

The Unicode® Standard

Version 9.0 – 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.

© 2016 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 9.0.

Includes bibliographical references and index.

ISBN 978-1-936213-13-9 (<http://www.unicode.org/versions/Unicode9.0.0/>)

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

QA268.U545 2016

ISBN 978-1-936213-13-9

Published in Mountain View, CA

July 2016

Contents

Figures	xxi
Tables	xxv
Preface	xxx
Why Unicode?	xxx
What's New?	xxx
Organization of This Standard	xxxii
Unicode Standard Annexes	xxxiii
The Unicode Character Database	xxxv
Unicode Code Charts	xxxv
Unicode Technical Standards and Unicode Technical Reports	xxxv
Updates and Errata	xxxvi
Acknowledgements	xxxvi
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 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	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	182
	Unicode Name Property	185

	Code Point Labels	187
	Use of Character Names in APIs and User Interfaces	188
4.9	Unicode 1.0 Names	189
4.10	Letters, Alphabetic, and Ideographic	190
4.11	Properties Related to Text Boundaries	191
4.12	Characters with Unusual Properties	192
5	Implementation Guidelines	197
5.1	Data Structures for Character Conversion	198
	Issues	198
	Multistage Tables	198
5.2	Programming Languages and Data Types	201
	Unicode Data Types for C	201
5.3	Unknown and Missing Characters	203
5.4	Handling Surrogate Pairs in UTF-16	205
5.5	Handling Numbers	207
5.6	Normalization	208
5.7	Compression	210
5.8	Newline Guidelines	211
	Definitions	211
	Line Separator and Paragraph Separator	212
	Recommendations	213
5.9	Regular Expressions	216
5.10	Language Information in Plain Text	217
	Requirements for Language Tagging	217
	Language Tags and Han Unification	217
5.11	Editing and Selection	219
	Consistent Text Elements	219
5.12	Strategies for Handling Nonspacing Marks	221
	Keyboard Input	222
	Truncation	222
5.13	Rendering Nonspacing Marks	224
	Canonical Equivalence	227
	Positioning Methods	228
5.14	Locating Text Element Boundaries	230
5.15	Identifiers	231
5.16	Sorting and Searching	232
	Culturally Expected Sorting and Searching	232
	Language-Insensitive Sorting	232
	Searching	233
	Sublinear Searching	233

5.17	Binary Order	235
	UTF-8 in UTF-16 Order	235
	UTF-16 in UTF-8 Order	236
5.18	Case Mappings	238
	Titlecasing	238
	Complications for Case Mapping	239
	Reversibility	241
	Caseless Matching	242
	Normalization and Casing	244
5.19	Mapping Compatibility Variants	246
5.20	Unicode Security	248
5.21	Ignoring Characters in Processing	251
	Characters Ignored in Text Segmentation	251
	Characters Ignored in Line Breaking	252
	Characters Ignored in Cursive Joining	252
	Characters Ignored in Identifiers	252
	Characters Ignored in Searching and Sorting	253
	Characters Ignored for Display	253
5.22	Best Practice for U+FFFD Substitution	257
6	Writing Systems and Punctuation	259
6.1	Writing Systems	260
6.2	General Punctuation	265
	Blocks Devoted to Punctuation	266
	Format Control Characters	267
	Space Characters	268
	Dashes and Hyphens	269
	Paired Punctuation	271
	Language-Based Usage of Quotation Marks	272
	Apostrophes	276
	Other Punctuation	276
	Archaic Punctuation and Editorial Marks	281
	Indic Punctuation	284
	CJK Punctuation	286
	Unknown or Unavailable Ideographs	287
	CJK Compatibility Forms	288
7	Europe-I	291
	Modern and Liturgical Scripts	
7.1	Latin	293
	Letters of Basic Latin: U+0041–U+007A	297
	Letters of the Latin-1 Supplement: U+00C0–U+00FF	297
	Latin Extended-A: U+0100–U+017F	297
	Latin Extended-B: U+0180–U+024F	298
	IPA Extensions: U+0250–U+02AF	300

