

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

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. For sales outside the United States please contact International Sales, international@pearsoned.com

Visit us on the Web: 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

List of Figures	<i>xxiii</i>
List of Tables	<i>xxvii</i>
Foreword by Mark Davis	<i>xxxi</i>
Preface	<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>
Acknowledgments	<i>xxxix</i>
Unicode Consortium Members	<i>xlvii</i>
Unicode Consortium Liaison Members	<i>xliv</i>
Unicode Consortium Board of Directors	<i>l</i>
1 Introduction	<i>1</i>
1.1 Coverage	<i>2</i>
Standards Coverage	<i>3</i>
New Characters	<i>3</i>
1.2 Design Goals	<i>4</i>
1.3 Text Handling	<i>5</i>
Text Elements	<i>6</i>
2 General Structure	<i>9</i>
2.1 Architectural Context	<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>
2.2 Unicode Design Principles	<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
2.3	Compatibility Characters	23
	Compatibility Variants	23
	Compatibility Decomposable Characters	24
	Mapping Compatibility Characters	25
2.4	Code Points and Characters	25
	Types of Code Points	26
2.5	Encoding Forms	28
	UTF-32	31
	UTF-16	31
	UTF-8	32
	Comparison of the Advantages of UTF-32, UTF-16, and UTF-8	33
2.6	Encoding Schemes	35
2.7	Unicode Strings	37
2.8	Unicode Allocation	38
	Planes	39
	Allocation Areas and Character Blocks	39
	Assignment of Code Points	41
2.9	Details of Allocation	41
	Plane 0 (BMP)	43
	Plane 1	43
	Plane 2	46
	Other Planes	46
2.10	Writing Direction	46
2.11	Combining Characters	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
2.12	Equivalent Sequences and Normalization	54
2.13	Special Characters and Noncharacters	57
	Special Noncharacter Code Points	57
	Byte Order Mark (BOM)	57
	Layout and Format Control Characters	58
	The Replacement Character	58
	Control Codes	58
2.14	Conforming to the Unicode Standard	59
	Characteristics of Conformant Implementations	59
	Unacceptable Behavior	60
	Acceptable Behavior	60
	Supported Subsets	61

3	Conformance	65
3.1	Versions of the Unicode Standard	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
3.2	Conformance Requirements	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
3.3	Semantics	76
	Definitions	76
	Character Identity and Semantics	76
3.4	Characters and Encoding	78
3.5	Properties	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
3.6	Combination	91
3.7	Decomposition	95
	Compatibility Decomposition	95
	Canonical Decomposition	96
3.8	Surrogates	97
3.9	Unicode Encoding Forms	98
	UTF-32	102
	UTF-16	102

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

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

5.19	Unicode Security	190
5.20	Default Ignorable Code Points	192
6	Writing Systems and Punctuation	197
6.1	Writing Systems	198
6.2	General Punctuation	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	European Alphabetic Scripts	225
7.1	Latin	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	Greek	237
	Greek: U+0370–U+03FF.....	237
	Greek Extended: U+1F00–U+1FFF.....	241
	Ancient Greek Numbers: U+10140–U+1018F.....	242
7.3	Coptic	243
7.4	Cyrillic	245
	Cyrillic: U+0400–U+04FF.....	245
	Cyrillic Supplement: U+0500–U+052F.....	246
7.5	Glagolitic	246
7.6	Armenian	247
7.7	Georgian	249

7.8	Modifier Letters	250
	Spacing Modifier Letters: U+02B0–U+02FF	250
	Modifier Tone Letters: U+A700–U+A71F	252
7.9	Combining Marks	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
8	Middle Eastern Scripts	263
8.1	Hebrew	264
	Hebrew: U+0590–U+05FF	264
	Alphabetic Presentation Forms: U+FB1D–U+FB4F	269
8.2	Arabic	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
8.3	Syriac	283
	Syriac: U+0700–U+074F	283
	Syriac Shaping	288
	Syriac Cursive Joining	288
	Syriac Ligatures	290
8.4	Thaana	291
9	South Asian Scripts-I	295
9.1	Devanagari	296
	Devanagari: U+0900–U+097F	296
	Principles of the Devanagari Script	297
	Rendering Devanagari	303
9.2	Bengali	312
9.3	Gurmukhi	317
9.4	Gujarati	321
9.5	Oriya	322
9.6	Tamil	324
	Tamil: U+0B80–U+0BFF	324
	Tamil Vowels	325
	Tamil Ligatures	327
9.7	Telugu	330
9.8	Kannada	331
	Kannada: U+0C80–U+0CFF	331

