

# The Unicode Standard

## Version 6.2 – 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.

Copyright © 1991–2012 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 Julie D. Allen ... [et al.]. — Version 6.2.

Includes bibliographical references and index.

ISBN 978-1-936213-07-8 (<http://www.unicode.org/versions/Unicode6.2.0/>)

1. Unicode (Computer character set) I. Allen, Julie D. II. Unicode Consortium.

QA268.U545 2012

ISBN 978-1-936213-07-8

Published in Mountain View, CA

September 2012

# Contents

<b>Figures</b> .....	xvii
<b>Tables</b> .....	xxi
<b>Preface</b> .....	xxv
Why Unicode? .....	xxv
What's New? .....	xxv
Organization of This Standard .....	xxv
Unicode Standard Annexes .....	xxvii
The Unicode Character Database .....	xxviii
Unicode Code Charts .....	xxviii
Unicode Technical Standards and Unicode Technical Reports .....	xxviii
Updates and Errata .....	xxviii
Acknowledgements .....	xxix
<b>1 Introduction</b> .....	1
<b>1.1 Coverage</b> .....	2
Standards Coverage .....	2
New Characters .....	3
<b>1.2 Design Goals</b> .....	3
<b>1.3 Text Handling</b> .....	4
Characters and Glyphs .....	5
Text Elements .....	5
<b>2 General Structure</b> .....	7
<b>2.1 Architectural Context</b> .....	7
Basic Text Processes .....	8
Text Elements, Characters, and Text Processes .....	8
Text Processes and Encoding .....	9
<b>2.2 Unicode Design Principles</b> .....	10
Universality .....	10
Efficiency .....	11
Characters, Not Glyphs .....	11
Semantics .....	14
Plain Text .....	14
Logical Order .....	15
Unification .....	17
Dynamic Composition .....	18
Stability .....	18
Convertibility .....	19
<b>2.3 Compatibility Characters</b> .....	19
Compatibility Variants .....	20
Compatibility Decomposable Characters .....	20
<b>2.4 Code Points and Characters</b> .....	21
Types of Code Points .....	22
<b>2.5 Encoding Forms</b> .....	24
UTF-32 .....	26
UTF-16 .....	27
UTF-8 .....	27
Comparison of the Advantages of UTF-32, UTF-16, and UTF-8 .....	28

<b>2.6</b>	<b>Encoding Schemes</b> .....	30
<b>2.7</b>	<b>Unicode Strings</b> .....	32
<b>2.8</b>	<b>Unicode Allocation</b> .....	33
	Planes .....	33
	Allocation Areas and Character Blocks .....	34
	Assignment of Code Points .....	35
<b>2.9</b>	<b>Details of Allocation</b> .....	35
	Plane 0 (BMP) .....	37
	Plane 1 (SMP) .....	38
	Plane 2 (SIP) .....	40
	Other Planes .....	40
<b>2.10</b>	<b>Writing Direction</b> .....	40
<b>2.11</b>	<b>Combining Characters</b> .....	41
	Sequence of Base Characters and Diacritics .....	42
	Multiple Combining Characters .....	43
	Ligated Multiple Base Characters .....	45
	Exhibiting Nonspacing Marks in Isolation .....	45
	“Characters” and Grapheme Clusters .....	46
<b>2.12</b>	<b>Equivalent Sequences and Normalization</b> .....	46
	Normalization .....	47
	Decompositions .....	48
	Non-decomposition of Overlaid Diacritics .....	49
<b>2.13</b>	<b>Special Characters and Noncharacters</b> .....	50
	Special Noncharacter Code Points .....	50
	Byte Order Mark (BOM) .....	50
	Layout and Format Control Characters .....	51
	The Replacement Character .....	51
	Control Codes .....	51
<b>2.14</b>	<b>Conforming to the Unicode Standard</b> .....	51
	Characteristics of Conformant Implementations .....	52
	Unacceptable Behavior .....	52
	Acceptable Behavior .....	53
	Supported Subsets .....	53
<b>3</b>	<b>Conformance</b> .....	55
<b>3.1</b>	<b>Versions of the Unicode Standard</b> .....	55
	Stability .....	56
	Version Numbering .....	56
	Errata and Corrigenda .....	57
	References to the Unicode Standard .....	57
	Precision in Version Citation .....	57
	References to Unicode Character Properties .....	58
	References to Unicode Algorithms .....	58
<b>3.2</b>	<b>Conformance Requirements</b> .....	59
	Code Points Unassigned to Abstract Characters .....	59
	Interpretation .....	59
	Modification .....	61
	Character Encoding Forms .....	62
	Character Encoding Schemes .....	62
	Bidirectional Text .....	63
	Normalization Forms .....	63
	Normative References .....	63
	Unicode Algorithms .....	63

