

### ***Electronic Edition***

This file is part of the electronic edition of *The Unicode Standard, Version 5.0*, provided for online access, content searching, and accessibility. It may not be printed. Bookmarks linking to specific chapters or sections of the whole Unicode Standard are available at

<http://www.unicode.org/versions/Unicode5.0.0/bookmarks.html>

### ***Purchasing the Book***

For convenient access to the full text of the standard as a useful reference book, we recommend purchasing the printed version. The book is available from the Unicode Consortium, the publisher, and booksellers. Purchase of the standard in book format contributes to the ongoing work of the Unicode Consortium. Details about the book publication and ordering information may be found at

<http://www.unicode.org/book/aboutbook.html>

### ***Joining Unicode***

You or your organization may benefit by joining the Unicode Consortium: for more information, see [Joining the Unicode Consortium](http://www.unicode.org/consortium/join.html) at

<http://www.unicode.org/consortium/join.html>

This PDF file is an excerpt from *The Unicode Standard, Version 5.0*, issued by the Unicode Consortium and published by Addison-Wesley. The material has been modified slightly for this electronic edition, however, the PDF files have not been modified to reflect the corrections found on the Updates and Errata page (<http://www.unicode.org/errata/>). For information on more recent versions of the standard, see <http://www.unicode.org/versions/enumeratedversions.html>.

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.

The Unicode® Consortium is a registered trademark, and Unicode™ is a trademark of Unicode, Inc. The Unicode logo is a trademark of Unicode, Inc., and may be registered in some jurisdictions.

The authors and publisher have taken care in the preparation of this book, 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. *Dai Kan-Wa Jiten*, used as the source of reference Kanji codes, was written by Tetsuji Morohashi and published by Taishukan Shoten.

Cover and CD-ROM label design: Steve Mehallo, [www.mehallo.com](http://www.mehallo.com)

The publisher offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales, which may include electronic versions and/or custom covers and content particular to your business, training goals, marketing focus, and branding interests. For more information, please contact U.S. Corporate and Government Sales, (800) 382-3419, [corpsales@pearsoned.com](mailto:corpsales@pearsoned.com). For sales outside the United States please contact International Sales, [international@pearsoned.com](mailto:international@pearsoned.com)

Visit us on the Web: [www.awprofessional.com](http://www.awprofessional.com)

*Library of Congress Cataloging-in-Publication Data*

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 5.0.  
p. cm.

Includes bibliographical references and index.

ISBN 0-321-48091-0 (hardcover : alk. paper)

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

II. Unicode Consortium.

QA268.U545 2007

005.7'22—dc22

2006023526

Copyright © 1991–2007 Unicode, Inc.

All rights reserved. Printed in the United States of America. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. For information regarding permissions, write to Pearson Education, Inc., Rights and Contracts Department, 75 Arlington Street, Suite 300, Boston, MA 02116. Fax: (617) 848-7047

ISBN 0-321-48091-0

Text printed in the United States on recycled paper at Courier in Westford, Massachusetts.

First printing, October 2006

# Contents

<b>List of Figures</b> .....	<i>xxiii</i>
<b>List of Tables</b> .....	<i>xxvii</i>
<b>Foreword by Mark Davis</b> .....	<i>xxxi</i>
<b>Preface</b> .....	<i>xxxiii</i>
Why Buy This Book .....	<i>xxxiii</i>
Why Upgrade to Version 5.0 .....	<i>xxxiv</i>
Organization of This Book .....	<i>xxxiv</i>
Unicode Standard Annexes .....	<i>xxxvi</i>
The Unicode Character Database .....	<i>xxxvii</i>
Unicode Technical Standards and Unicode Technical Reports .....	<i>xxxvii</i>
On the CD-ROM .....	<i>xxxvii</i>
Updates and Errata .....	<i>xxxviii</i>
<b>Acknowledgments</b> .....	<i>xxxix</i>
Unicode Consortium Members .....	<i>xlvii</i>
Unicode Consortium Liaison Members .....	<i>xliv</i>
Unicode Consortium Board of Directors .....	<i>l</i>
<b>1 Introduction</b> .....	<i>1</i>
<b>1.1 Coverage</b> .....	<i>2</i>
Standards Coverage .....	<i>3</i>
New Characters .....	<i>3</i>
<b>1.2 Design Goals</b> .....	<i>4</i>
<b>1.3 Text Handling</b> .....	<i>5</i>
Text Elements .....	<i>6</i>
<b>2 General Structure</b> .....	<i>9</i>
<b>2.1 Architectural Context</b> .....	<i>9</i>
Basic Text Processes .....	<i>10</i>
Text Elements, Characters, and Text Processes .....	<i>10</i>
Text Processes and Encoding .....	<i>11</i>
<b>2.2 Unicode Design Principles</b> .....	<i>13</i>
Universality .....	<i>14</i>
Efficiency .....	<i>14</i>
Characters, Not Glyphs .....	<i>14</i>
Semantics .....	<i>16</i>
Plain Text .....	<i>18</i>
Logical Order .....	<i>19</i>
Unification .....	<i>21</i>
Dynamic Composition .....	<i>22</i>

