

The Unicode® Standard

Version 13.0 – 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.

© 2020 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 13.0.

Includes index.

ISBN 978-1-936213-26-9 (<http://www.unicode.org/versions/Unicode13.0.0/>)

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

QA268.U545 2020

ISBN 978-1-936213-26-9

Published in Mountain View, CA

March 2020

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	xxvii
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	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
	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	58
	Exhibiting Nonspacing Marks in Isolation	59
	“Characters” and Grapheme Clusters	59
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	106
	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	128
3.11	Normalization Forms	132
	Normalization Stability	132
	Combining Classes	133
	Specification of Unicode Normalization Forms	134
	Starters	134
	Canonical Ordering Algorithm	135
	Canonical Composition Algorithm	136
	Definition of Normalization Forms	138
3.12	Conjoining Jamo Behavior	140
	Definitions	140
	Hangul Syllable Decomposition	142
	Hangul Syllable Composition	144
	Hangul Syllable Name Generation	145
	Sample Code for Hangul Algorithms	146
3.13	Default Case Algorithms	150
	Definitions	150
	Default Case Conversion	152
	Default Case Folding	153
	Default Case Detection	154
	Default Caseless Matching	155
4	Character Properties	157
4.1	Unicode Character Database	159
4.2	Case	162
	Definitions of Case and Casing	162
	Case Mapping	164
4.3	Combining Classes	166
	Reordrant, Split, and Subjoined Combining Marks	166
4.4	Directionality	169
4.5	General Category	170
4.6	Numeric Value	173
	Ideographic Numeric Values	174
4.7	Bidi Mirrored	176
4.8	Name	178
	Unicode Name Property	181
	Code Point Labels	184
	Use of Character Names in APIs and User Interfaces	184
4.9	Unicode 1.0 Names	186
4.10	Letters, Alphabetic, and Ideographic	187
4.11	Properties for Text Boundaries	188

4.12	Characters with Unusual Properties	189
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	230
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	240
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	U+FFFD Substitution in Conversion	252
6	Writing Systems and Punctuation	253
6.1	Writing Systems	254
6.2	General Punctuation	259
	Blocks Devoted to Punctuation	260
	Format Control Characters	261
	Space Characters	262
	Dashes and Hyphens	263
	Paired Punctuation	265
	Language-Based Usage of Quotation Marks	266
	Apostrophes	270
	Other Punctuation	270
	Archaic Punctuation and Editorial Marks	275
	Indic Punctuation	278
	CJK Punctuation	280
	Unknown or Unavailable Ideographs	282
	CJK Compatibility Forms	282
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	300

7.2	Greek	301
	Greek: U+0370–U+03FF	301
	Greek Extended: U+1F00–U+1FFF	305
	Ancient Greek Numbers: U+10140–U+1018F	306
7.3	Coptic	308
7.4	Cyrillic	311
	Cyrillic: U+0400–U+04FF	311
	Cyrillic Supplement: U+0500–U+052F	313
	Cyrillic Extended-A: U+2DE0–U+2DFF	313
	Cyrillic Extended-B: U+A640–U+A69F	314
	Cyrillic Extended-C: U+1C80–U+1C8F	314
7.5	Glagolitic	316
	Glagolitic: U+2C00–U+2C5F	316
	Glagolitic Supplement: U+1E000–U+1E02F	316
7.6	Armenian	317
7.7	Georgian	319
	Georgian: U+10A0–U+10FF	319
	Georgian Extended: U+1C90–U+1CBF	319
	Georgian Supplement: U+2D00–U+2D2F	319
7.8	Modifier Letters	321
	Spacing Modifier Letters: U+02B0–U+02FF	322
	Modifier Tone Letters: U+A700–U+A71F	324
7.9	Combining Marks	325
	Combining Diacritical Marks: U+0300–U+036F	330
	Combining Diacritical Marks Extended: U+1AB0–U+1AFF	331
	Combining Diacritical Marks Supplement: U+1DC0–U+1DFF	332
	Combining Diacritical Marks for Symbols: U+20D0–U+20FF	332
	Combining Half Marks: U+FE20–U+FE2F	333
	Combining Marks in Other Blocks	334
8	Europe-II	335
	Ancient and Other Scripts	
8.1	Linear A	337
8.2	Linear B	338
	Linear B Syllabary: U+10000–U+1007F	338
	Linear B Ideograms: U+10080–U+100FF	338
	Aegean Numbers: U+10100–U+1013F	338
8.3	Cypriot Syllabary	340
8.4	Ancient Anatolian Alphabets	341
	Lycian: U+10280–U+1029F	341
	Carian: U+102A0–U+102DF	341
	Lydian: U+10920–U+1093F	341
8.5	Old Italic	343