	Default Casing Algorithms	64
	Unicode Standard Annexes	64
<b>3.3</b>	<b>Semantics</b>	65
	Definitions	65
	Character Identity and Semantics	65
<b>3.4</b>	<b>Characters and Encoding</b>	67
<b>3.5</b>	<b>Properties</b>	70
	Types of Properties	70
	Property Values	71
	Default Property Values	72
	Classification of Properties by Their Values	73
	Property Status	73
	Context Dependence	76
	Stability of Properties	76
	Simple and Derived Properties	78
	Property Aliases	78
	Private Use	79
<b>3.6</b>	<b>Combination</b>	79
	Combining Character Sequences	79
	Grapheme Clusters	81
	Application of Combining Marks	83
<b>3.7</b>	<b>Decomposition</b>	87
	Compatibility Decomposition	87
	Canonical Decomposition	88
<b>3.8</b>	<b>Surrogates</b>	89
<b>3.9</b>	<b>Unicode Encoding Forms</b>	89
	UTF-32	93
	UTF-16	93
	UTF-8	94
	Encoding Form Conversion	95
	Constraints on Conversion Processes	96
<b>3.10</b>	<b>Unicode Encoding Schemes</b>	98
<b>3.11</b>	<b>Normalization Forms</b>	101
	Normalization Stability	102
	Combining Classes	102
	Specification of Unicode Normalization Forms	103
	Starters	103
	Canonical Ordering Algorithm	104
	Canonical Composition Algorithm	105
	Definition of Normalization Forms	107
<b>3.12</b>	<b>Conjoining Jamo Behavior</b>	107
	Definitions	107
	Hangul Syllable Decomposition	109
	Hangul Syllable Composition	111
	Hangul Syllable Name Generation	112
	Sample Code for Hangul Algorithms	112
<b>3.13</b>	<b>Default Case Algorithms</b>	115
	Definitions	115
	Default Case Conversion	117
	Default Case Folding	117
	Default Case Detection	118
	Default Caseless Matching	119

<b>4</b>	<b>Character Properties</b> .....	121
4.1	<b>Unicode Character Database</b> .....	122
4.2	<b>Case</b> .....	124
	Definitions of Case and Casing .....	124
	Case Mapping .....	126
4.3	<b>Combining Classes</b> .....	126
	Reordrant, Split, and Subjoined Combining Marks .....	127
4.4	<b>Directionality</b> .....	130
4.5	<b>General Category</b> .....	130
4.6	<b>Numeric Value</b> .....	133
	Ideographic Numeric Values .....	133
4.7	<b>Bidi Mirrored</b> .....	135
4.8	<b>Name</b> .....	135
	Unicode Name Property .....	137
	Code Point Labels .....	138
	Use of Character Names in APIs and User Interfaces .....	138
4.9	<b>Unicode 1.0 Names</b> .....	139
4.10	<b>Letters, Alphabetic, and Ideographic</b> .....	139
4.11	<b>Properties Related to Text Boundaries</b> .....	140
4.12	<b>Characters with Unusual Properties</b> .....	140
<b>5</b>	<b>Implementation Guidelines</b> .....	145
5.1	<b>Data Structures for Character Conversion</b> .....	145
	Issues .....	145
	Multistage Tables .....	146
5.2	<b>Programming Languages and Data Types</b> .....	147
	Unicode Data Types for C .....	147
5.3	<b>Unknown and Missing Characters</b> .....	148
5.4	<b>Handling Surrogate Pairs in UTF-16</b> .....	149
5.5	<b>Handling Numbers</b> .....	151
5.6	<b>Normalization</b> .....	152
5.7	<b>Compression</b> .....	153
5.8	<b>Newline Guidelines</b> .....	153
	Definitions .....	153
	Line Separator and Paragraph Separator .....	154
	Recommendations .....	155
5.9	<b>Regular Expressions</b> .....	156
5.10	<b>Language Information in Plain Text</b> .....	157
	Requirements for Language Tagging .....	157
	Language Tags and Han Unification .....	157
5.11	<b>Editing and Selection</b> .....	158
	Consistent Text Elements .....	158
5.12	<b>Strategies for Handling Nonspacing Marks</b> .....	159
	Keyboard Input .....	160
	Truncation .....	161
5.13	<b>Rendering Nonspacing Marks</b> .....	162
	Canonical Equivalence .....	164
	Positioning Methods .....	165
5.14	<b>Locating Text Element Boundaries</b> .....	167
5.15	<b>Identifiers</b> .....	167

