

# The Unicode® Standard

## Version 10.0 – 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.

© 2017 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 10.0.

Includes bibliographical references and index.

ISBN 978-1-936213-16-0 (<http://www.unicode.org/versions/Unicode10.0.0/>)

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

QA268.U545 2017

ISBN 978-1-936213-16-0

Published in Mountain View, CA

June 2017

# Contents

<b>Figures</b> .....	xxi
<b>Tables</b> .....	xxv
<b>Preface</b> .....	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
<b>1 Introduction</b> .....	1
<b>1.1 Coverage</b> .....	3
Standards Coverage .....	3
New Characters .....	3
<b>1.2 Design Goals</b> .....	4
<b>1.3 Text Handling</b> .....	6
Characters and Glyphs .....	6
Text Elements .....	6
<b>2 General Structure</b> .....	9
<b>2.1 Architectural Context</b> .....	10
Basic Text Processes .....	10
Text Elements, Characters, and Text Processes .....	10
Text Processes and Encoding .....	11
<b>2.2 Unicode Design Principles</b> .....	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

<b>2.3</b>	<b>Compatibility Characters</b> .....	25
	Compatibility Variants .....	26
	Compatibility Decomposable Characters .....	26
<b>2.4</b>	<b>Code Points and Characters</b> .....	29
	Types of Code Points .....	30
<b>2.5</b>	<b>Encoding Forms</b> .....	33
	UTF-32 .....	35
	UTF-16 .....	36
	UTF-8 .....	36
	Comparison of the Advantages of UTF-32, UTF-16, and UTF-8 .....	38
<b>2.6</b>	<b>Encoding Schemes</b> .....	40
<b>2.7</b>	<b>Unicode Strings</b> .....	43
<b>2.8</b>	<b>Unicode Allocation</b> .....	44
	Planes .....	44
	Allocation Areas and Blocks .....	45
	Assignment of Code Points .....	46
<b>2.9</b>	<b>Details of Allocation</b> .....	48
	Plane 0 (BMP) .....	49
	Plane 1 (SMP) .....	51
	Plane 2 (SIP) .....	52
	Other Planes .....	52
<b>2.10</b>	<b>Writing Direction</b> .....	53
<b>2.11</b>	<b>Combining Characters</b> .....	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
<b>2.12</b>	<b>Equivalent Sequences</b> .....	62
	Normalization .....	62
	Decompositions .....	63
	Non-decomposition of Certain Diacritics .....	65
<b>2.13</b>	<b>Special Characters</b> .....	67
	Special Noncharacter Code Points .....	67
	Byte Order Mark (BOM) .....	67
	Layout and Format Control Characters .....	68
	The Replacement Character .....	68
	Control Codes .....	68
<b>2.14</b>	<b>Conforming to the Unicode Standard</b> .....	69
	Characteristics of Conformant Implementations .....	69
	Unacceptable Behavior .....	70
	Acceptable Behavior .....	70
	Supported Subsets .....	71

<b>3</b>	<b>Conformance</b>	73
<b>3.1</b>	<b>Versions of the Unicode Standard</b>	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
<b>3.2</b>	<b>Conformance Requirements</b>	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
<b>3.3</b>	<b>Semantics</b>	87
	Definitions	87
	Character Identity and Semantics	87
<b>3.4</b>	<b>Characters and Encoding</b>	90
<b>3.5</b>	<b>Properties</b>	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
<b>3.6</b>	<b>Combination</b>	106
	Combining Character Sequences	106
	Grapheme Clusters	108
	Application of Combining Marks	110
<b>3.7</b>	<b>Decomposition</b>	116
	Compatibility Decomposition	116
	Canonical Decomposition	117

