The Unicode® Standard
Version 14.0 – Core Specification

To learn about the latest version of the Unicode Standard, see https://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.

© 2021 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 https://www.unicode.org/reporting.html. For information about the Unicode terms of use, please see https://www.unicode.org/copyright.html.

The Unicode Standard / the Unicode Consortium; edited by the Unicode Consortium. — Version 14.0.

Includes index.

ISBN 978-1-936213-29-0 (https://www.unicode.org/versions/Unicode14.0.0/)

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

QA268.U545 2021

ISBN 978-1-936213-29-0

Published in Mountain View, CA

September 2021

Contents

Preface xxi
Why Unicode? xxi
Organization of This Standard xxi
The Unicode Character Database xxii
Unicode Code Charts xxiii
Unicode Standard Annexes xxiii
Unicode Technical Standards and Unicode Technical Reports xxiv
Updates and Errata xxv
Acknowledgements xxv
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 23
Stability 23
Convertibility 24
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 ... 37
2.6 Encoding Schemes ... 39
2.7 Unicode Strings ... 42
2.8 Unicode Allocation ... 43
Planes ... 43
Allocation Areas and Blocks ... 44
Assignment of Code Points ... 45
2.9 Details of Allocation ... 47
Plane 0 (BMP) ... 48
Plane 1 (SMP) ... 50
Plane 2 (SIP) ... 51
Plane 3 (TIP) ... 51
Other Planes ... 51
2.10 Writing Direction ... 52
2.11 Combining Characters ... 54
Sequence of Base Characters and Diacritics ... 55
Multiple Combining Characters ... 56
Ligated Multiple Base Characters ... 59
Exhibiting Nonspacing Marks in Isolation ... 59
"Characters" and Grapheme Clusters ... 60
2.12 Equivalent Sequences ... 61
Normalization ... 61
Decompositions ... 62
Non-decomposition of Certain Diacritics ... 64
2.13 Special Characters ... 66
Special Noncharacter Code Points ... 66
Byte Order Mark (BOM) ... 66
Layout and Format Control Characters ... 67
The Replacement Character ... 67
Control Codes ... 67
2.14 Conforming to the Unicode Standard ... 68
Characteristics of Conformant Implementations ... 68
Unacceptable Behavior ... 69
Acceptable Behavior ... 69
Supported Subsets ... 70
3 Conformance ... 71
3.1 Versions of the Unicode Standard ... 72
Stability ... 72
Version Numbering ... 73
Errata and Corrigenda ... 74

References to the Unicode Standard 74
Precision in Version Citation 75
References to Unicode Character Properties 75
References to Unicode Algorithms 76
3.2 Conformance Requirements 77
Code Points Unassigned to Abstract Characters 77
Interpretation 78
Modification 79
Character Encoding Forms 80
Character Encoding Schemes 81
Bidirectional Text 82
Normalization Forms 82
Normative References 82
Unicode Algorithms 82
Default Casing Algorithms 83
Unicode Standard Annexes 83
3.3 Semantics 85
Definitions 85
Character Identity and Semantics 85
3.4 Characters and Encoding 88
3.5 Properties 93
Types of Properties 93
Property Values 94
Default Property Values 95
Classification of Properties by Their Values 96
Property Status 97
Context Dependence 100
Stability of Properties 100
Simple and Derived Properties 102
Property Aliases 102
Private Use 103
3.6 Combination 104
Combining Character Sequences 104
Grapheme Clusters 107
Application of Combining Marks 108
3.7 Decomposition 114
Compatibility Decomposition 114
Canonical Decomposition 115
3.8 Surrogates 117
3.9 Unicode Encoding Forms 118
UTF-32 122
UTF-16 123
UTF-8 123