5.16	<b>Sorting and Searching</b> .....	167
	Culturally Expected Sorting and Searching .....	168
	Language-Insensitive Sorting .....	168
	Searching .....	168
	Sublinear Searching .....	169
5.17	<b>Binary Order</b> .....	170
	UTF-8 in UTF-16 Order .....	170
	UTF-16 in UTF-8 Order .....	171
5.18	<b>Case Mappings</b> .....	172
	Titlecasing .....	172
	Complications for Case Mapping .....	172
	Reversibility .....	174
	Caseless Matching .....	175
	Normalization and Casing .....	177
5.19	<b>Mapping Compatibility Variants</b> .....	177
5.20	<b>Unicode Security</b> .....	179
5.21	<b>Ignoring Characters in Processing</b> .....	180
	Characters Ignored in Text Segmentation .....	181
	Characters Ignored in Line Breaking .....	181
	Characters Ignored in Cursive Joining .....	181
	Characters Ignored in Identifiers .....	182
	Characters Ignored in Searching and Sorting .....	182
	Characters Ignored for Display .....	183
5.22	<b>Best Practice for U+FFFD Substitution</b> .....	185
6	<b>Writing Systems and Punctuation</b> .....	187
6.1	<b>Writing Systems</b> .....	188
6.2	<b>General Punctuation</b> .....	191
	Blocks Devoted to Punctuation .....	193
	Format Control Characters .....	193
	Space Characters .....	194
	Dashes and Hyphens .....	195
	Paired Punctuation .....	197
	Language-Based Usage of Quotation Marks .....	197
	Apostrophes .....	200
	Other Punctuation .....	200
	Archaic Punctuation and Editorial Marks .....	204
	Indic Punctuation .....	206
	CJK Punctuation .....	207
	Unknown or Unavailable Ideographs .....	208
	CJK Compatibility Forms .....	209
7	<b>European Alphabetic Scripts</b> .....	211
7.1	<b>Latin</b> .....	212
	Letters of Basic Latin: U+0041–U+007A .....	215
	Letters of the Latin-1 Supplement: U+00C0–U+00FF .....	215
	Latin Extended-A: U+0100–U+017F .....	216
	Latin Extended-B: U+0180–U+024F .....	216
	IPA Extensions: U+0250–U+02AF .....	218
	Phonetic Extensions: U+1D00–U+1DBF .....	219
	Latin Extended Additional: U+1E00–U+1EFF .....	220
	Latin Extended-C: U+2C60–U+2C7F .....	220
	Latin Extended-D: U+A720–U+A7FF .....	221
	Latin Ligatures: U+FB00–U+FB06 .....	222