<b>3.8</b>	<b>Surrogates</b> .....	119
<b>3.9</b>	<b>Unicode Encoding Forms</b> .....	120
	UTF-32 .....	124
	UTF-16 .....	125
	UTF-8 .....	125
	Encoding Form Conversion .....	127
	Constraints on Conversion Processes .....	127
<b>3.10</b>	<b>Unicode Encoding Schemes</b> .....	131
<b>3.11</b>	<b>Normalization Forms</b> .....	135
	Normalization Stability .....	135
	Combining Classes .....	136
	Specification of Unicode Normalization Forms .....	137
	Starters .....	137
	Canonical Ordering Algorithm .....	138
	Canonical Composition Algorithm .....	139
	Definition of Normalization Forms .....	141
<b>3.12</b>	<b>Conjoining Jamo Behavior</b> .....	143
	Definitions .....	143
	Hangul Syllable Decomposition .....	145
	Hangul Syllable Composition .....	147
	Hangul Syllable Name Generation .....	148
	Sample Code for Hangul Algorithms .....	149
<b>3.13</b>	<b>Default Case Algorithms</b> .....	153
	Definitions .....	153
	Default Case Conversion .....	155
	Default Case Folding .....	156
	Default Case Detection .....	157
	Default Caseless Matching .....	158
<b>4</b>	<b>Character Properties</b> .....	161
<b>4.1</b>	<b>Unicode Character Database</b> .....	163
<b>4.2</b>	<b>Case</b> .....	166
	Definitions of Case and Casing .....	166
	Case Mapping .....	168
<b>4.3</b>	<b>Combining Classes</b> .....	170
	Reordrant, Split, and Subjoined Combining Marks .....	170
<b>4.4</b>	<b>Directionality</b> .....	173
<b>4.5</b>	<b>General Category</b> .....	174
<b>4.6</b>	<b>Numeric Value</b> .....	177
	Ideographic Numeric Values .....	178
<b>4.7</b>	<b>Bidi Mirrored</b> .....	180

4.8	<b>Name</b> .....	182
	Unicode Name Property .....	185
	Code Point Labels .....	187
	Use of Character Names in APIs and User Interfaces .....	188
4.9	<b>Unicode 1.0 Names</b> .....	189
4.10	<b>Letters, Alphabetic, and Ideographic</b> .....	190
4.11	<b>Properties Related to Text Boundaries</b> .....	191
4.12	<b>Characters with Unusual Properties</b> .....	192
5	<b>Implementation Guidelines</b> .....	197
5.1	<b>Data Structures for Character Conversion</b> .....	198
	Issues .....	198
	Multistage Tables .....	198
5.2	<b>Programming Languages and Data Types</b> .....	201
	Unicode Data Types for C .....	201
5.3	<b>Unknown and Missing Characters</b> .....	203
5.4	<b>Handling Surrogate Pairs in UTF-16</b> .....	205
5.5	<b>Handling Numbers</b> .....	207
5.6	<b>Normalization</b> .....	208
5.7	<b>Compression</b> .....	210
5.8	<b>Newline Guidelines</b> .....	211
	Definitions .....	211
	Line Separator and Paragraph Separator .....	212
	Recommendations .....	213
5.9	<b>Regular Expressions</b> .....	216
5.10	<b>Language Information in Plain Text</b> .....	217
	Requirements for Language Tagging .....	217
	Language Tags and Han Unification .....	217
5.11	<b>Editing and Selection</b> .....	219
	Consistent Text Elements .....	219
5.12	<b>Strategies for Handling Nonspacing Marks</b> .....	221
	Keyboard Input .....	222
	Truncation .....	222
5.13	<b>Rendering Nonspacing Marks</b> .....	224
	Canonical Equivalence .....	227
	Positioning Methods .....	228
5.14	<b>Locating Text Element Boundaries</b> .....	230
5.15	<b>Identifiers</b> .....	231
5.16	<b>Sorting and Searching</b> .....	232
	Culturally Expected Sorting and Searching .....	232
	Language-Insensitive Sorting .....	232