	Principles of the Kannada Script	332
	Rendering Kannada	333
9.9	Malayalam	334
10	South Asian Scripts-II	341
10.1	Sinhala	341
10.2	Tibetan	343
10.3	Phags-pa	353
10.4	Limbu	360
10.5	Syloti Nagri	363
10.6	Kharoshthi	364
	Kharoshthi: U+10A00–U+10A5F	364
	Rendering Kharoshthi	366
11	Southeast Asian Scripts	373
11.1	Thai	373
11.2	Lao	376
11.3	Myanmar	379
11.4	Khmer	382
	Khmer: U+1780–U+17FF	382
	Principles of the Khmer Script	382
	Khmer Symbols: U+19E0–U+19FF	392
11.5	Tai Le	393
11.6	New Tai Lue	394
11.7	Philippine Scripts	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
11.8	Buginese	397
11.9	Balinese	399
12	East Asian Scripts	407
12.1	Han	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
12.2	Ideographic Description Characters	427
12.3	Bopomofo	431
12.4	Hiragana and Katakana	433
	Hiragana: U+3040–U+309F	433
	Katakana: U+30A0–U+30FF	433
	Katakana Phonetic Extensions: U+31F0–U+31FF	434
12.5	Halfwidth and Fullwidth Forms	434
12.6	Hangul	435
	Hangul Jamo: U+1100–U+11FF	435
	Hangul Compatibility Jamo: U+3130–U+318F	436
	Hangul Syllables: U+AC00–U+D7A3	437
12.7	Yi	438
13	Additional Modern Scripts	445
13.1	Ethiopic	445
	Ethiopic: U+1200–U+137F	445
	Ethiopic Extensions: U+1380–U+139F, U+2D80–U+2DDF	448
13.2	Mongolian	448
13.3	Osmanya	457
13.4	Tifinagh	457
13.5	N’Ko	458
13.6	Cherokee	463
13.7	Canadian Aboriginal Syllabics	464
13.8	Deseret	465
13.9	Shavian	467
14	Archaic Scripts	471
14.1	Ogham	472
14.2	Old Italic	473
14.3	Runic	475
14.4	Gothic	477
14.5	Linear 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	Cypriot Syllabary	479
14.7	Phoenician	480
14.8	Ugaritic	482
14.9	Old Persian	483
14.10	Sumero-Akkadian	483
	Cuneiform: U+12000–U+123FF	483
	Cuneiform Numbers and Punctuation: U+12400–U+1247F	486
15	Symbols	489
15.1	Currency Symbols	490
15.2	Letterlike Symbols	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	Number Forms	498
	Number Forms: U+2150–U+218F	498
	CJK Number Forms	499
	Superscripts and Subscripts: U+2070–U+209F	501
15.4	Mathematical Symbols	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	Invisible Mathematical Operators	507
15.6	Technical Symbols	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	Geometrical Symbols	512
	Box Drawing and Box Elements	512
	Geometric Shapes: U+25A0–U+25FF	513
15.8	Miscellaneous Symbols and Dingbats	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	Enclosed and Square	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
15.10 Braille	519
15.11 Western Musical Symbols	520
15.12 Byzantine Musical Symbols	525
15.13 Ancient Greek Musical Notation	526
16 Special Areas and Format Characters	531
16.1 Control Codes	532
Representing Control Sequences	532
Specification of Control Code Semantics	533
16.2 Layout Controls	534
Line and Word Breaking	534
Cursive Connection and Ligatures	536
Combining Grapheme Joiner	540
Bidirectional Ordering Controls	542
16.3 Deprecated Format Characters	543
16.4 Variation Selectors	545
16.5 Private-Use Characters	546
Private Use Area: U+E000–U+F8FF	547
Supplementary Private Use Areas	548
16.6 Surrogates Area	548
16.7 Noncharacters	549
16.8 Specials	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
16.9 Tag Characters	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
17 Code Charts	563
17.1 Character Names List	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
17.2 CJK Unified Ideographs	569
17.3 Hangul Syllables	570
18 Han Radical-Stroke Index	1023
A Notational Conventions	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 Unicode Publications and Resources	1083
B.1 The Unicode Consortium	1083
The Unicode Technical Committee	1084
Other Activities	1084
B.2 Unicode Publications	1084
B.3 Unicode Technical Standards	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.4 Unicode Technical Reports	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.5 Unicode Technical Notes	1087
B.6 Other Unicode Online Resources	1088
Unicode Online Resources	1088
How to Contact the Unicode Consortium	1089

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

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