Encoding Form Conversion . . . 125
Constraints on Conversion Processes . . . 125
U+FFFD Substitution of Maximal Subparts . . . 126
3.10 Unicode Encoding Schemes . . . 129
3.11 Normalization Forms . . . 133
Normalization Stability . . . 133
Combining Classes . . . 134
Specification of Unicode Normalization Forms . . . 135
Starters . . . 135
Canonical Ordering Algorithm . . . 136
Canonical Composition Algorithm . . . 137
Definition of Normalization Forms . . . 139
3.12 Conjoining Jamo Behavior . . . 141
Definitions . . . 141
Hangul Syllable Decomposition . . . 143
Hangul Syllable Composition . . . 145
Hangul Syllable Name Generation . . . 146
Sample Code for Hangul Algorithms . . . 147
3.13 Default Case Algorithms . . . 151
Definitions . . . 151
Default Case Conversion . . . 153
Default Case Folding . . . 154
Default Case Detection . . . 155
Default Caseless Matching . . . 156
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 . . . 171
4.5 General Category . . . 172
4.6 Numeric Value . . . 175
Ideographic Numeric Values . . . 176
4.7 Bidi Mirrored . . . 178
4.8 Name . . . 180
Unicode Name Property . . . 183
Code Point Labels . . . 186
Use of Character Names in APIs and User Interfaces . . . 186
4.9 Unicode 1.0 Names . . . 188
4.10 Letters, Alphabetic, and Ideographic . . . 189

4.11 Properties for Text Boundaries 190
4.12 Characters with Unusual Properties 191
5 Implementation Guidelines 195
5.1 Data Structures for Character Conversion 196
Issues 196
Multistage Tables 196
5.2 Programming Languages and Data Types 199
Unicode Data Types for C 199
5.3 Unknown and Missing Characters 201
5.4 Handling Surrogate Pairs in UTF-16 203
5.5 Handling Numbers 205
5.6 Normalization 206
5.7 Compression 208
5.8 Newline Guidelines 209
Definitions 209
Line Separator and Paragraph Separator 210
Recommendations 211
5.9 Regular Expressions 214
5.10 Language Information in Plain Text 215
Requirements for Language Tagging 215
Language Tags and Han Unification 215
5.11 Editing and Selection 217
Consistent Text Elements 217
5.12 Strategies for Handling Nonspacing Marks 219
Keyboard Input 220
Truncation 220
5.13 Rendering Nonspacing Marks 222
Canonical Equivalence 225
Positioning Methods 226
5.14 Locating Text Element Boundaries 228
5.15 Identifiers 229
5.16 Sorting and Searching 230
Culturally Expected Sorting and Searching 230
Language-Insensitive Sorting 230
Searching 231
Sublinear Searching 232
5.17 Binary Order 233
UTF-8 in UTF-16 Order 233
UTF-16 in UTF-8 Order 234

5.18 Case Mappings 236
Titlecasing 236
Complications for Case Mapping 237
Reversibility 239
Caseless Matching 240
Normalization and Casing 242
5.19 Mapping Compatibility Variants 243
5.20 Unicode Security 245
5.21 Ignoring Characters in Processing 248
Characters Ignored in Text Segmentation 248
Characters Ignored in Line Breaking 249
Characters Ignored in Cursive Joining 249
Characters Ignored in Identifiers 249
Characters Ignored in Searching and Sorting 250
Characters Ignored for Display 250
5.22 U+FFFD Substitution in Conversion 255
6 Writing Systems and Punctuation 257
6.1 Writing Systems 258
6.2 General Punctuation 263
Blocks Devoted to Punctuation 264
Format Control Characters 265
Space Characters 266
Dashes and Hyphens 267
Paired Punctuation 269
Language-Based Usage of Quotation Marks 270
Apostrophes 274
Other Punctuation 275
Archaic Punctuation and Editorial Marks 280
Indic Punctuation 282
CJK Punctuation 284
Unknown or Unavailable Ideographs 286
CJK Compatibility Forms 287
7 Europe-I 289
Modern and Liturgical Scripts
7.1 Latin 291
Letters of Basic Latin: U+0041–U+007A 295
Letters of the Latin-1 Supplement: U+00C0–U+00FF 295
Latin Extended-A: U+0100–U+017F 295
Latin Extended-B: U+0180–U+024F 296
IPA Extensions: U+0250–U+02AF 298
Phonetic Extensions: U+1D00–U+1DBF 300
Latin Extended Additional: U+1E00–U+1EFF 301
Latin Extended-C: U+2C60–U+2C7F 301

