

The Unicode® Standard

Version 12.0 – 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.

© 2019 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 12.0.

Includes index.

ISBN 978-1-936213-22-1 (<http://www.unicode.org/versions/Unicode12.0.0/>)

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

QA268.U545 2019

ISBN 978-1-936213-22-1

Published in Mountain View, CA

March 2019

Contents

Preface	xxi
Why Unicode?	xxi
What's New?	xxi
Organization of This Standard	xxii
The Unicode Character Database	xxiv
Unicode Code Charts	xxiv
Unicode Standard Annexes	xxiv
Unicode Technical Standards and Unicode Technical Reports	xxvi
Updates and Errata	xxvi
Acknowledgements	xxvii
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	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 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 Certain Diacritics	65
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	77
	References to Unicode Character Properties	77
	References to Unicode Algorithms	78
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	95
	Types of Properties	95
	Property Values	96
	Default Property Values	97
	Classification of Properties by Their Values	98
	Property Status	99
	Context Dependence	102
	Stability of Properties	102
	Simple and Derived Properties	104
	Property Aliases	104
	Private Use	105
3.6	Combination	106
	Combining Character Sequences	106
	Grapheme Clusters	108
	Application of Combining Marks	110
3.7	Decomposition	116
	Compatibility Decomposition	116
	Canonical Decomposition	117
3.8	Surrogates	119
3.9	Unicode Encoding Forms	120
	UTF-32	124
	UTF-16	125
	UTF-8	125

	Encoding Form Conversion	127
	Constraints on Conversion Processes	127
	U+FFFD Substitution of Maximal Subparts	128
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	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	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 for Text Boundaries	189
4.12	Characters with Unusual Properties	190
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	254
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	280
	CJK Punctuation	282
	Unknown or Unavailable Ideographs	284
	CJK Compatibility Forms	284
7	Europe-I	287
	Modern and Liturgical Scripts	
7.1	Latin	289
	Letters of Basic Latin: U+0041–U+007A	293
	Letters of the Latin-1 Supplement: U+00C0–U+00FF	293
	Latin Extended-A: U+0100–U+017F	293
	Latin Extended-B: U+0180–U+024F	294
	IPA Extensions: U+0250–U+02AF	296
	Phonetic Extensions: U+1D00–U+1DBF	298
	Latin Extended Additional: U+1E00–U+1EFF	299
	Latin Extended-C: U+2C60–U+2C7F	299

	Latin Extended-D: U+A720–U+A7FF	300
	Latin Extended-E: U+AB30–U+AB6F	301
	Latin Ligatures: U+FB00–U+FB06	302
7.2	Greek	303
	Greek: U+0370–U+03FF	303
	Greek Extended: U+1F00–U+1FFF	307
	Ancient Greek Numbers: U+10140–U+1018F	308
7.3	Coptic	310
7.4	Cyrillic	313
	Cyrillic: U+0400–U+04FF	313
	Cyrillic Supplement: U+0500–U+052F	315
	Cyrillic Extended-A: U+2DE0–U+2DFF	315
	Cyrillic Extended-B: U+A640–U+A69F	316
	Cyrillic Extended-C: U+1C80–U+1C8F	316
7.5	Glagolitic	318
	Glagolitic: U+2C00–U+2C5F	318
	Glagolitic Supplement: U+1E000–U+1E02F	318
7.6	Armenian	319
7.7	Georgian	321
	Georgian: U+10A0–U+10FF	321
	Georgian Extended: U+1C90–U+1CBF	321
	Georgian Supplement: U+2D00–U+2D2F	321
7.8	Modifier Letters	323
	Spacing Modifier Letters: U+02B0–U+02FF	324
	Modifier Tone Letters: U+A700–U+A71F	326
7.9	Combining Marks	327
	Combining Diacritical Marks: U+0300–U+036F	332
	Combining Diacritical Marks Extended: U+1AB0–U+1AFF	333
	Combining Diacritical Marks Supplement: U+1DC0–U+1DFF	334
	Combining Diacritical Marks for Symbols: U+20D0–U+20FF	334
	Combining Half Marks: U+FE20–U+FE2F	335
	Combining Marks in Other Blocks	336
8	Europe-II	337
	Ancient and Other Scripts	
8.1	Linear A	339
8.2	Linear B	340
	Linear B Syllabary: U+10000–U+1007F	340
	Linear B Ideograms: U+10080–U+100FF	340
	Aegean Numbers: U+10100–U+1013F	340
8.3	Cypriot Syllabary	342
8.4	Ancient Anatolian Alphabets	343
	Lycian: U+10280–U+1029F	343