	Searching .....	233
	Sublinear Searching .....	233
<b>5.17</b>	<b>Binary Order</b> .....	235
	UTF-8 in UTF-16 Order .....	235
	UTF-16 in UTF-8 Order .....	236
<b>5.18</b>	<b>Case Mappings</b> .....	238
	Titlecasing .....	238
	Complications for Case Mapping .....	239
	Reversibility .....	241
	Caseless Matching .....	242
	Normalization and Casing .....	244
<b>5.19</b>	<b>Mapping Compatibility Variants</b> .....	245
<b>5.20</b>	<b>Unicode Security</b> .....	247
<b>5.21</b>	<b>Ignoring Characters in Processing</b> .....	250
	Characters Ignored in Text Segmentation .....	250
	Characters Ignored in Line Breaking .....	251
	Characters Ignored in Cursive Joining .....	251
	Characters Ignored in Identifiers .....	251
	Characters Ignored in Searching and Sorting .....	252
	Characters Ignored for Display .....	252
<b>5.22</b>	<b>Best Practice for U+FFFD Substitution</b> .....	256
<b>6</b>	<b>Writing Systems and Punctuation</b> .....	259
<b>6.1</b>	<b>Writing Systems</b> .....	260
<b>6.2</b>	<b>General Punctuation</b> .....	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 .....	288
	CJK Compatibility Forms .....	288
<b>7</b>	<b>Europe-I</b> .....	291
	<b>Modern and Liturgical Scripts</b>	
<b>7.1</b>	<b>Latin</b> .....	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
<b>7.2</b>	<b>Greek</b>	306
	Greek: U+0370–U+03FF	306
	Greek Extended: U+1F00–U+1FFF	310
	Ancient Greek Numbers: U+10140–U+1018F	311
<b>7.3</b>	<b>Coptic</b>	313
<b>7.4</b>	<b>Cyrillic</b>	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
	Cyrillic Extended-C: U+1C80–U+1C8F	319
<b>7.5</b>	<b>Glagolitic</b>	321
	Glagolitic: U+2C00–U+2C5F	321
	Glagolitic Supplement: U+1E000–U+1E02F	321
<b>7.6</b>	<b>Armenian</b>	322
<b>7.7</b>	<b>Georgian</b>	324
<b>7.8</b>	<b>Modifier Letters</b>	326
	Spacing Modifier Letters: U+02B0–U+02FF	327
	Modifier Tone Letters: U+A700–U+A71F	329
<b>7.9</b>	<b>Combining Marks</b>	330
	Combining Diacritical Marks: U+0300–U+036F	335
	Combining Diacritical Marks Extended: U+1AB0–U+1AFF	336
	Combining Diacritical Marks Supplement: U+1DC0–U+1DFF	337
	Combining Diacritical Marks for Symbols: U+20D0–U+20FF	337
	Combining Half Marks: U+FE20–U+FE2F	338
	Combining Marks in Other Blocks	339
<b>8</b>	<b>Europe-II</b>	341
	<b>Ancient and Other Scripts</b>	
<b>8.1</b>	<b>Linear A</b>	343
<b>8.2</b>	<b>Linear B</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
<b>8.3</b>	<b>Cypriot Syllabary</b>	346



<b>8.4</b>	<b>Ancient Anatolian Alphabets</b> .....	347
	Lycian: U+10280–U+1029F .....	347
	Carian: U+102A0–U+102DF .....	347
	Lydian: U+10920–U+1093F .....	347
<b>8.5</b>	<b>Old Italic</b> .....	349
<b>8.6</b>	<b>Runic</b> .....	352
<b>8.7</b>	<b>Old Hungarian</b> .....	355
<b>8.8</b>	<b>Gothic</b> .....	356
<b>8.9</b>	<b>Elbasan</b> .....	357
<b>8.10</b>	<b>Caucasian Albanian</b> .....	358
<b>8.11</b>	<b>Old Permic</b> .....	359
<b>8.12</b>	<b>Ogham</b> .....	360
<b>8.13</b>	<b>Shavian</b> .....	361
<b>9</b>	<b>Middle East-I</b> .....	363
	<b>Modern and Liturgical Scripts</b>	
<b>9.1</b>	<b>Hebrew</b> .....	365
	Hebrew: U+0590–U+05FF .....	365
	Alphabetic Presentation Forms: U+FB1D–U+FB4F .....	369
<b>9.2</b>	<b>Arabic</b> .....	371
	Arabic: U+0600–U+06FF .....	371
	Arabic Cursive Joining .....	379
	Arabic Ligatures .....	382
	Arabic Joining Groups .....	383
	Arabic Supplement: U+0750–U+077F .....	391
	Arabic Extended-A: U+08A0–U+08FF .....	391
	Arabic Presentation Forms-A: U+FB50–U+FDFF .....	391
	Arabic Presentation Forms-B: U+FE70–U+FEFF .....	393
<b>9.3</b>	<b>Syriac</b> .....	394
	Syriac: U+0700–U+074F .....	394
	Syriac Shaping .....	398
	Syriac Supplement: U+0860–U+086F .....	402
<b>9.4</b>	<b>Samaritan</b> .....	403
<b>9.5</b>	<b>Mandaic</b> .....	405
<b>10</b>	<b>Middle East-II</b> .....	409
	<b>Ancient Scripts</b>	
<b>10.1</b>	<b>Old North Arabian</b> .....	411
<b>10.2</b>	<b>Old South Arabian</b> .....	412
<b>10.3</b>	<b>Phoenician</b> .....	414
<b>10.4</b>	<b>Imperial Aramaic</b> .....	416
<b>10.5</b>	<b>Manichaean</b> .....	418