8.6	Runic	346
8.7	Old Hungarian	349
8.8	Gothic	350
8.9	Elbasan	351
8.10	Caucasian Albanian	352
8.11	Old Permic	353
8.12	Ogham	354
8.13	Shavian	355
9	Middle East-I	357
	Modern and Liturgical Scripts	
9.1	Hebrew	359
	Hebrew: U+0590–U+05FF	359
	Alphabetic Presentation Forms: U+FB1D–U+FB4F	363
9.2	Arabic	365
	Arabic: U+0600–U+06FF	365
	Arabic Cursive Joining	373
	Arabic Ligatures	376
	Arabic Joining Groups	377
	Combining Hamza	384
	Other Letters for Extended Arabic	386
	Arabic Supplement: U+0750–U+077F	386
	Arabic Extended-A: U+08A0–U+08FF	387
	Arabic Presentation Forms-A: U+FB50–U+FDFF	387
	Arabic Presentation Forms-B: U+FE70–U+FEFF	388
9.3	Syriac	390
	Syriac: U+0700–U+074F	390
	Syriac Shaping	394
	Syriac Supplement: U+0860–U+086F	398
9.4	Samaritan	399
9.5	Mandaic	401
9.6	Yezidi	404
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	Chorasmian	422
10.9	Elymaic	423
10.10	Nabataean	424
10.11	Palmyrene	425
10.12	Hatran	426
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	441
11.6	Anatolian Hieroglyphs	443
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	471
12.2	Bengali (Bangla)	473
12.3	Gurmukhi	480
12.4	Gujarati	485
12.5	Oriya (Odia)	487
12.6	Tamil	491
	Tamil: U+0B80–U+0BFF	491
	Tamil Vowels	492
	Tamil Ligatures	494
	Tamil Supplement: U+11FC0–U+11FFF	498
	Tamil Named Character Sequences	498
12.7	Telugu	501

12.8 Kannada	504
Kannada: U+0C80–U+0CFF	504
Principles of the Kannada Script	504
Rendering Kannada	506
12.9 Malayalam	508
Malayalam: U+0D00–U+0D7F	508
Malayalam Orthographic Reform	509
Rendering Malayalam	510
Malayalam Numbers and Punctuation	515
13 South and Central Asia-II	517
Other Modern Scripts	
13.1 Thaana	519
13.2 Sinhala	521
Sinhala: U+0D80–U+0DFF	521
Sinhala Archaic Numbers: U+111E0–U+111FF	522
13.3 Newa	523
13.4 Tibetan	526
13.5 Mongolian	537
Mongolian: U+1800–U+18AF	537
Mongolian Supplement: U+11660–U+1167F	546
13.6 Limbu	547
13.7 Meetei Mayek	551
Meetei Mayek: U+ABC0–U+ABFF	551
Meetei Mayek Extensions: U+AAE0–U+AAF6	552
13.8 Mro	553
13.9 Warang Citi	554
13.10 Ol Chiki	555
13.11 Chakma	557
13.12 Lepcha	558
13.13 Saurashtra	561
13.14 Masaram Gondi	562
13.15 Gunjala Gondi	564
13.16 Wancho	566
14 South and Central Asia-III	567
Ancient Scripts	
14.1 Brahmi	569
14.2 Kharoshthi	573
Kharoshthi: U+10A00–U+10A5F	573
Rendering Kharoshthi	575
14.3 Bhaiksuki	579

