Hangul Compatibility Jamo: U+3130–U+318F 682
Hangul Syllables: U+AC00–U+D7A3 683
18.7 Yi 685
18.8 Lisu 688
18.9 Miao 691
19 Africa 693
19.1 Ethiopic 694
Ethiopic: U+1200–U+137F 694
Ethiopic Extensions 696
19.2 Osmanya 698

19.3 Tifinagh 699
19.4 N'Ko 702
19.5 Vai 707
19.6 Bamum 709
Bamum: U+A6A0–U+A6FF 709
Bamum Supplement: U+16800–U+16A3F 709
19.7 Bassa Vah 711
19.8 Mende Kikakui 712
20 Americas 715
20.1 Cherokee 716
20.2 Canadian Aboriginal Syllabics 717
Canadian Aboriginal Syllabics: U+1400–U+167F 717
Canadian Aboriginal Syllabics Extended: U+18B0–U+18FF 718
20.3 Deseret 719
21 Notational Systems 723
Shorthands, Braille, Musical Symbols
21.1 Braille 724
21.2 Western Musical Symbols 726
21.3 Byzantine Musical Symbols 731
21.4 Ancient Greek Musical Notation 732
21.5 Duployan 735
22 Symbols 737
22.1 Currency Symbols 739
22.2 Letterlike Symbols 742
Letterlike Symbols: U+2100–U+214F 742
Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF 743
Mathematical Alphabets 744
Fonts Used for Mathematical Alphabets 746
Arabic Mathematical Alphabetic Symbols: U+1EE00–U+1EEFF 748
22.3 Numerals 749
Decimal Digits 749
Other Digits 751
Non-Decimal Radix Systems 753
Acrophonic Systems and Other Letter-based Numbers 754
Coptic Epact Numbers: U+102E0–U+102FF 755
Rumi Numeral Forms: U+10E60–U+10E7E 756
CJK Numerals 756
Fractions 758
Common Indic Number Forms: U+A830–U+A83F 758
22.4 Superscript and Subscript Symbols 760
Superscripts and Subscripts: U+2070–U+209F 760

22.5 Mathematical Symbols 762
Mathematical Operators: U+2200–U+22FF 763
Supplements to Mathematical Symbols and Arrows 766
Supplemental Mathematical Operators: U+2A00–U+2AFF 766
Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF 766
Miscellaneous Mathematical Symbols-B: U+2980–U+29FF 767
Miscellaneous Symbols and Arrows: U+2B00–U+2B7F 767
Arrows: U+2190–U+21FF 767
Supplemental Arrows 768
Standardized Variants of Mathematical Symbols 769
22.6 Invisible Mathematical Operators 770
22.7 Technical Symbols 771
Control Pictures: U+2400–U+243F 771
Miscellaneous Technical: U+2300–U+23FF 771
Optical Character Recognition: U+2440–U+245F 775
22.8 Geometrical Symbols 776
Box Drawing and Block Elements 776
Geometric Shapes: U+25A0–U+25FF 777
Geometric Shapes Extended: U+1F780–U+1F7FF 778
22.9 Miscellaneous Symbols 779
Miscellaneous Symbols: U+2600–U+26FF 780
Miscellaneous Symbols and Pictographs: U+1F300–U+1F5FF 780
Emoticons: U+1F600–U+1F64F 783
Transport and Map Symbols: U+1F680–U+1F6FF 783
Dingbats: U+2700–U+27BF 783
Ornamental Dingbats: U+1F650–U+1F67F 784
Alchemical Symbols: U+1F700–U+1F77F 785
Mahjong Tiles: U+1F000–U+1F02F 785
Domino Tiles: U+1F030–U+1F09F 786
Playing Cards: U+1F0A0–U+1F0FF 786
Yijing Hexagram Symbols: U+4DC0–U+4DFF 787
Tai Xuan Jing Symbols: U+1D300–U+1D356 787
Ancient Symbols: U+10190–U+101CF 788
Phaistos Disc Symbols: U+101D0–U+101FF 788
22.10 Enclosed and Square 790
Enclosed Alphanumerics: U+2460–U+24FF 792
Enclosed CJK Letters and Months: U+3200–U+32FF 792
CJK Compatibility: U+3300–U+33FF 793
Enclosed Alphanumeric Supplement: U+1F100–U+1F1FF 793
Enclosed Ideographic Supplement: U+1F200–U+1F2FF 794