<b>7.2</b>	<b>Greek</b> .....	222
	Greek: U+0370–U+03FF .....	222
	Greek Extended: U+1F00–U+1FFF .....	225
	Ancient Greek Numbers: U+10140–U+1018F .....	226
<b>7.3</b>	<b>Coptic</b> .....	227
<b>7.4</b>	<b>Cyrillic</b> .....	229
	Cyrillic: U+0400–U+04FF .....	230
	Cyrillic Supplement: U+0500–U+052F .....	230
	Cyrillic Extended-A: U+2DE0–U+2DFF .....	231
	Cyrillic Extended-B: U+A640–U+A69F .....	231
<b>7.5</b>	<b>Glagolitic</b> .....	231
<b>7.6</b>	<b>Armenian</b> .....	232
<b>7.7</b>	<b>Georgian</b> .....	233
<b>7.8</b>	<b>Modifier Letters</b> .....	235
	Spacing Modifier Letters: U+02B0–U+02FF .....	236
	Modifier Tone Letters: U+A700–U+A71F .....	237
<b>7.9</b>	<b>Combining Marks</b> .....	238
	Combining Diacritical Marks: U+0300–U+036F .....	242
	Combining Diacritical Marks Supplement: U+1DC0–U+1DFF .....	242
	Combining Marks for Symbols: U+20D0–U+20FF .....	242
	Combining Half Marks: U+FE20–U+FE2F .....	243
	Combining Marks in Other Blocks .....	243
<b>8</b>	<b>Middle Eastern Scripts</b> .....	245
<b>8.1</b>	<b>Hebrew</b> .....	246
	Hebrew: U+0590–U+05FF .....	246
	Alphabetic Presentation Forms: U+FB1D–U+FB4F .....	250
<b>8.2</b>	<b>Arabic</b> .....	250
	Arabic: U+0600–U+06FF .....	250
	Arabic Cursive Joining .....	256
	Arabic Ligatures .....	258
	Arabic Joining Groups .....	259
	Arabic Supplement: U+0750–U+077F .....	265
	Arabic Extended-A: U+08A0–U+08FF .....	265
	Arabic Presentation Forms-A: U+FB50–U+FDFF .....	265
	Arabic Presentation Forms-B: U+FE70–U+FEFF .....	266
<b>8.3</b>	<b>Syriac</b> .....	266
	Syriac: U+0700–U+074F .....	266
	Syriac Shaping .....	269
<b>8.4</b>	<b>Samaritan</b> .....	273
<b>8.5</b>	<b>Thaana</b> .....	274
<b>9</b>	<b>South Asian Scripts-I</b> .....	277
<b>9.1</b>	<b>Devanagari</b> .....	278
	Devanagari: U+0900–U+097F .....	278
	Principles of the Devanagari Script .....	279
	Rendering Devanagari .....	284
	Devanagari Digits, Punctuation, and Symbols .....	291
	Extensions in the Main Devanagari Block .....	292
	Devanagari Extended: U+A8E0–U+A8FF .....	293
	Vedic Extensions: U+1CD0–U+1CFF .....	294
<b>9.2</b>	<b>Bengali (Bangla)</b> .....	295
<b>9.3</b>	<b>Gurmukhi</b> .....	300