	Phonetic Extensions: U+1D00–U+1DBF	302
	Latin Extended Additional: U+1E00–U+1EFF	303
	Latin Extended-C: U+2C60–U+2C7F	303
	Latin Extended-D: U+A720–U+A7FF	304
	Latin Extended-E: U+AB30–U+AB6F	305
	Latin Ligatures: U+FB00–U+FB06	305
7.2	Greek	306
	Greek: U+0370–U+03FF	306
	Greek Extended: U+1F00–U+1FFF	310
	Ancient Greek Numbers: U+10140–U+1018F	311
7.3	Coptic	313
7.4	Cyrillic	316
	Cyrillic: U+0400–U+04FF	316
	Cyrillic Supplement: U+0500–U+052F	318
	Cyrillic Extended-A: U+2DE0–U+2DFF	318
	Cyrillic Extended-B: U+A640–U+A69F	319
7.5	Glagolitic	320
7.6	Armenian	321
7.7	Georgian	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 Marks for Symbols: U+20D0–U+20FF	336
	Combining Half Marks: U+FE20–U+FE2F	337
	Combining Marks in Other Blocks	337
8	Europe-II	339
	Ancient and Other Scripts	
8.1	Linear A	341
8.2	Linear B	342
	Linear B Syllabary: U+10000–U+1007F	342
	Linear B Ideograms: U+10080–U+100FF	342
	Aegean Numbers: U+10100–U+1013F	342
8.3	Cypriot Syllabary	344
8.4	Ancient Anatolian Alphabets	345
	Lycian: U+10280–U+1029F	345
	Carian: U+102A0–U+102DF	345
	Lydian: U+10920–U+1093F	345
8.5	Old Italic	347

8.6	Runic	350
8.7	Old Hungarian	353
8.8	Gothic	354
8.9	Elbasan	355
8.10	Caucasian Albanian	356
8.11	Old Permic	357
8.12	Ogham	358
8.13	Shavian	359
9	Middle East-I	361
	Modern and Liturgical Scripts	
9.1	Hebrew	363
	Hebrew: U+0590–U+05FF.....	363
	Alphabetic Presentation Forms: U+FB1D–U+FB4F.....	367
9.2	Arabic	369
	Arabic: U+0600–U+06FF.....	369
	Arabic Cursive Joining.....	377
	Arabic Ligatures.....	380
	Arabic Joining Groups.....	381
	Arabic Supplement: U+0750–U+077F.....	389
	Arabic Extended-A: U+08A0–U+08FF.....	389
	Arabic Presentation Forms-A: U+FB50–U+FDFF.....	389
	Arabic Presentation Forms-B: U+FE70–U+FEFF.....	391
9.3	Syriac	392
	Syriac: U+0700–U+074F.....	392
	Syriac Shaping.....	396
9.4	Samaritan	401
9.5	Mandaic	403
10	Middle East-II	407
	Ancient Scripts	
10.1	Old North Arabian	409
10.2	Old South Arabian	410
10.3	Phoenician	412
10.4	Imperial Aramaic	414
10.5	Manichaean	416
10.6	Pahlavi and Parthian	420
	Inscriptional Parthian: U+10B40–U+10B5F.....	420
	Inscriptional Pahlavi: U+10B60–U+10B7F.....	420
	Psalter Pahlavi: U+10B80–U+10BAF.....	421
10.7	Avestan	422
10.8	Nabataean	424

10.9	Palmyrene	425
10.10	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
11.5	Meroitic	439
11.6	Anatolian Hieroglyphs	441
12	South and Central Asia-I	443
	Official Scripts of India	
12.1	Devanagari	445
	Devanagari: U+0900–U+097F.....	445
	Principles of the Devanagari Script.....	446
	Rendering Devanagari.....	452
	Devanagari Digits, Punctuation, and Symbols.....	461
	Extensions in the Main Devanagari Block.....	462
	Devanagari Extended: U+A8E0–U+A8FF.....	464
	Vedic Extensions: U+1CD0–U+1CFF.....	465
12.2	Bengali (Bangla)	467
12.3	Gurmukhi	473
12.4	Gujarati	478
12.5	Oriya (Odia)	480
12.6	Tamil	483
	Tamil: U+0B80–U+0BFF.....	483
	Tamil Vowels.....	484
	Tamil Ligatures.....	485
	Tamil Named Character Sequences.....	489
12.7	Telugu	492
12.8	Kannada	495
	Kannada: U+0C80–U+0CFF.....	495
	Principles of the Kannada Script.....	495
	Rendering Kannada.....	497
12.9	Malayalam	499
	Malayalam Orthographic Reform.....	500
	Rendering Malayalam.....	501
	Malayalam Numbers and Punctuation.....	505