	Stability	22
	Convertibility	23
<b>2.3</b>	<b>Compatibility Characters</b>	23
	Compatibility Variants	23
	Compatibility Decomposable Characters	24
	Mapping Compatibility Characters	25
<b>2.4</b>	<b>Code Points and Characters</b>	25
	Types of Code Points	26
<b>2.5</b>	<b>Encoding Forms</b>	28
	UTF-32	31
	UTF-16	31
	UTF-8	32
	Comparison of the Advantages of UTF-32, UTF-16, and UTF-8	33
<b>2.6</b>	<b>Encoding Schemes</b>	35
<b>2.7</b>	<b>Unicode Strings</b>	37
<b>2.8</b>	<b>Unicode Allocation</b>	38
	Planes	39
	Allocation Areas and Character Blocks	39
	Assignment of Code Points	41
<b>2.9</b>	<b>Details of Allocation</b>	41
	Plane 0 (BMP)	43
	Plane 1	43
	Plane 2	46
	Other Planes	46
<b>2.10</b>	<b>Writing Direction</b>	46
<b>2.11</b>	<b>Combining Characters</b>	48
	Sequence of Base Characters and Diacritics	49
	Multiple Combining Characters	50
	Ligated Multiple Base Characters	52
	Exhibiting Nonspacing Marks in Isolation	53
	“Characters” and Grapheme Clusters	53
<b>2.12</b>	<b>Equivalent Sequences and Normalization</b>	54
<b>2.13</b>	<b>Special Characters and Noncharacters</b>	57
	Special Noncharacter Code Points	57
	Byte Order Mark (BOM)	57
	Layout and Format Control Characters	58
	The Replacement Character	58
	Control Codes	58
<b>2.14</b>	<b>Conforming to the Unicode Standard</b>	59
	Characteristics of Conformant Implementations	59
	Unacceptable Behavior	60
	Acceptable Behavior	60
	Supported Subsets	61

<b>3</b>	<b>Conformance</b> .....	65
<b>3.1</b>	<b>Versions of the Unicode Standard</b> .....	65
	Stability .....	66
	Version Numbering .....	66
	Errata and Corrigenda .....	67
	References to the Unicode Standard .....	68
	Precision in Version Citation .....	68
	References to Unicode Character Properties .....	69
	References to Unicode Algorithms .....	69
<b>3.2</b>	<b>Conformance Requirements</b> .....	70
	Code Points Unassigned to Abstract Characters .....	70
	Interpretation .....	71
	Modification .....	72
	Character Encoding Forms .....	73
	Character Encoding Schemes .....	74
	Bidirectional Text .....	74
	Normalization Forms .....	74
	Normative References .....	75
	Unicode Algorithms .....	75
	Default Casing Algorithms .....	75
	Unicode Standard Annexes .....	75
<b>3.3</b>	<b>Semantics</b> .....	76
	Definitions .....	76
	Character Identity and Semantics .....	76
<b>3.4</b>	<b>Characters and Encoding</b> .....	78
<b>3.5</b>	<b>Properties</b> .....	81
	Types of Properties .....	82
	Property Values .....	83
	Classification of Properties by Their Values .....	84
	Normative and Informative Properties .....	85
	Context Dependence .....	87
	Stability of Properties .....	88
	Simple and Derived Properties .....	89
	Property Aliases .....	90
	Private Use .....	91
<b>3.6</b>	<b>Combination</b> .....	91
<b>3.7</b>	<b>Decomposition</b> .....	95
	Compatibility Decomposition .....	95
	Canonical Decomposition .....	96
<b>3.8</b>	<b>Surrogates</b> .....	97
<b>3.9</b>	<b>Unicode Encoding Forms</b> .....	98
	UTF-32 .....	102
	UTF-16 .....	102