<b>10.6 Pahlavi and Parthian</b> .....	422
Inscriptional Parthian: U+10B40–U+10B5F .....	422
Inscriptional Pahlavi: U+10B60–U+10B7F .....	422
Psalter Pahlavi: U+10B80–U+10BAF .....	423
<b>10.7 Avestan</b> .....	424
<b>10.8 Nabataean</b> .....	426
<b>10.9 Palmyrene</b> .....	427
<b>10.10 Hatran</b> .....	428
<b>11 Cuneiform and Hieroglyphs</b> .....	429
<b>11.1 Sumero-Akkadian</b> .....	430
Cuneiform: U+12000–U+123FF .....	430
Cuneiform Numbers and Punctuation: U+12400–U+1247F .....	433
Early Dynastic Cuneiform: U+12480–U+1254F .....	433
<b>11.2 Ugaritic</b> .....	434
<b>11.3 Old Persian</b> .....	435
<b>11.4 Egyptian Hieroglyphs</b> .....	436
<b>11.5 Meroitic</b> .....	441
<b>11.6 Anatolian Hieroglyphs</b> .....	443
<b>12 South and Central Asia-I</b> .....	445
<b>Official Scripts of India</b>	
<b>12.1 Devanagari</b> .....	447
Devanagari: U+0900–U+097F .....	447
Principles of the Devanagari Script .....	448
Rendering Devanagari .....	454
Devanagari Digits, Punctuation, and Symbols .....	463
Extensions in the Main Devanagari Block .....	464
Devanagari Extended: U+A8E0–U+A8FF .....	466
Vedic Extensions: U+1CD0–U+1CFF .....	467
<b>12.2 Bengali (Bangla)</b> .....	469
<b>12.3 Gurmukhi</b> .....	475
<b>12.4 Gujarati</b> .....	480
<b>12.5 Oriya (Odia)</b> .....	482
<b>12.6 Tamil</b> .....	485
Tamil: U+0B80–U+0BFF .....	485
Tamil Vowels .....	486
Tamil Ligatures .....	487
Tamil Named Character Sequences .....	491
<b>12.7 Telugu</b> .....	494