9.4	<b>Gujarati</b> .....	303
9.5	<b>Oriya</b> .....	304
9.6	<b>Tamil</b> .....	306
	Tamil: U+0B80–U+0BFF .....	306
	Tamil Vowels .....	307
	Tamil Ligatures .....	308
	Tamil Named Character Sequences .....	311
9.7	<b>Telugu</b> .....	313
9.8	<b>Kannada</b> .....	315
	Kannada: U+0C80–U+0CFF .....	315
	Principles of the Kannada Script .....	315
	Rendering Kannada .....	317
9.9	<b>Malayalam</b> .....	317
10	<b>South Asian Scripts-II</b> .....	323
10.1	<b>Sinhala</b> .....	324
10.2	<b>Tibetan</b> .....	325
10.3	<b>Lepcha</b> .....	335
10.4	<b>Phags-pa</b> .....	336
10.5	<b>Limbu</b> .....	342
10.6	<b>Syloti Nagri</b> .....	344
10.7	<b>Kaithi</b> .....	345
10.8	<b>Saurashtra</b> .....	347
10.9	<b>Sharada</b> .....	348
10.10	<b>Takri</b> .....	349
10.11	<b>Chakma</b> .....	350
10.12	<b>Meetei Mayek</b> .....	351
10.13	<b>Ol Chiki</b> .....	353
10.14	<b>Sora Sompeng</b> .....	354
10.15	<b>Kharoshthi</b> .....	355
	Kharoshthi: U+10A00–U+10A5F .....	355
	Rendering Kharoshthi .....	356
10.16	<b>Brahmi</b> .....	359
11	<b>Southeast Asian Scripts</b> .....	363
11.1	<b>Thai</b> .....	364
11.2	<b>Lao</b> .....	366
11.3	<b>Myanmar</b> .....	368
	Myanmar: U+1000–U+109F .....	368
	Myanmar Extended-A: U+AA60–U+AA7F .....	371
	Khamti Shan .....	372
	Aiton and Phake .....	373
11.4	<b>Khmer</b> .....	374
	Khmer: U+1780–U+17FF .....	374
	Principles of the Khmer Script .....	374
	Khmer Symbols: U+19E0–U+19FF .....	383
11.5	<b>Tai Le</b> .....	383
11.6	<b>New Tai Lue</b> .....	384
11.7	<b>Tai Tham</b> .....	385
11.8	<b>Tai Viet</b> .....	387



<b>11.9 Kayah Li</b> .....	389
<b>11.10 Cham</b> .....	390
<b>11.11 Philippine Scripts</b> .....	392
Tagalog: U+1700–U+171F .....	392
Hanunóo: U+1720–U+173F .....	392
Buhid: U+1740–U+175F .....	392
Tagbanwa: U+1760–U+177F .....	392
Principles of the Philippine Scripts .....	392
<b>11.12 Buginese</b> .....	393
<b>11.13 Balinese</b> .....	394
<b>11.14 Javanese</b> .....	399
<b>11.15 Rejang</b> .....	401
<b>11.16 Batak</b> .....	402
<b>11.17 Sundanese</b> .....	403
<b>12 East Asian Scripts</b> .....	405
<b>12.1 Han</b> .....	406
CJK Unified Ideographs .....	406
CJK Standards .....	407
Blocks Containing Han Ideographs .....	409
General Characteristics of Han Ideographs .....	411
Principles of Han Unification .....	414
Unification Rules .....	415
Abstract Shape .....	416
Han Ideograph Arrangement .....	418
Radical-Stroke Indices .....	419
Mappings for Han Ideographs .....	419
CJK Unified Ideographs Extension B: U+20000–U+2A6D6 .....	420
CJK Unified Ideographs Extension C: U+2A700–U+2B734 .....	420
CJK Unified Ideographs Extension D: U+2B740–U+2B81D .....	420
CJK Compatibility Ideographs: U+F900–U+FAFF .....	420
CJK Compatibility Supplement: U+2F800–U+2FA1D .....	421
Kanbun: U+3190–U+319F .....	421
Symbols Derived from Han Ideographs .....	421
CJK and KangXi Radicals: U+2E80–U+2FD5 .....	421
CJK Additions from HKSCS and GB 18030 .....	422
CJK Strokes: U+31C0–U+31EF .....	423
<b>12.2 Ideographic Description Characters</b> .....	423
<b>12.3 Bopomofo</b> .....	426
<b>12.4 Hiragana and Katakana</b> .....	428
Hiragana: U+3040–U+309F .....	428
Katakana: U+30A0–U+30FF .....	428
Katakana Phonetic Extensions: U+31F0–U+31FF .....	429
Kana Supplement U+1B000–U+1B0FF .....	429
<b>12.5 Halfwidth and Fullwidth Forms</b> .....	429
<b>12.6 Hangul</b> .....	430
Hangul Jamo: U+1100–U+11FF .....	430
Hangul Jamo Extended-A: U+A960–U+A97F .....	431
Hangul Jamo Extended-B: U+D7B0–U+D7FF .....	431
Hangul Compatibility Jamo: U+3130–U+318F .....	431
Hangul Syllables: U+AC00–U+D7A3 .....	432
<b>12.7 Yi</b> .....	433