	Carian: U+102A0–U+102DF	343
	Lydian: U+10920–U+1093F	343
8.5	Old Italic	345
8.6	Runic	348
8.7	Old Hungarian	351
8.8	Gothic	352
8.9	Elbasan	353
8.10	Caucasian Albanian	354
8.11	Old Permic	355
8.12	Ogham	356
8.13	Shavian	357
9	Middle East-I	359
	Modern and Liturgical Scripts	
9.1	Hebrew	361
	Hebrew: U+0590–U+05FF	361
	Alphabetic Presentation Forms: U+FB1D–U+FB4F	365
9.2	Arabic	367
	Arabic: U+0600–U+06FF	367
	Arabic Cursive Joining	375
	Arabic Ligatures	378
	Arabic Joining Groups	379
	Combining Hamza	386
	Other Letters for Extended Arabic	387
	Arabic Supplement: U+0750–U+077F	388
	Arabic Extended-A: U+08A0–U+08FF	388
	Arabic Presentation Forms-A: U+FB50–U+FDFF	388
	Arabic Presentation Forms-B: U+FE70–U+FEFF	390
9.3	Syriac	391
	Syriac: U+0700–U+074F	391
	Syriac Shaping	395
	Syriac Supplement: U+0860–U+086F	399
9.4	Samaritan	400
9.5	Mandaic	402
10	Middle East-II	405
	Ancient Scripts	
10.1	Old North Arabian	407
10.2	Old South Arabian	408
10.3	Phoenician	410
10.4	Imperial Aramaic	412
10.5	Manichaean	414

10.6 Pahlavi and Parthian	418
Inscriptional Parthian: U+10B40–U+10B5F	418
Inscriptional Pahlavi: U+10B60–U+10B7F	418
Psalter Pahlavi: U+10B80–U+10BAF	419
10.7 Avestan	420
10.8 Elymaic	422
10.9 Nabataean	423
10.10 Palmyrene	424
10.11 Hatran	425
11 Cuneiform and Hieroglyphs	427
11.1 Sumero-Akkadian	428
Cuneiform: U+12000–U+123FF	428
Cuneiform Numbers and Punctuation: U+12400–U+1247F	431
Early Dynastic Cuneiform: U+12480–U+1254F	431
11.2 Ugaritic	432
11.3 Old Persian	433
11.4 Egyptian Hieroglyphs	434
Egyptian Hieroglyphs: U+13000–U+1342F	434
Egyptian Hieroglyph Format Controls: U+13430–U+1343F	436
11.5 Meroitic	440
11.6 Anatolian Hieroglyphs	442
12 South and Central Asia-I	445
Official Scripts of India	
12.1 Devanagari	447
Devanagari: U+0900–U+097F	447
Principles of the Devanagari Script	448
Rendering Devanagari	457
Devanagari Digits, Punctuation, and Symbols	466
Extensions in the Main Devanagari Block	467
Devanagari Extended: U+A8E0–U+A8FF	469
Vedic Extensions: U+1CD0–U+1CFF	470
12.2 Bengali (Bangla)	472
12.3 Gurmukhi	479
12.4 Gujarati	484
12.5 Oriya (Odia)	486
12.6 Tamil	489
Tamil: U+0B80–U+0BFF	489
Tamil Vowels	490
Tamil Ligatures	492
Tamil Supplement: U+11FC0–U+11FFF	496
Tamil Named Character Sequences	496

12.7 Telugu	499
12.8 Kannada	502
Kannada: U+0C80–U+0CFF	502
Principles of the Kannada Script	502
Rendering Kannada	504
12.9 Malayalam	506
Malayalam: U+0D00–U+0D7F	506
Malayalam Orthographic Reform	507
Rendering Malayalam	508
Malayalam Numbers and Punctuation	512
13 South and Central Asia-II	513
Other Modern Scripts	
13.1 Thaana	515
13.2 Sinhala	517
Sinhala: U+0D80–U+0DFF	517
Sinhala Archaic Numbers: U+111E0–U+111FF	518
13.3 Newa	519
13.4 Tibetan	521
13.5 Mongolian	532
Mongolian: U+1800–U+18AF	532
Mongolian Supplement: U+11660–U+1167F	541
13.6 Limbu	542
13.7 Meetei Mayek	546
Meetei Mayek: U+ABC0–U+ABFF	546
Meetei Mayak Extensions: U+AAE0–U+AAF6	547
13.8 Mro	548
13.9 Warang Citi	549
13.10 Ol Chiki	550
13.11 Chakma	552
13.12 Lepcha	553
13.13 Saurashtra	556
13.14 Masaram Gondi	557
13.15 Gunjala Gondi	559
13.16 Wancho	561
14 South and Central Asia-III	563
Ancient Scripts	
14.1 Brahmi	565
14.2 Kharoshthi	569
Kharoshthi: U+10A00–U+10A5F	569
Rendering Kharoshthi	571