<b>12.8</b>	<b>Kannada</b> .....	497
	Kannada: U+0C80–U+0CFF .....	497
	Principles of the Kannada Script .....	497
	Rendering Kannada .....	499
<b>12.9</b>	<b>Malayalam</b> .....	501
	Malayalam: U+0D00–U+0D7F .....	501
	Malayalam Orthographic Reform .....	503
	Rendering Malayalam .....	504
	Malayalam Numbers and Punctuation .....	508
<b>13</b>	<b>South and Central Asia-II</b> .....	509
	<b>Other Modern Scripts</b>	
<b>13.1</b>	<b>Thaana</b> .....	511
<b>13.2</b>	<b>Sinhala</b> .....	513
	Sinhala: U+0D80–U+0DFF .....	513
	Sinhala Archaic Numbers: U+111E0–U+111FF .....	514
<b>13.3</b>	<b>Newa</b> .....	515
<b>13.4</b>	<b>Tibetan</b> .....	517
<b>13.5</b>	<b>Mongolian</b> .....	528
	Mongolian: U+1800–U+18AF .....	528
	Mongolian Supplement: U+11660–U+1167F .....	536
<b>13.6</b>	<b>Limbu</b> .....	537
<b>13.7</b>	<b>Meetei Mayek</b> .....	541
	Meetei Mayek: U+ABC0–U+ABFF .....	541
	Meetei Mayek Extensions: U+AAE0–U+AAF6 .....	542
<b>13.8</b>	<b>Mro</b> .....	543
<b>13.9</b>	<b>Warang Citi</b> .....	544
<b>13.10</b>	<b>Ol Chiki</b> .....	545
<b>13.11</b>	<b>Chakma</b> .....	547
<b>13.12</b>	<b>Lepcha</b> .....	548
<b>13.13</b>	<b>Saurashtra</b> .....	551
<b>13.14</b>	<b>Masaram Gondi</b> .....	552
<b>14</b>	<b>South and Central Asia-III</b> .....	555
	<b>Ancient Scripts</b>	
<b>14.1</b>	<b>Brahmi</b> .....	556
<b>14.2</b>	<b>Kharoshthi</b> .....	560
	Kharoshthi: U+10A00–U+10A5F .....	560
	Rendering Kharoshthi .....	562
<b>14.3</b>	<b>Bhaiksuki</b> .....	566
<b>14.4</b>	<b>Phags-pa</b> .....	568
<b>14.5</b>	<b>Marchen</b> .....	575

14.6	Zanabazar Square	576
14.7	Soyombo	579
14.8	Old Turkic	581
15	South and Central Asia-IV	583
	<b>Other Historic Scripts</b>	
15.1	Syloti Nagri	585
15.2	Kaithi	587
15.3	Sharada	590
15.4	Takri	592
15.5	Siddham	594
15.6	Mahajani	596
15.7	Khojki	598
15.8	Khudawadi	600
15.9	Multani	602
15.10	Tirhuta	603
15.11	Modi	606
15.12	Grantha	609
	Grantha: U+11300–U+1137F	609
	Rendering Grantha	609
15.13	Ahom	612
15.14	Sora Sompeng	614
16	Southeast Asia	615
16.1	Thai	617
16.2	Lao	621
16.3	Myanmar	624
	Myanmar: U+1000–U+109F	624
	Myanmar Extended-A: U+AA60–U+AA7F	628
	Khamti Shan	628
	Aiton and Phake	630
	Myanmar Extended-B: U+A9E0–U+A9FF	630
16.4	Khmer	631
	Khmer: U+1780–U+17FF	631
	Principles of the Khmer Script	631
	Khmer Symbols: U+19E0–U+19FF	641
16.5	Tai Le	642
16.6	New Tai Lue	644
16.7	Tai Tham	647
16.8	Tai Viet	650
16.9	Kayah Li	653
16.10	Cham	654

<b>16.11 Pahawh Hmong</b> .....	656
<b>16.12 Pau Cin Hau</b> .....	658
<b>17 Indonesia and Oceania</b> .....	659
<b>17.1 Philippine Scripts</b> .....	660
Tagalog: U+1700–U+171F .....	660
Hanunóo: U+1720–U+173F .....	660
Buhid: U+1740–U+175F .....	660
Tagbanwa: U+1760–U+177F .....	660
Principles of the Philippine Scripts .....	660
<b>17.2 Buginese</b> .....	663
<b>17.3 Balinese</b> .....	665
<b>17.4 Javanese</b> .....	671
<b>17.5 Rejang</b> .....	675
<b>17.6 Batak</b> .....	676
<b>17.7 Sundanese</b> .....	677
Sundanese: U+1B80–U+1BBF .....	677
Sundanese Supplement: U+1CC0–U+1CCF .....	678
<b>18 East Asia</b> .....	679
<b>18.1 Han</b> .....	682
CJK Unified Ideographs .....	682
Blocks Containing Han Ideographs .....	683
General Characteristics of Han Ideographs .....	685
Principles of Han Unification .....	689
Unification Rules .....	690
Abstract Shape .....	692
Han Ideograph Arrangement .....	694
Radical-Stroke Indices .....	695
Mappings for Han Ideographs .....	696
CJK Unified Ideographs Extension B: U+20000–U+2A6D6 .....	696
CJK Unified Ideographs Extension C: U+2A700–U+2B734 .....	697
CJK Unified Ideographs Extension D: U+2B740–U+2B81D .....	697
CJK Unified Ideographs Extension E: U+2B820–U+2CEA1 .....	697
CJK Unified Ideographs Extension F: U+2CEB0–U+2EBE0 .....	697
CJK Compatibility Ideographs: U+F900–U+FAFF .....	697
CJK Compatibility Supplement: U+2F800–U+2FA1D .....	698
Kanbun: U+3190–U+319F .....	698
Symbols Derived from Han Ideographs .....	699
CJK and KangXi Radicals: U+2E80–U+2FD5 .....	699
CJK Additions from HKSCS and GB 18030 .....	700
CJK Strokes: U+31C0–U+31EF .....	700
<b>18.2 Ideographic Description Characters</b> .....	701
<b>18.3 Bopomofo</b> .....	705