<b>13 Additional Modern Scripts</b> .....	437
<b>13.1 Ethiopic</b> .....	438
Ethiopic: U+1200–U+137F .....	438
Ethiopic Extensions .....	440
<b>13.2 Mongolian</b> .....	440
<b>13.3 Osmanya</b> .....	447
<b>13.4 Tifinagh</b> .....	448
<b>13.5 N’Ko</b> .....	450
<b>13.6 Vai</b> .....	454
<b>13.7 Bamum</b> .....	455
Bamum: U+A6A0–U+A6FF .....	455
Bamum Supplement: U+16800–U+16A3F .....	456
<b>13.8 Cherokee</b> .....	456
<b>13.9 Canadian Aboriginal Syllabics</b> .....	457
Canadian Aboriginal Syllabics: U+1400–U+167F .....	457
Canadian Aboriginal Syllabics Extended: U+18B0–U+18FF .....	458
<b>13.10 Deseret</b> .....	459
<b>13.11 Shavian</b> .....	461
<b>13.12 Lisu</b> .....	461
<b>13.13 Miao</b> .....	463
<b>14 Additional Ancient and Historic Scripts</b> .....	465
<b>14.1 Ogham</b> .....	466
<b>14.2 Old Italic</b> .....	467
<b>14.3 Runic</b> .....	469
<b>14.4 Gothic</b> .....	471
<b>14.5 Old Turkic</b> .....	472
<b>14.6 Linear B</b> .....	473
Linear B Syllabary: U+10000–U+1007F .....	473
Linear B Ideograms: U+10080–U+100FF .....	473
Aegean Numbers: U+10100–U+1013F .....	473
<b>14.7 Cypriot Syllabary</b> .....	474
<b>14.8 Ancient Anatolian Alphabets</b> .....	474
Lycian: U+10280–U+1029F .....	474
Carian: U+102A0–U+102DF .....	474
Lydian: U+10920–U+1093F .....	474
<b>14.9 Old South Arabian</b> .....	475
<b>14.10 Phoenician</b> .....	477
<b>14.11 Imperial Aramaic</b> .....	478
<b>14.12 Mandaic</b> .....	479
<b>14.13 Inscriptional Parthian and Inscriptional Pahlavi</b> .....	481
<b>14.14 Avestan</b> .....	482
<b>14.15 Ugaritic</b> .....	483
<b>14.16 Old Persian</b> .....	484
<b>14.17 Sumero-Akkadian</b> .....	485
Cuneiform: U+12000–U+123FF .....	485
Cuneiform Numbers and Punctuation: U+12400–U+1247F .....	487
<b>14.18 Egyptian Hieroglyphs</b> .....	487
<b>14.19 Meroitic Hieroglyphs and Meroitic Cursive</b> .....	491