14.3	Bhaiksuki	575
14.4	Phags-pa	577
14.5	Marchen	584
14.6	Zanabazar Square	585
14.7	Soyombo	588
14.8	Old Turkic	590
14.9	Old Sogdian	591
14.10	Sogdian	592
15	South and Central Asia-IV	593
	Other Historic Scripts	
15.1	Syloti Nagri	595
15.2	Kaithi	597
15.3	Sharada	600
15.4	Takri	602
15.5	Siddham	604
15.6	Mahajani	606
15.7	Khojki	608
15.8	Khudawadi	610
15.9	Multani	612
15.10	Tirhuta	613
15.11	Modi	616
15.12	Nandinagari	619
	Nandinagari: U+119A0–U+119FF	619
15.13	Grantha	621
	Grantha: U+11300–U+1137F	621
	Rendering Grantha	621
15.14	Ahom	624
15.15	Sora Sompeng	626
15.16	Dogra	627
16	Southeast Asia	629
16.1	Thai	631
16.2	Lao	635
16.3	Myanmar	638
	Myanmar: U+1000–U+109F	638
	Myanmar Extended-A: U+AA60–U+AA7F	642
	Khamti Shan	642
	Aiton and Phake	644
	Myanmar Extended-B: U+A9E0–U+A9FF	644
16.4	Khmer	645
	Khmer: U+1780–U+17FF	645

	Principles of the Khmer Script	645
	Khmer Symbols: U+19E0–U+19FF	656
16.5	Tai Le	657
16.6	New Tai Lue	659
16.7	Tai Tham	662
16.8	Tai Viet	665
16.9	Kayah Li	668
16.10	Cham	669
16.11	Pahawh Hmong	671
16.12	Nyiakeng Puachue Hmong	673
16.13	Pau Cin Hau	675
16.14	Hanifi Rohingya	676
17	Indonesia and Oceania	677
17.1	Philippine Scripts	678
	Tagalog: U+1700–U+171F	678
	Hanunóo: U+1720–U+173F	678
	Buhid: U+1740–U+175F	678
	Tagbanwa: U+1760–U+177F	678
	Principles of the Philippine Scripts	678
17.2	Buginese	681
17.3	Balinese	683
17.4	Javanese	689
17.5	Rejang	693
17.6	Batak	694
17.7	Sundanese	696
	Sundanese: U+1B80–U+1BBF	696
	Sundanese Supplement: U+1CC0–U+1CCF	697
17.8	Makasar	698
18	East Asia	701
18.1	Han	704
	CJK Unified Ideographs	704
	Blocks Containing Han Ideographs	705
	General Characteristics of Han Ideographs	707
	Principles of Han Unification	711
	Unification Rules	712
	Abstract Shape	714
	Han Ideograph Arrangement	716
	Radical-Stroke Indices	717
	Mappings for Han Ideographs	718
	CJK Unified Ideographs Extension B: U+20000–U+2A6D6	718
	CJK Unified Ideographs Extension C: U+2A700–U+2B734	719