<b>18.4 Hiragana and Katakana</b> .....	708
Hiragana: U+3040–U+309F .....	708
Katakana: U+30A0–U+30FF .....	708
Katakana Phonetic Extensions: U+31F0–U+31FF .....	709
Kana Supplement: U+1B000–U+1B0FF .....	709
Kana Extended-A: U+1B100–U+1B12F .....	709
<b>18.5 Halfwidth and Fullwidth Forms</b> .....	712
<b>18.6 Hangul</b> .....	713
Hangul Jamo: U+1100–U+11FF .....	713
Hangul Jamo Extended-A: U+A960–U+A97F .....	714
Hangul Jamo Extended-B: U+D7B0–U+D7FF .....	714
Hangul Compatibility Jamo: U+3130–U+318F .....	714
Hangul Syllables: U+AC00–U+D7A3 .....	715
<b>18.7 Yi</b> .....	717
<b>18.8 Nüshu</b> .....	720
<b>18.9 Lisu</b> .....	721
<b>18.10 Miao</b> .....	724
<b>18.11 Tangut</b> .....	726
Tangut: U+17000–U+187FF .....	726
Tangut Components: U+18800–U+18AFF .....	727
<b>19 Africa</b> .....	729
<b>19.1 Ethiopic</b> .....	730
Ethiopic: U+1200–U+137F .....	730
Ethiopic Extensions .....	732
<b>19.2 Osmanya</b> .....	734
<b>19.3 Tifinagh</b> .....	735
<b>19.4 N’Ko</b> .....	738
<b>19.5 Vai</b> .....	743
<b>19.6 Bamum</b> .....	745
Bamum: U+A6A0–U+A6FF .....	745
Bamum Supplement: U+16800–U+16A3F .....	745
<b>19.7 Bassa Vah</b> .....	747
<b>19.8 Mende Kikakui</b> .....	748
<b>19.9 Adlam</b> .....	750
<b>20 Americas</b> .....	753
<b>20.1 Cherokee</b> .....	754
<b>20.2 Canadian Aboriginal Syllabics</b> .....	756
Canadian Aboriginal Syllabics: U+1400–U+167F .....	756
Canadian Aboriginal Syllabics Extended: U+18B0–U+18FF .....	757
<b>20.3 Osage</b> .....	758
<b>20.4 Deseret</b> .....	759