Latin Extended-D: U+A720–U+A7FF 302
Latin Extended-E: U+AB30–U+AB6F 303
Latin Extended-F: U+10780–U+107BF 304
Latin Ligatures: U+FB00–U+FB06 304
7.2 Greek 305
Greek: U+0370–U+03FF 305
Greek Extended: U+1F00–U+1FFF 309
Ancient Greek Numbers: U+10140–U+1018F 310
7.3 Coptic 312
7.4 Cyrillic 315
Cyrillic: U+0400–U+04FF 315
Cyrillic Supplement: U+0500–U+052F 317
Cyrillic Extended-A: U+2DE0–U+2DFF 317
Cyrillic Extended-B: U+A640–U+A69F 318
Cyrillic Extended-C: U+1C80–U+1C8F 318
7.5 Glagolitic 320
Glagolitic: U+2C00–U+2C5F 320
Glagolitic Supplement: U+1E000–U+1E02F 320
7.6 Armenian 321
7.7 Georgian 323
Georgian: U+10A0–U+10FF 323
Georgian Extended: U+1C90–U+1CBF 323
Georgian Supplement: U+2D00–U+2D2F 323
7.8 Modifier Letters 325
Spacing Modifier Letters: U+02B0–U+02FF 326
Modifier Tone Letters: U+A700–U+A71F 328
7.9 Combining Marks 329
Combining Diacritical Marks: U+0300–U+036F 334
Combining Diacritical Marks Extended: U+1AB0–U+1AFF 335
Combining Diacritical Marks Supplement: U+1DC0–U+1DFF 336
Combining Diacritical Marks for Symbols: U+20D0–U+20FF 336
Combining Half Marks: U+FE20–U+FE2F 337
Combining Marks in Other Blocks 338
8 Europe-II 341
Ancient and Other Scripts
8.1 Linear A 343
8.2 Linear B 344
Linear B Syllabary: U+10000–U+1007F 344
Linear B Ideograms: U+10080–U+100FF 344
Aegean Numbers: U+10100–U+1013F 344
8.3 Cypriot Syllabary 346
8.4 Cypro-Minoan 347

8.5 Ancient Anatolian Alphabets ... 348
Lycian: U+10280–U+1029F ... 348
Carian: U+102A0–U+102DF ... 348
Lydian: U+10920–U+1093F ... 348
8.6 Old Italic ... 350
8.7 Runic ... 353
8.8 Old Hungarian ... 356
8.9 Gothic ... 357
8.10 Elbasan ... 358
8.11 Caucasian Albanian ... 359
8.12 Vithkuqi ... 360
8.13 Old Permic ... 361
8.14 Ogham ... 362
8.15 Shavian ... 363
9 Middle East-I ... 365
Modern and Liturgical Scripts
9.1 Hebrew ... 367
Hebrew: U+0590–U+05FF ... 367
Alphabetic Presentation Forms: U+FB1D–U+FB4F ... 371
9.2 Arabic ... 373
Arabic: U+0600–U+06FF ... 373
Arabic Cursive Joining ... 381
Arabic Ligatures ... 384
Arabic Joining Groups ... 385
Combining Hamza ... 392
Other Letters for Extended Arabic ... 394
Arabic Letters in Other Languages ... 395
Arabic Supplement: U+0750–U+077F ... 395
Arabic Extended-A: U+08A0–U+08FF ... 395
Arabic Presentation Forms-A: U+FB50–U+FDFF ... 396
Arabic Presentation Forms-B: U+FE70–U+FEFF ... 397
9.3 Syriac ... 399
Syriac: U+0700–U+074F ... 399
Syriac Shaping ... 403
Syriac Supplement: U+0860–U+086F ... 408
9.4 Samaritan ... 409
9.5 Mandaic ... 411
9.6 Yezidi ... 414
10 Middle East-II ... 415
Ancient Scripts
10.1 Old North Arabian ... 417