	CJK Unified Ideographs Extension D: U+2B740–U+2B81D	719
	CJK Unified Ideographs Extension E: U+2B820–U+2CEA1	719
	CJK Unified Ideographs Extension F: U+2CEB0–U+2EBE0	719
	CJK Compatibility Ideographs: U+F900–U+FAFF	719
	CJK Compatibility Supplement: U+2F800–U+2FA1D	720
	Kanbun: U+3190–U+319F	720
	Symbols Derived from Han Ideographs	721
	CJK and KangXi Radicals: U+2E80–U+2FD5	721
	CJK Additions from HKSCS and GB 18030	722
	CJK Strokes: U+31C0–U+31EF	722
18.2	Ideographic Description Characters	723
18.3	Bopomofo	727
18.4	Hiragana and Katakana	730
	Hiragana: U+3040–U+309F	730
	Katakana: U+30A0–U+30FF	730
	Katakana Phonetic Extensions: U+31F0–U+31FF	731
	Small Kana Extension: U+1B130–U+1B16F	731
	Kana Supplement: U+1B000–U+1B0FF	731
	Kana Extended-A: U+1B100–U+1B12F	731
18.5	Halfwidth and Fullwidth Forms	734
18.6	Hangul	735
	Hangul Jamo: U+1100–U+11FF	735
	Hangul Jamo Extended-A: U+A960–U+A97F	736
	Hangul Jamo Extended-B: U+D7B0–U+D7FF	736
	Hangul Compatibility Jamo: U+3130–U+318F	736
	Hangul Syllables: U+AC00–U+D7A3	737
18.7	Yi	739
18.8	Nüshu	742
18.9	Lisu	743
18.10	Miao	746
18.11	Tangut	748
	Tangut: U+17000–U+187FF	748
	Tangut Components: U+18800–U+18AFF	749
19	Africa	751
19.1	Ethiopic	752
	Ethiopic: U+1200–U+137F	752
	Ethiopic Extensions	754
19.2	Osmanya	756
19.3	Tifinagh	757
19.4	N’Ko	760
19.5	Vai	765

19.6 Bamum	767
Bamum: U+A6A0–U+A6FF	767
Bamum Supplement: U+16800–U+16A3F	767
19.7 Bassa Vah	769
19.8 Mende Kikakui	770
19.9 Adlam	772
19.10 Medefaidrin	774
20 Americas	775
20.1 Cherokee	776
20.2 Canadian Aboriginal Syllabics	778
Canadian Aboriginal Syllabics: U+1400–U+167F	778
Canadian Aboriginal Syllabics Extended: U+18B0–U+18FF	779
20.3 Osage	780
20.4 Deseret	781
21 Notational Systems	785
21.1 Braille	786
21.2 Western Musical Symbols	788
21.3 Byzantine Musical Symbols	794
21.4 Ancient Greek Musical Notation	795
21.5 Duployan	798
Duployan: U+1BC00–U+1BC9F	798
Shorthand Format Controls: U+1BCA0–U+1BCAF	799
21.6 Sutton SignWriting	800
Sutton SignWriting: U+1D800–U+1DAAF	800
22 Symbols	803
22.1 Currency Symbols	805
Currency Symbols: U+20A0–U+20CF	805
22.2 Letterlike Symbols	809
Letterlike Symbols: U+2100–U+214F	809
Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF	810
Mathematical Alphabets	811
Fonts Used for Mathematical Alphabets	813
Arabic Mathematical Alphabetic Symbols: U+1EE00–U+1EEFF	815
22.3 Numerals	816
Decimal Digits	816
Other Digits	818
Non-Decimal Radix Systems	820
Acrophonic Systems and Other Letter-based Numbers	822
Coptic Epact Numbers: U+102E0–U+102FF	823
Rumi Numeral Symbols: U+10E60–U+10E7E	824
Siyaq Numerical Notation Systems	824