<b>21</b>	<b>Notational Systems</b> .....	763
<b>21.1</b>	<b>Braille</b> .....	764
<b>21.2</b>	<b>Western Musical Symbols</b> .....	766
<b>21.3</b>	<b>Byzantine Musical Symbols</b> .....	771
<b>21.4</b>	<b>Ancient Greek Musical Notation</b> .....	772
<b>21.5</b>	<b>Duployan</b> .....	775
	Duployan: U+1BC00–U+1BC9F.....	775
	Shorthand Format Controls: U+1BCA0–U+1BCAF.....	776
<b>21.6</b>	<b>Sutton SignWriting</b> .....	777
<b>22</b>	<b>Symbols</b> .....	779
<b>22.1</b>	<b>Currency Symbols</b> .....	781
	Currency Symbols: U+20A0–U+20CF.....	781
<b>22.2</b>	<b>Letterlike Symbols</b> .....	785
	Letterlike Symbols: U+2100–U+214F.....	785
	Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF.....	786
	Mathematical Alphabets.....	787
	Fonts Used for Mathematical Alphabets.....	789
	Arabic Mathematical Alphabetic Symbols: U+1EE00–U+1EEFF.....	791
<b>22.3</b>	<b>Numerals</b> .....	792
	Decimal Digits.....	792
	Other Digits.....	794
	Non-Decimal Radix Systems.....	796
	Acrophonic Systems and Other Letter-based Numbers.....	797
	Coptic Epact Numbers: U+102E0–U+102FF.....	798
	Rumi Numeral Symbols: U+10E60–U+10E7E.....	799
	CJK Numerals.....	800
	Fractions.....	801
	Common Indic Number Forms: U+A830–U+A83F.....	802
<b>22.4</b>	<b>Superscript and Subscript Symbols</b> .....	803
	Superscripts and Subscripts: U+2070–U+209F.....	803
<b>22.5</b>	<b>Mathematical Symbols</b> .....	805
	Mathematical Operators: U+2200–U+22FF.....	806
	Supplements to Mathematical Symbols and Arrows.....	809
	Supplemental Mathematical Operators: U+2A00–U+2AFF.....	809
	Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF.....	809
	Miscellaneous Mathematical Symbols-B: U+2980–U+29FF.....	810
	Miscellaneous Symbols and Arrows: U+2B00–U+2B7F.....	811
	Arrows: U+2190–U+21FF.....	811
	Supplemental Arrows.....	811
	Standardized Variants of Mathematical Symbols.....	812
<b>22.6</b>	<b>Invisible Mathematical Operators</b> .....	813

<b>22.7</b>	<b>Technical Symbols</b> .....	814
	Control Pictures: U+2400–U+243F .....	814
	Miscellaneous Technical: U+2300–U+23FF .....	814
	Optical Character Recognition: U+2440–U+245F .....	818
<b>22.8</b>	<b>Geometrical Symbols</b> .....	819
	Box Drawing and Block Elements .....	819
	Geometric Shapes: U+25A0–U+25FF .....	820
	Geometric Shapes Extended: U+1F780–U+1F7FF .....	821
<b>22.9</b>	<b>Miscellaneous Symbols</b> .....	822
	Miscellaneous Symbols and Pictographs .....	823
	Emoticons: U+1F600–U+1F64F .....	826
	Transport and Map Symbols: U+1F680–U+1F6FF .....	826
	Dingbats: U+2700–U+27BF .....	826
	Ornamental Dingbats: U+1F650–U+1F67F .....	828
	Alchemical Symbols: U+1F700–U+1F77F .....	828
	Mahjong Tiles: U+1F000–U+1F02F .....	829
	Domino Tiles: U+1F030–U+1F09F .....	829
	Playing Cards: U+1F0A0–U+1F0FF .....	829
	Yijing Hexagram Symbols: U+4DC0–U+4DFF .....	830
	Tai Xuan Jing Symbols: U+1D300–U+1D356 .....	831
	Ancient Symbols: U+10190–U+101CF .....	831
	Phaistos Disc Symbols: U+101D0–U+101FF .....	832
<b>22.10</b>	<b>Enclosed and Square</b> .....	833
	Enclosed Alphanumerics: U+2460–U+24FF .....	835
	Enclosed CJK Letters and Months: U+3200–U+32FF .....	835
	CJK Compatibility: U+3300–U+33FF .....	836
	Enclosed Alphanumeric Supplement: U+1F100–U+1F1FF .....	836
	Enclosed Ideographic Supplement: U+1F200–U+1F2FF .....	837
<b>23</b>	<b>Special Areas and Format Characters</b> .....	839
<b>23.1</b>	<b>Control Codes</b> .....	840
	Representing Control Sequences .....	840
	Specification of Control Code Semantics .....	841
<b>23.2</b>	<b>Layout Controls</b> .....	843
	Line and Word Breaking .....	843
	Cursive Connection and Ligatures .....	845
	Combining Grapheme Joiner .....	849
	Bidirectional Ordering Controls .....	851
	Stateful Format Controls .....	852
<b>23.3</b>	<b>Deprecated Format Characters</b> .....	854
<b>23.4</b>	<b>Variation Selectors</b> .....	856
<b>23.5</b>	<b>Private-Use Characters</b> .....	859
	Private Use Area: U+E000–U+F8FF .....	860
	Supplementary Private Use Areas .....	861