10.2 Old South Arabian 418
10.3 Phoenician 420
10.4 Imperial Aramaic 422
10.5 Manichaean 424
10.6 Pahlavi and Parthian 428
Inscriptional Parthian: U+10B40–U+10B5F 428
Inscriptional Pahlavi: U+10B60–U+10B7F 428
Psalter Pahlavi: U+10B80–U+10BAF 429
10.7 Avestan 430
10.8 Chorasmian 432
10.9 Elymaic 433
10.10 Nabataean 434
10.11 Palmyrene 435
10.12 Hatran 436
11 Cuneiform and Hieroglyphs 437
11.1 Sumero-Akkadian 438
Cuneiform: U+12000–U+123FF 438
Cuneiform Numbers and Punctuation: U+12400–U+1247F 441
Early Dynastic Cuneiform: U+12480–U+1254F 441
11.2 Ugaritic 442
11.3 Old Persian 443
11.4 Egyptian Hieroglyphs 444
Egyptian Hieroglyphs: U+13000–U+1342F 444
Egyptian Hieroglyph Format Controls: U+13430–U+1343F 446
11.5 Meroitic 451
11.6 Anatolian Hieroglyphs 453
12 South and Central Asia-I 455
Official Scripts of India
12.1 Devanagari 457
Devanagari: U+0900–U+097F 457
Principles of the Devanagari Script 458
Rendering Devanagari 467
Devanagari Digits, Punctuation, and Symbols 477
Extensions in the Main Devanagari Block 477
Devanagari Extended: U+A8E0–U+A8FF 479
Vedic Extensions: U+1CD0–U+1CFF 481
12.2 Bengali (Bangla) 483
12.3 Gurmukhi 490
12.4 Gujarati 495
12.5 Oriya (Odia) 497

12.6 Tamil 501
Tamil: U+0B80–U+0BFF 501
Tamil Vowels 502
Tamil Ligatures 504
Tamil Supplement: U+11FC0–U+11FFF 508
Tamil Named Character Sequences 508
12.7 Telugu 511
12.8 Kannada 514
Kannada: U+0C80–U+0CFF 514
Principles of the Kannada Script 514
Rendering Kannada 516
12.9 Malayalam 518
Malayalam: U+0D00–U+0D7F 518
Malayalam Orthographic Reform 519
Rendering Malayalam 520
Malayalam Numbers and Punctuation 525
13 South and Central Asia-II 527
Other Modern Scripts
13.1 Thaana 529
13.2 Sinhala 531
Sinhala: U+0D80–U+0DFF 531
Sinhala Archaic Numbers: U+111E0–U+111FF 534
13.3 Newa 535
13.4 Tibetan 538
13.5 Mongolian 549
Mongolian: U+1800–U+18AF 549
Mongolian Supplement: U+11660–U+1167F 558
13.6 Limbu 559
13.7 Meetei Mayek 563
Meetei Mayek: U+ABC0–U+ABFF 563
Meetei Mayek Extensions: U+AAE0–U+AAF6 564
13.8 Mro 565
13.9 Warang Citi 566
13.10 Ol Chiki 567
13.11 Chakma 569
13.12 Lepcha 570
13.13 Saurashtra 573
13.14 Masaram Gondi 574
13.15 Gunjala Gondi 576
13.16 Wancho 578

13.17 Toto 579
13.18 Tangsa 580
14 South and Central Asia-III 581
Ancient Scripts
14.1 Brahmi 583
14.2 Kharoshthi 587
Kharoshthi: U+10A00–U+10A5F 587
Rendering Kharoshthi 589
14.3 Bhaiksuki 593
14.4 Phags-pa 595
14.5 Marchen 602
14.6 Zanabazar Square 603
14.7 Soyombo 606
14.8 Old Turkic 608
14.9 Old Sogdian 609
14.10 Sogdian 610
14.11 Old Uyghur 611
15 South and Central Asia-IV 613
Other Historic Scripts
15.1 Syloti Nagri 616
15.2 Kaithi 618
15.3 Sharada 621
15.4 Takri 623
15.5 Siddham 625
15.6 Mahajani 627
15.7 Khojki 629
15.8 Khudawadi 631
15.9 Multani 633
15.10 Tirhuta 634
15.11 Modi 637
15.12 Nandinagari 640
Nandinagari: U+119A0–U+119FF 640
15.13 Grantha 642
Grantha: U+11300–U+1137F 642
Rendering Grantha 642
15.14 Dives Akuru 646
15.15 Ahom 648
15.16 Sora Sompeng 650
15.17 Dogra 651