	UTF-8 .....	103
	Encoding Form Conversion .....	104
<b>3.10</b>	<b>Unicode Encoding Schemes</b> .....	<b>105</b>
<b>3.11</b>	<b>Canonical Ordering Behavior</b> .....	<b>109</b>
	Application of Combining Marks .....	109
	Combining Classes .....	114
	Canonical Ordering .....	115
<b>3.12</b>	<b>Conjoining Jamo Behavior</b> .....	<b>117</b>
	Definitions .....	117
	Hangul Syllable Boundary Determination .....	119
	Standard Korean Syllables .....	120
	Hangul Syllable Composition .....	121
	Hangul Syllable Decomposition .....	122
	Hangul Syllable Name Generation .....	123
<b>3.13</b>	<b>Default Case Algorithms</b> .....	<b>123</b>
	Definitions .....	123
	Default Case Conversion .....	125
	Default Case Detection .....	125
	Default Caseless Matching .....	126
<b>4</b>	<b>Character Properties</b> .....	<b>129</b>
<b>4.1</b>	<b>Unicode Character Database</b> .....	<b>130</b>
<b>4.2</b>	<b>Case—Normative</b> .....	<b>132</b>
	Case Mapping .....	133
<b>4.3</b>	<b>Combining Classes—Normative</b> .....	<b>133</b>
	Reordrant, Split, and Subjoined Combining Marks .....	134
<b>4.4</b>	<b>Directionality—Normative</b> .....	<b>138</b>
<b>4.5</b>	<b>General Category—Normative</b> .....	<b>138</b>
<b>4.6</b>	<b>Numeric Value—Normative</b> .....	<b>139</b>
	Ideographic Numeric Values .....	140
<b>4.7</b>	<b>Bidi Mirrored—Normative</b> .....	<b>141</b>
<b>4.8</b>	<b>Name—Normative</b> .....	<b>142</b>
<b>4.9</b>	<b>Unicode 1.0 Names</b> .....	<b>144</b>
<b>4.10</b>	<b>Letters, Alphabetic, and Ideographic</b> .....	<b>144</b>
<b>4.11</b>	<b>Properties Related to Text Boundaries</b> .....	<b>145</b>
<b>4.12</b>	<b>Characters with Unusual Properties</b> .....	<b>145</b>
<b>5</b>	<b>Implementation Guidelines</b> .....	<b>151</b>
<b>5.1</b>	<b>Transcoding to Other Standards</b> .....	<b>151</b>
	Issues .....	151
	Multistage Tables .....	152
<b>5.2</b>	<b>Programming Languages and Data Types</b> .....	<b>153</b>
	Unicode Data Types for C .....	154