13 South and Central Asia-II	507
Other Modern Scripts	
13.1 Thaana	509
13.2 Sinhala	511
Sinhala Archaic Numbers: U+111E0–U+111FF	512
13.3 Newa	513
13.4 Tibetan	515
13.5 Mongolian	527
13.6 Limbu	536
13.7 Meetei Mayek	540
13.8 Mro	542
13.9 Warang Citi	543
13.10 Ol Chiki	544
13.11 Chakma	546
13.12 Lepcha	547
13.13 Saurashtra	550
14 South and Central Asia-III	551
Ancient Scripts	
14.1 Brahmi	552
14.2 Kharoshthi	556
Kharoshthi: U+10A00–U+10A5F	556
Rendering Kharoshthi	558
14.3 Bhaiksuki	562
14.4 Phags-pa	564
14.5 Marchen	571
14.6 Old Turkic	572
15 South and Central Asia-IV	573
Other Historic Scripts	
15.1 Syloti Nagri	575
15.2 Kaithi	577
15.3 Sharada	580
15.4 Takri	582
15.5 Siddham	584
15.6 Mahajani	586
15.7 Khojki	588
15.8 Khudawadi	590
15.9 Multani	592
15.10 Tirhuta	593
15.11 Modi	596

15.12	Grantha	599
	Rendering Behavior	599
15.13	Ahom	602
15.14	Sora Sompeng	604
16	Southeast Asia	605
16.1	Thai	607
16.2	Lao	611
16.3	Myanmar	614
	Myanmar: U+1000–U+109F	614
	Myanmar Extended-A: U+AA60–U+AA7F	618
	Khamti Shan	618
	Aiton and Phake	620
	Myanmar Extended-B: U+A9E0–U+A9FF	620
16.4	Khmer	621
	Khmer: U+1780–U+17FF	621
	Principles of the Khmer Script	621
	Khmer Symbols: U+19E0–U+19FF	631
16.5	Tai Le	632
16.6	New Tai Lue	634
16.7	Tai Tham	637
16.8	Tai Viet	640
16.9	Kayah Li	643
16.10	Cham	644
16.11	Pahawh Hmong	646
16.12	Pau Cin Hau	648
17	Indonesia and Oceania	649
17.1	Philippine Scripts	650
	Tagalog: U+1700–U+171F	650
	Hanunóo: U+1720–U+173F	650
	Buhid: U+1740–U+175F	650
	Tagbanwa: U+1760–U+177F	650
	Principles of the Philippine Scripts	650
17.2	Buginese	653
17.3	Balinese	655
17.4	Javanese	661
17.5	Rejang	665
17.6	Batak	666
17.7	Sundanese	667
	Sundanese: U+1B80–U+1BBF	667
	Sundanese Supplement: U+1CC0–U+1CCF	668

18 East Asia	669
18.1 Han	671
CJK Unified Ideographs	671
Blocks Containing Han Ideographs	672
General Characteristics of Han Ideographs	674
Principles of Han Unification	678
Unification Rules	679
Abstract Shape	680
Han Ideograph Arrangement	682
Radical-Stroke Indices	683
Mappings for Han Ideographs	684
CJK Unified Ideographs Extension B: U+20000–U+2A6D6	685
CJK Unified Ideographs Extension C: U+2A700–U+2B734	685
CJK Unified Ideographs Extension D: U+2B740–U+2B81D	685
CJK Unified Ideographs Extension E: U+2B820–U+2CEA1	685
CJK Compatibility Ideographs: U+F900–U+FAFF	685
CJK Compatibility Supplement: U+2F800–U+2FA1D	686
Kanbun: U+3190–U+319F	686
Symbols Derived from Han Ideographs	687
CJK and KangXi Radicals: U+2E80–U+2FD5	687
CJK Additions from HKSCS and GB 18030	688
CJK Strokes: U+31C0–U+31EF	688
18.2 Ideographic Description Characters	689
18.3 Bopomofo	693
18.4 Hiragana and Katakana	696
Hiragana: U+3040–U+309F	696
Katakana: U+30A0–U+30FF	696
Katakana Phonetic Extensions: U+31F0–U+31FF	697
Kana Supplement U+1B000–U+1B0FF	697
18.5 Halfwidth and Fullwidth Forms	698
18.6 Hangul	699
Hangul Jamo: U+1100–U+11FF	699
Hangul Jamo Extended-A: U+A960–U+A97F	700
Hangul Jamo Extended-B: U+D7B0–U+D7FF	700
Hangul Compatibility Jamo: U+3130–U+318F	700
Hangul Syllables: U+AC00–U+D7A3	701
18.7 Yi	703
18.8 Lisu	706
18.9 Miao	709
18.10 Tangut	711
Tangut: U+17000–U+187FF	711
Tangut Components: U+18800–U+18AFF	712