16 Southeast Asia 653
16.1 Thai 655
16.2 Lao 659
16.3 Myanmar 662
Myanmar: U+1000–U+109F 662
Myanmar Extended-A: U+AA60–U+AA7F 666
Khamti Shan 666
Aiton and Phake 668
Myanmar Extended-B: U+A9E0–U+A9FF 668
16.4 Khmer 669
Khmer: U+1780–U+17FF 669
Principles of the Khmer Script 669
Khmer Symbols: U+19E0–U+19FF 680
16.5 Tai Le 681
16.6 New Tai Lue 683
16.7 Tai Tham 686
16.8 Tai Viet 689
16.9 Kayah Li 692
16.10 Cham 693
16.11 Pahawh Hmong 695
16.12 Nyiakeng Puachue Hmong 697
16.13 Pau Cin Hau 699
16.14 Hanifi Rohingya 700
17 Indonesia and Oceania 701
17.1 Philippine Scripts 702
Tagalog: U+1700–U+171F 702
Hanunóo: U+1720–U+173F 702
Buhid: U+1740–U+175F 702
Tagbanwa: U+1760–U+177F 702
Principles of the Philippine Scripts 702
17.2 Buginese 705
17.3 Balinese 707
17.4 Javanese 713
17.5 Rejang 717
17.6 Batak 718
17.7 Sundanese 720
Sundanese: U+1B80–U+1BBF 720
Sundanese Supplement: U+1CC0–U+1CCF 721
17.8 Makasar 722

18 East Asia 725
18.1 Han 728
CJK Unified Ideographs 728
Blocks Containing Han Ideographs 729
General Characteristics of Han Ideographs 732
Principles of Han Unification 736
Unification Rules 737
Abstract Shape 738
Han Ideograph Arrangement 740
Radical-Stroke Indices 741
Mappings for Han Ideographs 742
CJK Compatibility Ideographs: U+F900–U+FAFF 743
CJK Compatibility Supplement: U+2F800–U+2FA1D 744
Kanbun: U+3190–U+319F 744
Symbols Derived from Han Ideographs 744
CJK and Kangxi Radicals: U+2E80–U+2FD5 745
CJK Additions from HKSCS and GB 18030 746
CJK Strokes: U+31C0–U+31EF 746
Ideographic Symbols and Punctuation: U+16FE0–U+16FFF 747
18.2 Ideographic Description Characters 748
18.3 Bopomofo 752
18.4 Hiragana and Katakana 755
Hiragana: U+3040–U+309F 755
Katakana: U+30A0–U+30FF 755
Katakana Phonetic Extensions: U+31F0–U+31FF 756
Small Kana Extension: U+1B130-U+1B16F 756
Kana Supplement: U+1B000–U+1B0FF 756
Kana Extended-A: U+1B100–U+1B12F 756
Kana Extended-B: U+AFF0-U+1AFFF 758
18.5 Halfwidth and Fullwidth Forms 761
18.6 Hangul 762
Hangul Jamo: U+1100–U+11FF 762
Hangul Jamo Extended-A: U+A960–U+A97F 763
Hangul Jamo Extended-B: U+D7B0–U+D7FF 763
Hangul Compatibility Jamo: U+3130–U+318F 763
Hangul Syllables: U+AC00–U+D7A3 764
18.7 Yi 766
18.8 Nüshu 769
18.9 Lisu 770
18.10 Miao 773