14.4	Phags-pa	581
14.5	Marchen	588
14.6	Zanabazar Square	589
14.7	Soyombo	592
14.8	Old Turkic	594
14.9	Old Sogdian	595
14.10	Sogdian	596
15	South and Central Asia-IV	597
	Other Historic Scripts	
15.1	Syloti Nagri	600
15.2	Kaithi	602
15.3	Sharada	605
15.4	Takri	607
15.5	Siddham	609
15.6	Mahajani	611
15.7	Khojki	613
15.8	Khudawadi	615
15.9	Multani	617
15.10	Tirhuta	618
15.11	Modi	621
15.12	Nandinagari	624
	Nandinagari: U+119A0–U+119FF	624
15.13	Grantha	626
	Grantha: U+11300–U+1137F	626
	Rendering Grantha	626
15.14	Dives Akuru	630
15.15	Ahom	632
15.16	Sora Sompeng	634
15.17	Dogra	635
16	Southeast Asia	637
16.1	Thai	639
16.2	Lao	643
16.3	Myanmar	646
	Myanmar: U+1000–U+109F	646
	Myanmar Extended-A: U+AA60–U+AA7F	650
	Khamti Shan	650
	Aiton and Phake	652
	Myanmar Extended-B: U+A9E0–U+A9FF	652
16.4	Khmer	653
	Khmer: U+1780–U+17FF	653

	Principles of the Khmer Script	653
	Khmer Symbols: U+19E0–U+19FF	664
16.5	Tai Le	665
16.6	New Tai Lue	667
16.7	Tai Tham	670
16.8	Tai Viet	673
16.9	Kayah Li	676
16.10	Cham	677
16.11	Pahawh Hmong	679
16.12	Nyiakeng Puachue Hmong	681
16.13	Pau Cin Hau	683
16.14	Hanifi Rohingya	684
17	Indonesia and Oceania	685
17.1	Philippine Scripts	686
	Tagalog: U+1700–U+171F	686
	Hanunóo: U+1720–U+173F	686
	Buhid: U+1740–U+175F	686
	Tagbanwa: U+1760–U+177F	686
	Principles of the Philippine Scripts	686
17.2	Buginese	689
17.3	Balinese	691
17.4	Javanese	697
17.5	Rejang	701
17.6	Batak	702
17.7	Sundanese	704
	Sundanese: U+1B80–U+1BBF	704
	Sundanese Supplement: U+1CC0–U+1CCF	705
17.8	Makasar	706
18	East Asia	709
18.1	Han	712
	CJK Unified Ideographs	712
	Blocks Containing Han Ideographs	713
	General Characteristics of Han Ideographs	716
	Principles of Han Unification	720
	Unification Rules	721
	Abstract Shape	722
	Han Ideograph Arrangement	724
	Radical-Stroke Indices	725
	Mappings for Han Ideographs	726
	CJK Compatibility Ideographs: U+F900–U+FAFF	727
	CJK Compatibility Supplement: U+2F800–U+2FA1D	727

	Kanbun: U+3190–U+319F	728
	Symbols Derived from Han Ideographs	728
	CJK and Kangxi Radicals: U+2E80–U+2FD5	729
	CJK Additions from HKSCS and GB 18030	730
	CJK Strokes: U+31C0–U+31EF	730
	Ideographic Symbols and Punctuation: U+16FE0–U+16FFF	731
18.2	Ideographic Description Characters	732
18.3	Bopomofo	736
18.4	Hiragana and Katakana	739
	Hiragana: U+3040–U+309F	739
	Katakana: U+30A0–U+30FF	739
	Katakana Phonetic Extensions: U+31F0–U+31FF	740
	Small Kana Extension: U+1B130–U+1B16F	740
	Kana Supplement: U+1B000–U+1B0FF	740
	Kana Extended-A: U+1B100–U+1B12F	740
18.5	Halfwidth and Fullwidth Forms	743
18.6	Hangul	744
	Hangul Jamo: U+1100–U+11FF	744
	Hangul Jamo Extended-A: U+A960–U+A97F	745
	Hangul Jamo Extended-B: U+D7B0–U+D7FF	745
	Hangul Compatibility Jamo: U+3130–U+318F	745
	Hangul Syllables: U+AC00–U+D7A3	746
18.7	Yi	748
18.8	Nüshu	751
18.9	Lisu	752
18.10	Miao	755
18.11	Tangut	757
	Tangut: U+17000–U+187FF	757
	Tangut Supplement: U+18D00–U+18D8F	757
	Tangut Components: U+18800–U+18AFF	758
18.12	Khitan Small Script	760
19	Africa	763
19.1	Ethiopic	764
	Ethiopic: U+1200–U+137F	764
	Ethiopic Extensions	766
19.2	Osmanya	768
19.3	Tifinagh	769
19.4	N’Ko	772
19.5	Vai	777
19.6	Bamum	779
	Bamum: U+A6A0–U+A6FF	779
	Bamum Supplement: U+16800–U+16A3F	779