19 Africa	713
19.1 Ethiopic	714
Ethiopic: U+1200–U+137F	714
Ethiopic Extensions	716
19.2 Osmanya	718
19.3 Tifinagh	719
19.4 N’Ko	722
19.5 Vai	727
19.6 Bamum	729
Bamum: U+A6A0–U+A6FF	729
Bamum Supplement: U+16800–U+16A3F	729
19.7 Bassa Vah	731
19.8 Mende Kikakui	732
19.9 Adlam	734
20 Americas	737
20.1 Cherokee	738
20.2 Canadian Aboriginal Syllabics	740
Canadian Aboriginal Syllabics: U+1400–U+167F	740
Canadian Aboriginal Syllabics Extended: U+18B0–U+18FF	741
20.3 Osage	742
20.4 Deseret	743
21 Notational Systems	747
21.1 Braille	748
21.2 Western Musical Symbols	750
21.3 Byzantine Musical Symbols	755
21.4 Ancient Greek Musical Notation	756
21.5 Duployan	759
Shorthand Format Controls: U+1BCA0–U+1BCAF	760
21.6 Sutton SignWriting	761
22 Symbols	763
22.1 Currency Symbols	765
22.2 Letterlike Symbols	768
Letterlike Symbols: U+2100–U+214F	768
Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF	769
Mathematical Alphabets	770
Fonts Used for Mathematical Alphabets	772
Arabic Mathematical Alphabetic Symbols: U+1EE00–U+1EEFF	774
22.3 Numerals	775
Decimal Digits	775
Other Digits	777

Non-Decimal Radix Systems	779
Acrophonic Systems and Other Letter-based Numbers	780
Coptic Epact Numbers: U+102E0–U+102FF	781
Rumi Numeral Symbols: U+10E60–U+10E7E	782
CJK Numerals	783
Fractions	784
Common Indic Number Forms: U+A830–U+A83F	785
22.4 Superscript and Subscript Symbols	786
Superscripts and Subscripts: U+2070–U+209F	786
22.5 Mathematical Symbols	788
Mathematical Operators: U+2200–U+22FF	789
Supplements to Mathematical Symbols and Arrows	792
Supplemental Mathematical Operators: U+2A00–U+2AFF	792
Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF	792
Miscellaneous Mathematical Symbols-B: U+2980–U+29FF	793
Miscellaneous Symbols and Arrows: U+2B00–U+2B7F	794
Arrows: U+2190–U+21FF	794
Supplemental Arrows	794
Standardized Variants of Mathematical Symbols	795
22.6 Invisible Mathematical Operators	796
22.7 Technical Symbols	797
Control Pictures: U+2400–U+243F	797
Miscellaneous Technical: U+2300–U+23FF	797
Optical Character Recognition: U+2440–U+245F	801
22.8 Geometrical Symbols	802
Box Drawing and Block Elements	802
Geometric Shapes: U+25A0–U+25FF	803
Geometric Shapes Extended: U+1F780–U+1F7FF	804
22.9 Miscellaneous Symbols	805
Miscellaneous Symbols and Pictographs	806
Emoticons: U+1F600–U+1F64F	809
Transport and Map Symbols: U+1F680–U+1F6FF	809
Dingbats: U+2700–U+27BF	809
Ornamental Dingbats: U+1F650–U+1F67F	811
Alchemical Symbols: U+1F700–U+1F77F	811
Mahjong Tiles: U+1F000–U+1F02F	812
Domino Tiles: U+1F030–U+1F09F	812
Playing Cards: U+1F0A0–U+1F0FF	812
Yijing Hexagram Symbols: U+4DC0–U+4DFF	813
Tai Xuan Jing Symbols: U+1D300–U+1D356	814
Ancient Symbols: U+10190–U+101CF	814
Phaistos Disc Symbols: U+101D0–U+101FF	815