<b>15 Symbols</b> .....	493
<b>15.1 Currency Symbols</b> .....	494
<b>15.2 Letterlike Symbols</b> .....	496
Letterlike Symbols: U+2100–U+214F .....	496
Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF .....	498
Mathematical Alphabets .....	498
Fonts Used for Mathematical Alphabets .....	500
Arabic Mathematical Alphabetic Symbols: U+1EE00–U+1EEFF .....	501
<b>15.3 Numerals</b> .....	502
Decimal Digits .....	502
Other Digits .....	504
Non-Decimal Radix Systems .....	505
Acrophonic Systems and Other Letter-based Numbers .....	506
Rumi Numeral Forms: U+10E60–U+10E7E .....	507
CJK Numerals .....	507
Fractions .....	509
Common Indic Number Forms: U+A830–U+A83F .....	509
<b>15.4 Superscript and Subscript Symbols</b> .....	510
Superscripts and Subscripts: U+2070–U+209F .....	510
<b>15.5 Mathematical Symbols</b> .....	511
Mathematical Operators: U+2200–U+22FF .....	512
Supplements to Mathematical Symbols and Arrows .....	513
Supplemental Mathematical Operators: U+2A00–U+2AFF .....	514
Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF .....	514
Miscellaneous Mathematical Symbols-B: U+2980–U+29FF .....	514
Miscellaneous Symbols and Arrows: U+2B00–U+2B7F .....	515
Arrows: U+2190–U+21FF .....	515
Supplemental Arrows .....	515
Standardized Variants of Mathematical Symbols .....	516
<b>15.6 Invisible Mathematical Operators</b> .....	516
<b>15.7 Technical Symbols</b> .....	517
Control Pictures: U+2400–U+243F .....	517
Miscellaneous Technical: U+2300–U+23FF .....	517
Optical Character Recognition: U+2440–U+245F .....	520
<b>15.8 Geometrical Symbols</b> .....	520
Box Drawing and Block Elements .....	521
Geometric Shapes: U+25A0–U+25FF .....	521
<b>15.9 Miscellaneous Symbols</b> .....	522
Miscellaneous Symbols: U+2600–U+26FF .....	523
Miscellaneous Symbols and Pictographs: U+1F300–U+1F5FF .....	523
Emoticons: U+1F600–U+1F64F .....	525
Transport and Map Symbols: U+1F680–U+1F6FF .....	526
Dingbats: U+2700–U+27BF .....	526
Alchemical Symbols: U+1F700–U+1F77F .....	527
Mahjong Tiles: U+1F000–U+1F02F .....	527
Domino Tiles: U+1F030–U+1F09F .....	528
Playing Cards: U+1F0A0–U+1F0FF .....	528
Yijing Hexagram Symbols: U+4DC0–U+4DFF .....	529
Tai Xuan Jing Symbols: U+1D300–U+1D356 .....	529
Ancient Symbols: U+10190–U+101CF .....	530
Phaistos Disc Symbols: U+101D0–U+101FF .....	530

<b>15.10 Enclosed and Square</b> .....	531
Enclosed Alphanumerics: U+2460–U+24FF .....	533
Enclosed CJK Letters and Months: U+3200–U+32FF .....	533
CJK Compatibility: U+3300–U+33FF .....	533
Enclosed Alphanumeric Supplement: U+1F100–U+1F1FF .....	534
Enclosed Ideographic Supplement: U+1F200–U+1F2FF .....	534
<b>15.11 Braille</b> .....	534
<b>15.12 Western Musical Symbols</b> .....	536
<b>15.13 Byzantine Musical Symbols</b> .....	540
<b>15.14 Ancient Greek Musical Notation</b> .....	540
<b>16 Special Areas and Format Characters</b> .....	543
<b>16.1 Control Codes</b> .....	544
Representing Control Sequences .....	544
Specification of Control Code Semantics .....	545
<b>16.2 Layout Controls</b> .....	545
Line and Word Breaking .....	546
Cursive Connection and Ligatures .....	548
Combining Grapheme Joiner .....	551
Bidirectional Ordering Controls .....	553
Stateful Format Controls .....	553
<b>16.3 Deprecated Format Characters</b> .....	554
<b>16.4 Variation Selectors</b> .....	556
<b>16.5 Private-Use Characters</b> .....	557
Private Use Area: U+E000–U+F8FF .....	558
Supplementary Private Use Areas .....	559
<b>16.6 Surrogates Area</b> .....	559
<b>16.7 Noncharacters</b> .....	560
<b>16.8 Specials</b> .....	561
Byte Order Mark (BOM): U+FEFF .....	561
Specials: U+FFF0–U+FFF8 .....	563
Annotation Characters: U+FFF9–U+FFFB .....	563
Replacement Characters: U+FFFC–U+FFFD .....	564
<b>16.9 Deprecated Tag Characters</b> .....	565
Deprecated Tag Characters: U+E0000–U+E007F .....	565
Syntax for Embedding Tags .....	565
Working with Language Tags .....	567
Unicode Conformance Issues .....	568
Formal Tag Syntax .....	569
<b>17 About the Code Charts</b> .....	571
<b>17.1 Character Names List</b> .....	571
Images in the Code Charts and Character Lists .....	572
Special Characters and Code Points .....	573
Character Names .....	574
Informative Aliases .....	574
Normative Aliases .....	575
Cross References .....	575
Information About Languages .....	575
Case Mappings .....	576
Decompositions .....	576
Standardized Variation Sequences .....	577
Subheads .....	577