	CJK Numerals	825
	Fractions	827
	Common Indic Number Forms: U+A830–U+A83F	827
22.4	Superscript and Subscript Symbols	829
	Superscripts and Subscripts: U+2070–U+209F	829
22.5	Mathematical Symbols	831
	Mathematical Operators: U+2200–U+22FF	832
	Supplements to Mathematical Symbols and Arrows	835
	Supplemental Mathematical Operators: U+2A00–U+2AFF	835
	Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF	835
	Miscellaneous Mathematical Symbols-B: U+2980–U+29FF	836
	Miscellaneous Symbols and Arrows: U+2B00–U+2B7F	837
	Arrows: U+2190–U+21FF	837
	Supplemental Arrows	838
	Standardized Variants of Mathematical Symbols	838
22.6	Invisible Mathematical Operators	839
22.7	Technical Symbols	840
	Control Pictures: U+2400–U+243F	840
	Miscellaneous Technical: U+2300–U+23FF	840
	Optical Character Recognition: U+2440–U+245F	844
22.8	Geometrical Symbols	845
	Box Drawing and Block Elements	845
	Geometric Shapes: U+25A0–U+25FF	846
	Geometric Shapes Extended: U+1F780–U+1F7FF	847
22.9	Miscellaneous Symbols	849
	Miscellaneous Symbols and Pictographs	850
	Emoticons: U+1F600–U+1F64F	853
	Transport and Map Symbols: U+1F680–U+1F6FF	854
	Dingbats: U+2700–U+27BF	854
	Ornamental Dingbats: U+1F650–U+1F67F	855
	Alchemical Symbols: U+1F700–U+1F77F	855
	Mahjong Tiles: U+1F000–U+1F02F	856
	Domino Tiles: U+1F030–U+1F09F	856
	Playing Cards: U+1F0A0–U+1F0FF	857
	Chess Symbols: U+1FA00–U+1FA6F	858
	Yijing Hexagram Symbols: U+4DC0–U+4DFF	858
	Tai Xuan Jing Symbols: U+1D300–U+1D356	859
	Ancient Symbols: U+10190–U+101CF	859
	Phaistos Disc Symbols: U+101D0–U+101FF	860
22.10	Enclosed and Square	861
	Enclosed Alphanumerics: U+2460–U+24FF	863
	Enclosed CJK Letters and Months: U+3200–U+32FF	863
	CJK Compatibility: U+3300–U+33FF	864

	Enclosed Alphanumeric Supplement: U+1F100–U+1F1FF	864
	Enclosed Ideographic Supplement: U+1F200–U+1F2FF	865
23	Special Areas and Format Characters	867
23.1	Control Codes	868
	Representing Control Sequences	868
	Specification of Control Code Semantics	869
23.2	Layout Controls	871
	Line and Word Breaking	871
	Cursive Connection and Ligatures	873
	Combining Grapheme Joiner	877
	Bidirectional Ordering Controls	879
	Stateful Format Controls	880
23.3	Deprecated Format Characters	882
23.4	Variation Selectors	884
23.5	Private-Use Characters	887
	Private Use Area: U+E000–U+F8FF	888
	Supplementary Private Use Areas	889
23.6	Surrogates Area	890
23.7	Noncharacters	891
23.8	Specials	893
	Byte Order Mark (BOM): U+FEFF	893
	Specials: U+FFF0–U+FFF8	895
	Annotation Characters: U+FFF9–U+FFFB	895
	Replacement Characters: U+FFFC–U+FFFD	897
23.9	Tag Characters	898
	Tag Characters: U+E0000–U+E007F	898
	Deprecated Use for Language Tagging	898
	Syntax for Embedding Tags	898
	Working with Language Tags	900
	Unicode Conformance Issues	902
	Formal Tag Syntax	902
24	About the Code Charts	903
24.1	Character Names List	904
	Images in the Code Charts and Character Lists	904
	Special Characters and Code Points	906
	Character Names	908
	Informative Aliases	908
	Normative Aliases	909
	Cross References	910
	Information About Languages	911
	Case Mappings	912
	Decompositions	912
	Standardized Variation Sequences	913

	Positional Forms	915
	Block Headers	916
	Subheads	916
24.2	CJK Ideographs	917
	CJK Unified Ideographs	917
	Compatibility Ideographs	919
24.3	Hangul Syllables	920
A	Notational Conventions	921
	Code Points	921
	Character Names	921
	Character Blocks	921
	Sequences	922
	Rendering	922
	Properties and Property Values	923
	Miscellaneous	923
	Extended BNF	923
	Operators	925
B	Unicode Publications and Resources	927
B.1	The Unicode Consortium	928
	The Unicode Technical Committee	928
	Other Activities	928
B.2	Unicode Publications	929
B.3	Other Unicode Online Resources	930
	Unicode Online Resources	930
	How to Contact the Unicode Consortium	932
C	Relationship to ISO/IEC 10646	933
C.1	History	934
C.2	Encoding Forms in ISO/IEC 10646	942
	Zero Extending	942
C.3	UTF-8 and UTF-16	943
	UTF-8	943
	UTF-16	943
C.4	Synchronization of the Standards	944
C.5	Identification of Features for Unicode	945
C.6	Character Names	946
C.7	Character Functional Specifications	947
D	Version History of the Standard	949
E	Han Unification History	955
E.1	Development of the URO	956
E.2	Ideographic Rapporteur Group	958
E.3	CJK Sources	960

F	Documentation of CJK Strokes	961
I	Index	969