<b>23.6</b>	<b>Surrogates Area</b> .....	862
<b>23.7</b>	<b>Noncharacters</b> .....	863
<b>23.8</b>	<b>Specials</b> .....	865
	Byte Order Mark (BOM): U+FEFF .....	865
	Specials: U+FFF0–U+FFF8 .....	867
	Annotation Characters: U+FFF9–U+FFFB .....	867
	Replacement Characters: U+FFFC–U+FFFD .....	869
<b>23.9</b>	<b>Tag Characters</b> .....	870
	Tag Characters: U+E000–U+E007F .....	870
	Deprecated Use for Language Tagging .....	870
	Syntax for Embedding Tags .....	870
	Working with Language Tags .....	872
	Unicode Conformance Issues .....	874
	Formal Tag Syntax .....	874
<b>24</b>	<b>About the Code Charts</b> .....	875
<b>24.1</b>	<b>Character Names List</b> .....	876
	Images in the Code Charts and Character Lists .....	876
	Special Characters and Code Points .....	878
	Character Names .....	880
	Informative Aliases .....	880
	Normative Aliases .....	881
	Cross References .....	882
	Information About Languages .....	883
	Case Mappings .....	884
	Decompositions .....	884
	Standardized Variation Sequences .....	885
	Positional Forms .....	887
	Block Headers .....	887
	Subheads .....	888
<b>24.2</b>	<b>CJK Ideographs</b> .....	889
	CJK Unified Ideographs .....	889
	Compatibility Ideographs .....	891
<b>24.3</b>	<b>Hangul Syllables</b> .....	892
<b>A</b>	<b>Notational Conventions</b> .....	893
	Code Points .....	893
	Character Names .....	893
	Character Blocks .....	893
	Sequences .....	894
	Rendering .....	894
	Properties and Property Values .....	895
	Miscellaneous .....	895
	Extended BNF .....	895
	Operators .....	897

<b>B</b>	<b>Unicode Publications and Resources</b> .....	899
<b>B.1</b>	<b>The Unicode Consortium</b> .....	900
	The Unicode Technical Committee .....	900
	Other Activities .....	900
<b>B.2</b>	<b>Unicode Publications</b> .....	901
<b>B.3</b>	<b>Other Unicode Online Resources</b> .....	902
	Unicode Online Resources .....	902
	How to Contact the Unicode Consortium .....	904
<b>C</b>	<b>Relationship to ISO/IEC 10646</b> .....	905
<b>C.1</b>	<b>History</b> .....	906
<b>C.2</b>	<b>Encoding Forms in ISO/IEC 10646</b> .....	913
	Zero Extending .....	913
<b>C.3</b>	<b>UTF-8 and UTF-16</b> .....	914
	UTF-8 .....	914
	UTF-16 .....	914
<b>C.4</b>	<b>Synchronization of the Standards</b> .....	915
<b>C.5</b>	<b>Identification of Features for Unicode</b> .....	916
<b>C.6</b>	<b>Character Names</b> .....	917
<b>C.7</b>	<b>Character Functional Specifications</b> .....	918
<b>D</b>	<b>Version History of the Standard</b> .....	919
<b>E</b>	<b>Han Unification History</b> .....	925
<b>E.1</b>	<b>Development of the URO</b> .....	926
<b>E.2</b>	<b>Ideographic Rapporteur Group</b> .....	928
<b>E.3</b>	<b>CJK Sources</b> .....	930
<b>F</b>	<b>Documentation of CJK Strokes</b> .....	931
<b>R</b>	<b>References</b> .....	939
<b>R.1</b>	<b>Source Standards and Specifications</b> .....	940
<b>R.2</b>	<b>Source Dictionaries for Han</b> .....	949
<b>R.3</b>	<b>Other Script Sources</b> .....	950
<b>R.4</b>	<b>Selected Resources: Technical</b> .....	979
<b>R.5</b>	<b>Selected Resources: Other</b> .....	982
<b>I</b>	<b>Index</b> .....	987