19.7	Bassa Vah	781
19.8	Mende Kikakui	782
19.9	Adlam	784
19.10	Medefaidrin	786
20	Americas	787
20.1	Cherokee	788
20.2	Canadian Aboriginal Syllabics	790
	Canadian Aboriginal Syllabics: U+1400–U+167F	790
	Canadian Aboriginal Syllabics Extended: U+18B0–U+18FF	791
20.3	Osage	792
20.4	Deseret	793
21	Notational Systems	797
21.1	Braille	798
21.2	Western Musical Symbols	800
21.3	Byzantine Musical Symbols	806
21.4	Ancient Greek Musical Notation	807
21.5	Duployan	810
	Duployan: U+1BC00–U+1BC9F	810
	Shorthand Format Controls: U+1BCA0–U+1BCAF	811
21.6	Sutton SignWriting	812
	Sutton SignWriting: U+1D800–U+1DAAF	812
22	Symbols	815
22.1	Currency Symbols	817
	Currency Symbols: U+20A0–U+20CF	817
22.2	Letterlike Symbols	821
	Letterlike Symbols: U+2100–U+214F	821
	Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF	822
	Mathematical Alphabets	823
	Fonts Used for Mathematical Alphabets	825
	Arabic Mathematical Alphabetic Symbols: U+1EE00–U+1EEFF	827
22.3	Numerals	828
	Decimal Digits	828
	Other Digits	830
	Non-Decimal Radix Systems	832
	Acrophonic Systems and Other Letter-based Numbers	834
	Coptic E pact Numbers: U+102E0–U+102FF	835
	Rumi Numeral Symbols: U+10E60–U+10E7E	836
	Siyaq Numerical Notation Systems	836
	CJK Numerals	838
	Fractions	839
	Common Indic Number Forms: U+A830–U+A83F	840

22.4	Superscript and Subscript Symbols	841
	Superscripts and Subscripts: U+2070–U+209F	841
22.5	Mathematical Symbols	843
	Mathematical Operators: U+2200–U+22FF	844
	Supplements to Mathematical Symbols and Arrows	847
	Supplemental Mathematical Operators: U+2A00–U+2AFF	847
	Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF	847
	Miscellaneous Mathematical Symbols-B: U+2980–U+29FF	848
	Miscellaneous Symbols and Arrows: U+2B00–U+2B7F	849
	Arrows: U+2190–U+21FF	849
	Supplemental Arrows	850
	Standardized Variants of Mathematical Symbols	850
22.6	Invisible Mathematical Operators	851
22.7	Technical Symbols	852
	Control Pictures: U+2400–U+243F	852
	Miscellaneous Technical: U+2300–U+23FF	852
	Optical Character Recognition: U+2440–U+245F	855
	Symbols for Legacy Computing: U+1FB00–U+1FBFF	856
22.8	Geometrical Symbols	857
	Box Drawing and Block Elements	857
	Geometric Shapes: U+25A0–U+25FF	858
	Geometric Shapes Extended: U+1F780–U+1F7FF	859
22.9	Miscellaneous Symbols	861
	Miscellaneous Symbols and Pictographs	863
	Emoticons: U+1F600–U+1F64F	866
	Transport and Map Symbols: U+1F680–U+1F6FF	866
	Dingbats: U+2700–U+27BF	866
	Ornamental Dingbats: U+1F650–U+1F67F	867
	Alchemical Symbols: U+1F700–U+1F77F	868
	Mahjong Tiles: U+1F000–U+1F02F	868
	Domino Tiles: U+1F030–U+1F09F	869
	Playing Cards: U+1F0A0–U+1F0FF	869
	Chess Symbols: U+1FA00–U+1FA6F	870
	Yijing Hexagram Symbols: U+4DC0–U+4DFF	871
	Tai Xuan Jing Symbols: U+1D300–U+1D356	871
	Ancient Symbols: U+10190–U+101CF	872
	Phaistos Disc Symbols: U+101D0–U+101FF	872
22.10	Enclosed and Square	874
	Enclosed Alphanumerics: U+2460–U+24FF	876
	Enclosed CJK Letters and Months: U+3200–U+32FF	876
	CJK Compatibility: U+3300–U+33FF	877
	Enclosed Alphanumeric Supplement: U+1F100–U+1F1FF	877
	Enclosed Ideographic Supplement: U+1F200–U+1F2FF	878