22.10 Enclosed and Square	816
Enclosed Alphanumerics: U+2460–U+24FF	818
Enclosed CJK Letters and Months: U+3200–U+32FF	818
CJK Compatibility: U+3300–U+33FF	819
Enclosed Alphanumeric Supplement: U+1F100–U+1F1FF	819
Enclosed Ideographic Supplement: U+1F200–U+1F2FF	820
23 Special Areas and Format Characters	821
23.1 Control Codes	822
Representing Control Sequences	822
Specification of Control Code Semantics	823
23.2 Layout Controls	825
Line and Word Breaking	825
Cursive Connection and Ligatures	827
Combining Grapheme Joiner	831
Bidirectional Ordering Controls	833
Stateful Format Controls	834
23.3 Deprecated Format Characters	836
23.4 Variation Selectors	838
23.5 Private-Use Characters	841
Private Use Area: U+E000–U+F8FF	842
Supplementary Private Use Areas	843
23.6 Surrogates Area	844
23.7 Noncharacters	845
23.8 Specials	847
Byte Order Mark (BOM): U+FEFF	847
Specials: U+FFF0–U+FFF8	849
Annotation Characters: U+FFF9–U+FFFB	849
Replacement Characters: U+FFFC–U+FFFD	851
23.9 Tag Characters	852
Tag Characters: U+E0000–U+E007F	852
Deprecated Use for Language Tagging	852
Syntax for Embedding Tags	852
Working with Language Tags	854
Unicode Conformance Issues	856
Formal Tag Syntax	856
24 About the Code Charts	857
24.1 Character Names List	858
Images in the Code Charts and Character Lists	858
Special Characters and Code Points	860
Character Names	862
Informative Aliases	862
Normative Aliases	863
Cross References	864

	Information About Languages	865
	Case Mappings	866
	Decompositions	866
	Standardized Variation Sequences	867
	Subheads	869
24.2	CJK Ideographs	870
	CJK Unified Ideographs	870
	Compatibility Ideographs	872
24.3	Hangul Syllables	873
A	Notational Conventions	875
	Code Points	875
	Character Names	875
	Character Blocks	875
	Sequences	876
	Rendering	876
	Properties and Property Values	877
	Miscellaneous	877
	Extended BNF	877
	Operators	879
B	Unicode Publications and Resources	881
B.1	The Unicode Consortium	882
	The Unicode Technical Committee	882
	Other Activities	882
B.2	Unicode Publications	883
B.3	Unicode Technical Standards	884
	UTS #6: A Standard Compression Scheme for Unicode	884
	UTS #10: Unicode Collation Algorithm	884
	UTS #18: Unicode Regular Expressions	884
	UTS #22: Character Mapping Markup Language (CharMapML)	884
	UTS #35: Unicode Locale Data Markup Language (LDML)	884
	UTS #37: Unicode Ideographic Variation Database	884
	UTS #39: Unicode Security Mechanisms	884
	UTS #46: Unicode IDNA Compatibility Processing	885
B.4	Unicode Technical Reports	886
	UTR #16: UTF-EBCDIC	886
	UTR #17: Unicode Character Encoding Model	886
	UTR #23: The Unicode Character Property Model	886
	UTR #25: Unicode Support for Mathematics	886
	UTR #26: Compatibility Encoding Scheme for UTF-16: 8-Bit (CESU-8)	886
	UTR #33: Unicode Conformance Model	886
	UTR #36: Unicode Security Considerations	887
	UTR #50: Unicode Vertical Text Layout	887
	UTR #51: Unicode Emoji	887

B.5	Unicode Technical Notes	888
B.6	Other Unicode Online Resources	889
	Unicode Online Resources	889
	How to Contact the Unicode Consortium	891
C	Relationship to ISO/IEC 10646	893
C.1	History	894
C.2	Encoding Forms in ISO/IEC 10646	900
	Zero Extending	900
C.3	UTF-8 and UTF-16	901
	UTF-8	901
	UTF-16	901
C.4	Synchronization of the Standards	902
C.5	Identification of Features for Unicode	903
C.6	Character Names	904
C.7	Character Functional Specifications	905
D	Version History of the Standard	907
E	Han Unification History	913
E.1	Development of the URO	914
E.2	Ideographic Rapporteur Group	916
E.3	CJK Sources	918
F	Documentation of CJK Strokes	923
R	References	931
R.1	Source Standards and Specifications	932
R.2	Source Dictionaries for Han	941
R.3	Other Script Sources	942
R.4	Selected Resources: Technical	970
R.5	Selected Resources: Other	973
I	Index	979