18.11 Tangut 775
Tangut: U+17000–U+187FF 775
Tangut Supplement: U+18D00–U+18D8F 775
Tangut Components: U+18800–U+18AFF 776
18.12 Khitan Small Script 778
19 Africa 781
19.1 Ethiopic 782
Ethiopic: U+1200–U+137F 782
Ethiopic Extensions 785
19.2 Osmanya 786
19.3 Tifinagh 787
19.4 N'Ko 790
19.5 Vai 795
19.6 Bamum 797
Bamum: U+A6A0–U+A6FF 797
Bamum Supplement: U+16800–U+16A3F 797
19.7 Bassa Vah 799
19.8 Mende Kikakui 800
19.9 Adlam 802
19.10 Medefaidrin 804
20 Americas 805
20.1 Cherokee 806
20.2 Canadian Aboriginal Syllabics 808
Unified Canadian Aboriginal Syllabics: U+1400–U+167F 808
Unified Canadian Aboriginal Syllabics Extended: U+18B0–U+18FF 809
Unified Canadian Aboriginal Syllabics Extended-A: U+11AB0–U+11ABF 809
20.3 Osage 810
20.4 Deseret 811
21 Notational Systems 815
21.1 Braille 816
21.2 Western Musical Symbols 818
21.3 Byzantine Musical Symbols 824
21.4 Znamenny Musical Notation 825
21.5 Ancient Greek Musical Notation 826
21.6 Duployan 829
Duployan: U+1BC00–U+1BC9F 829
Shorthand Format Controls: U+1BCA0–U+1BCAF 830
21.7 Sutton SignWriting 831
Sutton SignWriting: U+1D800–U+1DAAF 831

22 Symbols 833
22.1 Currency Symbols 835
Currency Symbols: U+20A0–U+20CF 835
22.2 Letterlike Symbols 839
Letterlike Symbols: U+2100–U+214F 839
Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF 840
Mathematical Alphabets 841
Fonts Used for Mathematical Alphabets 844
Arabic Mathematical Alphabetic Symbols: U+1EE00–U+1EEFF 845
22.3 Numerals 847
Decimal Digits 847
Other Digits 849
Non-Decimal Radix Systems 852
Acrophonic Systems and Other Letter-based Numbers 853
Coptic Epact Numbers: U+102E0–U+102FF 854
Rumi Numeral Symbols: U+10E60–U+10E7E 855
Siyaq Numerical Notation Systems 855
CJK Numerals 857
Fractions 858
Common Indic Number Forms: U+A830–U+A83F 859
22.4 Superscript and Subscript Symbols 860
Superscripts and Subscripts: U+2070–U+209F 860
22.5 Mathematical Symbols 862
Mathematical Operators: U+2200–U+22FF 863
Supplements to Mathematical Symbols and Arrows 866
Supplemental Mathematical Operators: U+2A00–U+2AFF 866
Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF 866
Miscellaneous Mathematical Symbols-B: U+2980–U+29FF 867
Miscellaneous Symbols and Arrows: U+2B00–U+2B7F 868
Arrows: U+2190–U+21FF 868
Supplemental Arrows 869
Standardized Variants of Mathematical Symbols 869
22.6 Invisible Mathematical Operators 870
22.7 Technical Symbols 871
Control Pictures: U+2400–U+243F 871
Miscellaneous Technical: U+2300–U+23FF 871
Optical Character Recognition: U+2440–U+245F 874
Symbols for Legacy Computing: U+1FB00-U+1FBFF 875
22.8 Geometrical Symbols 876
Box Drawing and Block Elements 876
Geometric Shapes: U+25A0–U+25FF 877
Geometric Shapes Extended: U+1F780–U+1F7FF 878