<b>17.2</b>	<b>CJK Unified and Compatibility Ideographs</b> .....	577
	CJK Unified Ideographs .....	578
	Compatibility Ideographs .....	579
<b>17.3</b>	<b>Hangul Syllables</b> .....	579
<b>A</b>	<b>Notational Conventions</b> .....	581
	Code Points .....	581
	Character Names .....	581
	Character Blocks .....	581
	Sequences .....	582
	Rendering .....	582
	Properties and Property Values .....	582
	Miscellaneous .....	582
	Extended BNF .....	583
	Operators .....	584
<b>B</b>	<b>Unicode Publications and Resources</b> .....	585
<b>B.1</b>	<b>The Unicode Consortium</b> .....	585
	The Unicode Technical Committee .....	585
	Other Activities .....	586
<b>B.2</b>	<b>Unicode Publications</b> .....	586
<b>B.3</b>	<b>Unicode Technical Standards</b> .....	586
	UTS #6: A Standard Compression Scheme for Unicode .....	586
	UTS #10: Unicode Collation Algorithm .....	587
	UTS #18: Unicode Regular Expressions .....	587
	UTS #22: Character Mapping Markup Language (CharMapML) .....	587
	UTS #35: Unicode Locale Data Markup Language (LDML) .....	587
	UTS #37: Unicode Ideographic Variation Database .....	587
	UTS #39: Unicode Security Mechanisms .....	587
<b>B.4</b>	<b>Unicode Technical Reports</b> .....	587
	UTR #16: UTF-EBCDIC .....	587
	UTR #17: Unicode Character Encoding Model .....	588
	UTR #20: Unicode in XML and Other Markup Languages .....	588
	UTR #23: The Unicode Character Property Model .....	588
	UTR #25: Unicode Support for Mathematics .....	588
	UTR #26: Compatibility Encoding Scheme for UTF-16: 8-Bit (CESU-8) ..	588
	UTR #33: Unicode Conformance Model .....	588
	UTR #36: Unicode Security Considerations .....	588
<b>B.5</b>	<b>Unicode Technical Notes</b> .....	589
<b>B.6</b>	<b>Other Unicode Online Resources</b> .....	589
	Unicode Online Resources .....	589
	How to Contact the Unicode Consortium .....	591
<b>C</b>	<b>Relationship to ISO/IEC 10646</b> .....	593
<b>C.1</b>	<b>History</b> .....	593
<b>C.2</b>	<b>Encoding Forms in ISO/IEC 10646</b> .....	597
	Zero Extending .....	597
<b>C.3</b>	<b>UTF-8 and UTF-16</b> .....	598
	UTF-8 .....	598
	UTF-16 .....	598
<b>C.4</b>	<b>Synchronization of the Standards</b> .....	598
<b>C.5</b>	<b>Identification of Features for the Unicode Standard</b> .....	598
<b>C.6</b>	<b>Character Names</b> .....	599
<b>C.7</b>	<b>Character Functional Specifications</b> .....	599

<b>D</b>	<b>Changes from Previous Versions</b> .....	601
<b>D.1</b>	<b>Versions of the Unicode Standard</b> .....	601
<b>D.2</b>	<b>Clause and Definition Updates</b> .....	603
<b>E</b>	<b>Han Unification History</b> .....	605
<b>E.1</b>	<b>Development of the URO</b> .....	605
<b>E.2</b>	<b>Ideographic Rapporteur Group</b> .....	606
<b>F</b>	<b>Documentation of CJK Strokes</b> .....	609
<b>R</b>	<b>References</b> .....	615
<b>R.1</b>	<b>Source Standards and Specifications</b> .....	615
<b>R.2</b>	<b>Source Dictionaries for Han Unification</b> .....	622
<b>R.3</b>	<b>Other Sources for the Unicode Standard</b> .....	622
<b>R.4</b>	<b>Selected Resources: Technical</b> .....	638
<b>R.5</b>	<b>Selected Resources: Other</b> .....	640
<b>I</b>	<b>General Index</b> .....	645