<b>5.3</b>	<b>Unknown and Missing Characters</b> .....	155
	Reserved and Private-Use Character Codes .....	155
	Interpretable but Unrenderable Characters .....	156
	Default Property Values .....	156
	Default Ignorable Code Points .....	156
	Interacting with Downlevel Systems .....	156
<b>5.4</b>	<b>Handling Surrogate Pairs in UTF-16</b> .....	157
<b>5.5</b>	<b>Handling Numbers</b> .....	158
<b>5.6</b>	<b>Normalization</b> .....	160
<b>5.7</b>	<b>Compression</b> .....	161
<b>5.8</b>	<b>Newline Guidelines</b> .....	161
	Definitions .....	162
	Line Separator and Paragraph Separator .....	163
	Recommendations .....	164
<b>5.9</b>	<b>Regular Expressions</b> .....	166
<b>5.10</b>	<b>Language Information in Plain Text</b> .....	166
	Requirements for Language Tagging .....	166
	Language Tags and Han Unification .....	166
<b>5.11</b>	<b>Editing and Selection</b> .....	167
	Consistent Text Elements .....	167
<b>5.12</b>	<b>Strategies for Handling Nonspacing Marks</b> .....	169
	Keyboard Input .....	170
	Truncation .....	171
<b>5.13</b>	<b>Rendering Nonspacing Marks</b> .....	172
	Canonical Equivalence .....	175
	Positioning Methods .....	176
<b>5.14</b>	<b>Locating Text Element Boundaries</b> .....	178
<b>5.15</b>	<b>Identifiers</b> .....	179
<b>5.16</b>	<b>Sorting and Searching</b> .....	179
	Culturally Expected Sorting and Searching .....	179
	Language-Insensitive Sorting .....	180
	Searching .....	180
	Sublinear Searching .....	181
<b>5.17</b>	<b>Binary Order</b> .....	181
	UTF-8 in UTF-16 Order .....	182
	UTF-16 in UTF-8 Order .....	183
<b>5.18</b>	<b>Case Mappings</b> .....	184
	Titlecasing .....	184
	Complications for Case Mapping .....	185
	Reversibility .....	187
	Caseless Matching .....	187
	Normalization .....	189

5.19	<b>Unicode Security</b> .....	190
5.20	<b>Default Ignorable Code Points</b> .....	192
6	<b>Writing Systems and Punctuation</b> .....	197
6.1	<b>Writing Systems</b> .....	198
6.2	<b>General Punctuation</b> .....	202
	Blocks Devoted to Punctuation.....	204
	Format Control Characters.....	204
	Space Characters.....	205
	Dashes and Hyphens.....	206
	Paired Punctuation.....	208
	Language-Based Usage of Quotation Marks.....	209
	Apostrophes.....	211
	Other Punctuation.....	212
	Archaic Punctuation and Editorial Marks.....	216
	Indic Punctuation.....	218
	CJK Punctuation.....	219
	Unknown or Unavailable Ideographs.....	220
	CJK Compatibility Forms.....	220
7	<b>European Alphabetic Scripts</b> .....	225
7.1	<b>Latin</b> .....	226
	Letters of Basic Latin: U+0041–U+007A.....	230
	Letters of the Latin-1 Supplement: U+00C0–U+00FF.....	230
	Latin Extended-A: U+0100–U+017F.....	230
	Latin Extended-B: U+0180–U+024F.....	231
	IPA Extensions: U+0250–U+02AF.....	233
	Phonetic Extensions: U+1D00–U+1DBF.....	234
	Latin Extended Additional: U+1E00–U+1EFF.....	235
	Latin Extended-C: U+2C60–U+2C7F.....	236
	Latin Extended-D: U+A720–U+A7FF.....	236
	Latin Ligatures: U+FB00–U+FB06.....	236
7.2	<b>Greek</b> .....	237
	Greek: U+0370–U+03FF.....	237
	Greek Extended: U+1F00–U+1FFF.....	241
	Ancient Greek Numbers: U+10140–U+1018F.....	242
7.3	<b>Coptic</b> .....	243
7.4	<b>Cyrillic</b> .....	245
	Cyrillic: U+0400–U+04FF.....	245
	Cyrillic Supplement: U+0500–U+052F.....	246
7.5	<b>Glagolitic</b> .....	246
7.6	<b>Armenian</b> .....	247
7.7	<b>Georgian</b> .....	249