22.9 Miscellaneous Symbols 880
Miscellaneous Symbols and Pictographs 882
Emoticons: U+1F600–U+1F64F 885
Transport and Map Symbols: U+1F680–U+1F6FF 885
Dingbats: U+2700–U+27BF 885
Ornamental Dingbats: U+1F650–U+1F67F 886
Alchemical Symbols: U+1F700–U+1F77F 887
Mahjong Tiles: U+1F000–U+1F02F 887
Domino Tiles: U+1F030–U+1F09F 888
Playing Cards: U+1F0A0–U+1F0FF 888
Chess Symbols: U+1FA00–U+1FA6F 889
Yijing Hexagram Symbols: U+4DC0–U+4DFF 890
Tai Xuan Jing Symbols: U+1D300–U+1D356 890
Ancient Symbols: U+10190–U+101CF 891
Phaistos Disc Symbols: U+101D0–U+101FF 891
22.10 Enclosed and Square 893
Enclosed Alphanumerics: U+2460–U+24FF 895
Enclosed CJK Letters and Months: U+3200–U+32FF 895
CJK Compatibility: U+3300–U+33FF 896
Enclosed Alphanumeric Supplement: U+1F100–U+1F1FF 896
Enclosed Ideographic Supplement: U+1F200–U+1F2FF 897
23 Special Areas and Format Characters 899
23.1 Control Codes 900
Representing Control Sequences 900
Specification of Control Code Semantics 901
23.2 Layout Controls 903
Line and Word Breaking 903
Cursive Connection and Ligatures 905
Prepended Concatenation Marks 909
Combining Grapheme Joiner 910
Bidirectional Ordering Controls 912
Stateful Format Controls 913
23.3 Deprecated Format Characters 915
23.4 Variation Selectors 917
23.5 Private-Use Characters 920
Private Use Area: U+E000–U+F8FF 921
Supplementary Private Use Areas 922
23.6 Surrogates Area 923
23.7 Noncharacters 924

23.8 Specials ... 926
Byte Order Mark (BOM): U+FEFF ... 926
Specials: U+FFF0–U+FFF8 ... 928
Annotation Characters: U+FFF9–U+FFFB ... 928
Replacement Characters: U+FFFC–U+FFFD ... 930
23.9 Tag Characters ... 931
Tag Characters: U+E0000–U+E007F ... 931
Deprecated Use for Language Tagging ... 931
Syntax for Embedding Tags ... 931
Working with Language Tags ... 933
Unicode Conformance Issues ... 935
Formal Tag Syntax ... 935
24 About the Code Charts ... 937
24.1 Character Names List ... 938
Images in the Code Charts and Character Lists ... 938
Special Characters and Code Points ... 940
Character Names ... 942
Informative Aliases ... 942
Normative Aliases ... 943
Cross References ... 944
Information About Languages ... 945
Case Mappings ... 946
Decompositions ... 946
Standardized Variation Sequences ... 947
Positional Forms ... 949
Block Headers ... 949
Subheads ... 950
24.2 CJK and Other Ideographs ... 951
CJK Unified Ideographs ... 951
Compatibility Ideographs ... 953
Tangut Ideographs ... 953
24.3 Hangul Syllables ... 955
A Notational Conventions ... 957
Code Points ... 957
Character Names ... 957
Character Blocks ... 958
Sequences ... 958
Rendering ... 958
Properties and Property Values ... 959
Miscellaneous ... 959
Extended BNF ... 959
Operators ... 961

B Unicode Publications and Resources ... 963
B.1 The Unicode Consortium ... 964
The Unicode Technical Committee ... 964
Other Activities ... 964
B.2 Unicode Publications ... 965
B.3 Other Unicode Online Resources ... 966
Unicode Online Resources ... 966
How to Contact the Unicode Consortium ... 968
C Relationship to ISO/IEC 10646 ... 969
C.1 History ... 970
C.2 Encoding Forms in ISO/IEC 10646 ... 978
Zero Extending ... 978
C.3 UTF-8 and UTF-16 ... 979
UTF-8 ... 979
UTF-16 ... 979
C.4 Synchronization of the Standards ... 980
C.5 Identification of Features for Unicode ... 981
C.6 Character Names ... 982
C.7 Character Functional Specifications ... 983
D Version History of the Standard ... 985
E Han Unification History ... 987
E.1 Development of the URO ... 988
E.2 Continuing Research on Ideographs ... 990
Ideographic Rapporteur Group ... 990
Ideographic Research Group ... 991
E.3 CJK Sources ... 992
F Documentation of CJK Strokes ... 993
I Index ... 1001