23 Special Areas and Format Characters 795
23.1 Control Codes 796
Representing Control Sequences 796
Specification of Control Code Semantics 797
23.2 Layout Controls 799
Line and Word Breaking 799
Cursive Connection and Ligatures 801
Combining Grapheme Joiner 805
Bidirectional Ordering Controls 807
Stateful Format Controls 808
23.3 Deprecated Format Characters 810
23.4 Variation Selectors 812
23.5 Private-Use Characters 815
Private Use Area: U+E000–U+F8FF 816
Supplementary Private Use Areas 817
23.6 Surrogates Area 818
23.7 Noncharacters 819
23.8 Specials 821
Byte Order Mark (BOM): U+FEFF 821
Specials: U+FFF0–U+FFF8 823
Annotation Characters: U+FFF9–U+FFFB 823
Replacement Characters: U+FFFC–U+FFFD 825
23.9 Deprecated Tag Characters 826
Deprecated Tag Characters: U+E0000–U+E007F 826
Syntax for Embedding Tags 826
Working with Language Tags 828
Unicode Conformance Issues 830
Formal Tag Syntax 830
24 About the Code Charts 831
24.1 Character Names List 832
Images in the Code Charts and Character Lists 832
Special Characters and Code Points 834
Character Names 836
Informative Aliases 836
Normative Aliases 836
Cross References 836
Information About Languages 837
Case Mappings 837
Decompositions 838
Standardized Variation Sequences 839
Subheads 840

24.2 CJK Unified and Compatibility Ideographs 841
CJK Unified Ideographs 841
Compatibility Ideographs 842
24.3 Hangul Syllables 844
A Notational Conventions 845
Code Points 845
Character Names 845
Character Blocks 845
Sequences 846
Rendering 846
Properties and Property Values 847
Miscellaneous 847
Extended BNF 847
Operators 849
B Unicode Publications and Resources 851
B.1 The Unicode Consortium 852
The Unicode Technical Committee 852
Other Activities 852
B.2 Unicode Publications 853
B.3 Unicode Technical Standards 854
UTS #6: A Standard Compression Scheme for Unicode 854
UTS #10: Unicode Collation Algorithm 854
UTS #18: Unicode Regular Expressions 854
UTS #22: Character Mapping Markup Language (CharMapML) 854
UTS #35: Unicode Locale Data Markup Language (LDML) 854
UTS #37: Unicode Ideographic Variation Database 854
UTS #39: Unicode Security Mechanisms 854
UTS #46: Unicode IDNA Compatibility Processing 855
B.4 Unicode Technical Reports 856
UTR #16: UTF-EBCDIC 856
UTR #17: Unicode Character Encoding Model 856
UTR #20: Unicode in XML and Other Markup Languages 856
UTR #23: The Unicode Character Property Model 856
UTR #25: Unicode Support for Mathematics 856
UTR #26: Compatibility Encoding Scheme for UTF-16: 8-Bit (CESU-8) . . 856
UTR #33: Unicode Conformance Model 857
UTR #36: Unicode Security Considerations 857
UTR #50: Unicode Vertical Text Layout 857
B.5 Unicode Technical Notes 858
B.6 Other Unicode Online Resources 859
Unicode Online Resources 859
How to Contact the Unicode Consortium 860

C Relationship to ISO/IEC 10646 861
C.1 History 862
C.2 Encoding Forms in ISO/IEC 10646 868
Zero Extending 868
C.3 UTF-8 and UTF-16 869
UTF-8 869
UTF-16 869
C.4 Synchronization of the Standards 870
C.5 Identification of Features for the Unicode Standard 871
C.6 Character Names 872
C.7 Character Functional Specifications 873
D Changes from Previous Versions 875
D.1 Versions of the Unicode Standard 876
D.2 Clause and Definition Updates 880
E Han Unification History 881
E.1 Development of the URO 882
E.2 Ideographic Rapporteur Group 884
E.3 CJK Sources 886
F Documentation of CJK Strokes 891
R References 899
R.1 Source Standards and Specifications 900
R.2 Source Dictionaries for Han 909
R.3 Other Script Sources 910
R.4 Selected Resources: Technical 935
R.5 Selected Resources: Other 938
I General Index 943