<b>7.8</b>	<b>Modifier Letters</b> .....	250
	Spacing Modifier Letters: U+02B0–U+02FF .....	250
	Modifier Tone Letters: U+A700–U+A71F .....	252
<b>7.9</b>	<b>Combining Marks</b> .....	252
	Combining Diacritical Marks: U+0300–U+036F .....	257
	Combining Diacritical Marks Supplement: U+1DC0–U+1DFF .....	257
	Combining Marks for Symbols: U+20D0–U+20FF .....	257
	Combining Half Marks: U+FE20–U+FE2F .....	258
	Combining Marks in Other Blocks .....	259
<b>8</b>	<b>Middle Eastern Scripts</b> .....	263
<b>8.1</b>	<b>Hebrew</b> .....	264
	Hebrew: U+0590–U+05FF .....	264
	Alphabetic Presentation Forms: U+FB1D–U+FB4F .....	269
<b>8.2</b>	<b>Arabic</b> .....	269
	Arabic: U+0600–U+06FF .....	269
	Arabic Cursive Joining .....	275
	Arabic Ligatures .....	278
	Arabic Supplement: U+0750–U+077F .....	282
	Arabic Presentation Forms-A: U+FB50–U+FDFF .....	282
	Arabic Presentation Forms-B: U+FE70–U+FEFF .....	282
<b>8.3</b>	<b>Syriac</b> .....	283
	Syriac: U+0700–U+074F .....	283
	Syriac Shaping .....	288
	Syriac Cursive Joining .....	288
	Syriac Ligatures .....	290
<b>8.4</b>	<b>Thaana</b> .....	291
<b>9</b>	<b>South Asian Scripts-I</b> .....	295
<b>9.1</b>	<b>Devanagari</b> .....	296
	Devanagari: U+0900–U+097F .....	296
	Principles of the Devanagari Script .....	297
	Rendering Devanagari .....	303
<b>9.2</b>	<b>Bengali</b> .....	312
<b>9.3</b>	<b>Gurmukhi</b> .....	317
<b>9.4</b>	<b>Gujarati</b> .....	321
<b>9.5</b>	<b>Oriya</b> .....	322
<b>9.6</b>	<b>Tamil</b> .....	324
	Tamil: U+0B80–U+0BFF .....	324
	Tamil Vowels .....	325
	Tamil Ligatures .....	327
<b>9.7</b>	<b>Telugu</b> .....	330
<b>9.8</b>	<b>Kannada</b> .....	331
	Kannada: U+0C80–U+0CFF .....	331

	Principles of the Kannada Script .....	332
	Rendering Kannada .....	333
<b>9.9</b>	<b>Malayalam</b> .....	334
<b>10</b>	<b>South Asian Scripts-II</b> .....	341
<b>10.1</b>	<b>Sinhala</b> .....	341
<b>10.2</b>	<b>Tibetan</b> .....	343
<b>10.3</b>	<b>Phags-pa</b> .....	353
<b>10.4</b>	<b>Limbu</b> .....	360
<b>10.5</b>	<b>Syloti Nagri</b> .....	363
<b>10.6</b>	<b>Kharoshthi</b> .....	364
	Kharoshthi: U+10A00–U+10A5F .....	364
	Rendering Kharoshthi .....	366
<b>11</b>	<b>Southeast Asian Scripts</b> .....	373
<b>11.1</b>	<b>Thai</b> .....	373
<b>11.2</b>	<b>Lao</b> .....	376
<b>11.3</b>	<b>Myanmar</b> .....	379
<b>11.4</b>	<b>Khmer</b> .....	382
	Khmer: U+1780–U+17FF .....	382
	Principles of the Khmer Script .....	382
	Khmer Symbols: U+19E0–U+19FF .....	392
<b>11.5</b>	<b>Tai Le</b> .....	393
<b>11.6</b>	<b>New Tai Lue</b> .....	394
<b>11.7</b>	<b>Philippine Scripts</b> .....	395
	Tagalog: U+1700–U+171F .....	395
	Hanunóo: U+1720–U+173F .....	395
	Buhid: U+1740–U+175F .....	395
	Tagbanwa: U+1760–U+177F .....	395
	Principles of the Philippine Scripts .....	396
<b>11.8</b>	<b>Buginese</b> .....	397
<b>11.9</b>	<b>Balinese</b> .....	399
<b>12</b>	<b>East Asian Scripts</b> .....	407
<b>12.1</b>	<b>Han</b> .....	408
	CJK Unified Ideographs .....	408
	CJK Standards .....	409
	Blocks Containing Han Ideographs .....	411
	General Characteristics of Han Ideographs .....	413
	Principles of Han Unification .....	417
	Unification Rules .....	418
	Abstract Shape .....	419
	Han Ideograph Arrangement .....	420
	Mappings for Han Ideographs .....	422