23	Special Areas and Format Characters	881
23.1	Control Codes	882
	Representing Control Sequences	882
	Specification of Control Code Semantics	883
23.2	Layout Controls	885
	Line and Word Breaking	885
	Cursive Connection and Ligatures	887
	Combining Grapheme Joiner	891
	Bidirectional Ordering Controls	893
	Stateful Format Controls	894
23.3	Deprecated Format Characters	896
23.4	Variation Selectors	898
23.5	Private-Use Characters	901
	Private Use Area: U+E000–U+F8FF	902
	Supplementary Private Use Areas	903
23.6	Surrogates Area	904
23.7	Noncharacters	905
23.8	Specials	907
	Byte Order Mark (BOM): U+FEFF	907
	Specials: U+FFF0–U+FFF8	909
	Annotation Characters: U+FFF9–U+FFFB	909
	Replacement Characters: U+FFFC–U+FFFD	911
23.9	Tag Characters	912
	Tag Characters: U+E000–U+E007F	912
	Deprecated Use for Language Tagging	912
	Syntax for Embedding Tags	912
	Working with Language Tags	914
	Unicode Conformance Issues	916
	Formal Tag Syntax	916
24	About the Code Charts	917
24.1	Character Names List	918
	Images in the Code Charts and Character Lists	918
	Special Characters and Code Points	920
	Character Names	922
	Informative Aliases	922
	Normative Aliases	923
	Cross References	924
	Information About Languages	925
	Case Mappings	926
	Decompositions	926
	Standardized Variation Sequences	927
	Positional Forms	929
	Block Headers	929

	Subheads	930
24.2	CJK Ideographs	931
	CJK Unified Ideographs	931
	Compatibility Ideographs	933
24.3	Hangul Syllables	935
A	Notational Conventions	937
	Code Points	937
	Character Names	937
	Character Blocks	938
	Sequences	938
	Rendering	938
	Properties and Property Values	939
	Miscellaneous	939
	Extended BNF	939
	Operators	941
B	Unicode Publications and Resources	943
B.1	The Unicode Consortium	944
	The Unicode Technical Committee	944
	Other Activities	944
B.2	Unicode Publications	945
B.3	Other Unicode Online Resources	946
	Unicode Online Resources	946
	How to Contact the Unicode Consortium	948
C	Relationship to ISO/IEC 10646	949
C.1	History	950
C.2	Encoding Forms in ISO/IEC 10646	958
	Zero Extending	958
C.3	UTF-8 and UTF-16	959
	UTF-8	959
	UTF-16	959
C.4	Synchronization of the Standards	960
C.5	Identification of Features for Unicode	961
C.6	Character Names	962
C.7	Character Functional Specifications	963
D	Version History of the Standard	965
E	Han Unification History	967
E.1	Development of the URO	968
E.2	Continuing Research on Ideographs	970
	Ideographic Rapporteur Group	970
	Ideographic Research Group	971
E.3	CJK Sources	972

F	Documentation of CJK Strokes	973
I	Index	981