	CJK Unified Ideographs Extension B: U+20000–U+2A6D6	423
	CJK Compatibility Ideographs: U+F900–U+FAFF	424
	CJK Compatibility Supplement: U+2F800–U+2FA1D	425
	Kanbun: U+3190–U+319F	425
	Symbols Derived from Han Ideographs	425
	CJK and KangXi Radicals: U+2E80–U+2FD5	425
	CJK Additions from HKSCS and GB 18030	426
	CJK Strokes: U+31C0–U+31EF	427
<b>12.2</b>	<b>Ideographic Description Characters</b>	427
<b>12.3</b>	<b>Bopomofo</b>	431
<b>12.4</b>	<b>Hiragana and Katakana</b>	433
	Hiragana: U+3040–U+309F	433
	Katakana: U+30A0–U+30FF	433
	Katakana Phonetic Extensions: U+31F0–U+31FF	434
<b>12.5</b>	<b>Halfwidth and Fullwidth Forms</b>	434
<b>12.6</b>	<b>Hangul</b>	435
	Hangul Jamo: U+1100–U+11FF	435
	Hangul Compatibility Jamo: U+3130–U+318F	436
	Hangul Syllables: U+AC00–U+D7A3	437
<b>12.7</b>	<b>Yi</b>	438
<b>13</b>	<b>Additional Modern Scripts</b>	445
<b>13.1</b>	<b>Ethiopic</b>	445
	Ethiopic: U+1200–U+137F	445
	Ethiopic Extensions: U+1380–U+139F, U+2D80–U+2DDF	448
<b>13.2</b>	<b>Mongolian</b>	448
<b>13.3</b>	<b>Osmanya</b>	457
<b>13.4</b>	<b>Tifinagh</b>	457
<b>13.5</b>	<b>N’Ko</b>	458
<b>13.6</b>	<b>Cherokee</b>	463
<b>13.7</b>	<b>Canadian Aboriginal Syllabics</b>	464
<b>13.8</b>	<b>Deseret</b>	465
<b>13.9</b>	<b>Shavian</b>	467
<b>14</b>	<b>Archaic Scripts</b>	471
<b>14.1</b>	<b>Ogham</b>	472
<b>14.2</b>	<b>Old Italic</b>	473
<b>14.3</b>	<b>Runic</b>	475
<b>14.4</b>	<b>Gothic</b>	477
<b>14.5</b>	<b>Linear B</b>	478
	Linear B Syllabary: U+10000–U+1007F	478
	Linear B Ideograms: U+10080–U+108FF	479
	Aegean Numbers: U+10100–U+1013F	479

14.6	<b>Cypriot Syllabary</b> .....	479
14.7	<b>Phoenician</b> .....	480
14.8	<b>Ugaritic</b> .....	482
14.9	<b>Old Persian</b> .....	483
14.10	<b>Sumero-Akkadian</b> .....	483
	Cuneiform: U+12000–U+123FF .....	483
	Cuneiform Numbers and Punctuation: U+12400–U+1247F .....	486
15	<b>Symbols</b> .....	489
15.1	<b>Currency Symbols</b> .....	490
15.2	<b>Letterlike Symbols</b> .....	492
	Letterlike Symbols: U+2100–U+214F .....	492
	Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF .....	494
	Mathematical Alphabets .....	494
	Fonts Used for Mathematical Alphabets .....	497
15.3	<b>Number Forms</b> .....	498
	Number Forms: U+2150–U+218F .....	498
	CJK Number Forms .....	499
	Superscripts and Subscripts: U+2070–U+209F .....	501
15.4	<b>Mathematical Symbols</b> .....	502
	Mathematical Operators: U+2200–U+22FF .....	503
	Supplements to Mathematical Symbols and Arrows .....	505
	Supplemental Mathematical Operators: U+2A00–U+2AFF .....	505
	Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF .....	505
	Miscellaneous Mathematical Symbols-B: U+2980–U+29FF .....	505
	Arrows: U+2190–U+21FF .....	506
	Supplemental Arrows .....	506
	Standardized Variants of Mathematical Symbols .....	507
15.5	<b>Invisible Mathematical Operators</b> .....	507
15.6	<b>Technical Symbols</b> .....	508
	Control Pictures: U+2400–U+243F .....	508
	Miscellaneous Technical: U+2300–U+23FF .....	508
	Optical Character Recognition: U+2440–U+245F .....	511
15.7	<b>Geometrical Symbols</b> .....	512
	Box Drawing and Box Elements .....	512
	Geometric Shapes: U+25A0–U+25FF .....	513
15.8	<b>Miscellaneous Symbols and Dingbats</b> .....	514
	Miscellaneous Symbols: U+2600–U+26FF .....	514
	Dingbats: U+2700–U+27BF .....	515
	Yijing Hexagram Symbols: U+4DC0–U+4DFF .....	516
	Tai Xuan Jing Symbols: U+1D300–U+1D356 .....	517
15.9	<b>Enclosed and Square</b> .....	517
	Enclosed Alphanumerics: U+2460–U+24FF .....	517

Enclosed CJK Letters and Months: U+3200–U+32FF	518
CJK Compatibility: U+3300–U+33FF	518
<b>15.10 Braille</b>	519
<b>15.11 Western Musical Symbols</b>	520
<b>15.12 Byzantine Musical Symbols</b>	525
<b>15.13 Ancient Greek Musical Notation</b>	526
<b>16 Special Areas and Format Characters</b>	531
<b>16.1 Control Codes</b>	532
Representing Control Sequences	532
Specification of Control Code Semantics	533
<b>16.2 Layout Controls</b>	534
Line and Word Breaking	534
Cursive Connection and Ligatures	536
Combining Grapheme Joiner	540
Bidirectional Ordering Controls	542
<b>16.3 Deprecated Format Characters</b>	543
<b>16.4 Variation Selectors</b>	545
<b>16.5 Private-Use Characters</b>	546
Private Use Area: U+E000–U+F8FF	547
Supplementary Private Use Areas	548
<b>16.6 Surrogates Area</b>	548
<b>16.7 Noncharacters</b>	549
<b>16.8 Specials</b>	550
Byte Order Mark (BOM): U+FEFF	550
Specials: U+FFF0–U+FFF8	552
Annotation Characters: U+FFF9–U+FFFB	552
Replacement Characters: U+FFFC–U+FFFD	554
<b>16.9 Tag Characters</b>	554
Tag Characters: U+E000–U+E07F	554
Syntax for Embedding Tags	555
Working with Language Tags	557
Unicode Conformance Issues	559
Formal Tag Syntax	559
<b>17 Code Charts</b>	563
<b>17.1 Character Names List</b>	563
Images in the Code Charts and Character Lists	564
Character Names	565
Informative Aliases	565
Normative Aliases	566
Cross References	566
Information About Languages	566
Case Mappings	567

Decompositions	567
Reserved Characters	568
Noncharacters	568
Subheads	569
<b>17.2 CJK Unified Ideographs</b>	569
<b>17.3 Hangul Syllables</b>	570
<b>18 Han Radical-Stroke Index</b>	1023
<b>A Notational Conventions</b>	1077
Code Points	1077
Character Names	1077
Character Blocks	1077
Sequences	1078
Rendering	1078
Properties and Property Values	1079
Miscellaneous	1079
Extended BNF	1079
Operators	1081
<b>B Unicode Publications and Resources</b>	1083
<b>B.1 The Unicode Consortium</b>	1083
The Unicode Technical Committee	1084
Other Activities	1084
<b>B.2 Unicode Publications</b>	1084
<b>B.3 Unicode Technical Standards</b>	1085
UTS #6: A Standard Compression Scheme for Unicode	1085
UTS #10: Unicode Collation Algorithm	1085
UTS #18: Unicode Regular Expression Guidelines	1085
UTS #22: Character Mapping Markup Language (CharMapML)	1085
UTS #35: Locale Data Markup Language (LDML)	1085
UTS #37: Ideographic Variation Database	1086
UTS #39: Unicode Security Mechanisms	1086
<b>B.4 Unicode Technical Reports</b>	1086
UTR #16: UTF-EBCDIC	1086
UTR #17: Character Encoding Model	1086
UTR #20: Unicode in XML and Other Markup Languages	1086
UTR #23: The Unicode Character Property Model	1086
UTR #25: Unicode Support for Mathematics	1086
UTR #26: Compatibility Encoding Scheme for UTF-16: 8-Bit (CESU-8)	1087
UTR #30: Character Foldings	1087
UTR #36: Unicode Security Considerations	1087
<b>B.5 Unicode Technical Notes</b>	1087
<b>B.6 Other Unicode Online Resources</b>	1088
Unicode Online Resources	1088
How to Contact the Unicode Consortium	1089

<b>C</b>	<b>Relationship to ISO/IEC 10646</b> .....	1091
	<b>C.1 History</b> .....	1091
	<b>C.2 Encoding Forms in ISO/IEC 10646</b> .....	1095
	Zero Extending .....	1095
	<b>C.3 UCS Transformation Formats</b> .....	1096
	UTF-8 .....	1096
	UTF-16 .....	1096
	<b>C.4 Synchronization of the Standards</b> .....	1097
	<b>C.5 Identification of Features for the Unicode Standard</b> .....	1097
	<b>C.6 Character Names</b> .....	1098
	<b>C.7 Character Functional Specifications</b> .....	1098
<b>D</b>	<b>Changes from Previous Versions</b> .....	1099
	<b>D.1 Improvements to the Standard</b> .....	1099
	<b>D.2 Versions of the Unicode Standard</b> .....	1100
	<b>D.3 Clause and Definition Numbering Changes</b> .....	1102
	<b>D.4 Changes from Version 4.1 to Version 5.0</b> .....	1104
	<b>D.5 Changes from Version 4.0 to Version 4.1</b> .....	1106
	<b>D.6 Changes from Unicode Version 3.2 to Version 4.0</b> .....	1109
	<b>D.7 Changes from Unicode Version 3.1 to Version 3.2</b> .....	1111
	<b>D.8 Changes from Unicode Version 3.0 to Version 3.1</b> .....	1113
<b>E</b>	<b>Han Unification History</b> .....	1115
	<b>E.1 Development of the URO</b> .....	1115
	<b>E.2 Ideographic Rapporteur Group</b> .....	1116
<b>F</b>	<b>Unicode Encoding Stability Policies</b> .....	1119
	<b>F.1 Encoding Stability Policies for the Unicode Standard</b> .....	1119
	Encoding Stability .....	1119
	Name Stability .....	1120
	Formal Name Alias Stability .....	1120
	Named Character Sequence Stability .....	1120
	Name Uniqueness .....	1121
	Normalization Stability .....	1121
	Identity Stability .....	1122
	Property Value Stability .....	1122
	Identifier Stability .....	1123
	Case Folding Stability .....	1124
<b>G</b>	<b>Glossary</b> .....	1125
<b>R</b>	<b>References</b> .....	1153
	<b>R.1 Source Standards and Specifications</b> .....	1153
	<b>R.2 Source Dictionaries for Han Unification</b> .....	1161
	<b>R.3 Other Sources for the Unicode Standard</b> .....	1161

<b>R.4</b>	<b>Selected Resources: Technical</b> .....	1171
<b>R.5</b>	<b>Selected Resources: Scripts and Languages</b> .....	1173
<b>I</b>	<b>Indices</b> .....	1179
<b>I.1</b>	<b>Unicode Names Index</b> .....	1179
<b>I.2</b>	<b>General Index</b> .....	1231
<b>X</b>	<b>Annexes</b> .....	1251
<b>UAX 9:</b>	<b>The Bidirectional Algorithm</b> .....	1251
<b>UAX 11:</b>	<b>East Asian Width</b> .....	1275
<b>UAX 14:</b>	<b>Line Breaking Properties</b> .....	1283
<b>UAX 15:</b>	<b>Unicode Normalization Forms</b> .....	1333
<b>UAX 24:</b>	<b>Script Names</b> .....	1365
<b>UAX 29:</b>	<b>Text Boundaries</b> .....	1373
<b>UAX 31:</b>	<b>Identifier and Pattern Syntax</b> .....	1393
<b>UAX 34:</b>	<b>Unicode Named Character Sequences</b> .....	1405
<b>UAX 41:</b>	<b>Common References for Unicode Standard Annexes</b> ...	1411