

This PDF file is an excerpt from *The Unicode Standard, Version 4.0*, issued by the Unicode Consortium and published by Addison-Wesley. The material has been modified slightly for this online 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/standard/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 Addison-Wesley was aware of a trademark claim, the designations have been printed in initial capital letters. However, not all words in initial capital letters are trademark designations.

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 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, <http://www.mehallo.com>

The publisher offers discounts on this book when ordered in quantity for bulk purchases and special sales. For more information, customers in the U.S. please contact U.S. Corporate and Government Sales, (800) 382-3419, [corpsales@pearsontechgroup.com](mailto:corpsales@pearsontechgroup.com). For sales outside of the U.S., please contact International Sales, +1 317 581 3793, [international@pearsontechgroup.com](mailto:international@pearsontechgroup.com)

Visit Addison-Wesley on the Web: <http://www.awprofessional.com>

*Library of Congress Cataloging-in-Publication Data*

The Unicode Standard, Version 4.0 : the Unicode Consortium /Joan Aliprand... [et al.].

p. cm.

Includes bibliographical references and index.

ISBN 0-321-18578-1 (alk. paper)

1. Unicode (Computer character set). I. Aliprand, Joan.

QA268.U545 2004

005.7'2—dc21

2003052158

Copyright © 1991–2003 by Unicode, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher or Unicode, Inc. Printed in the United States of America. Published simultaneously in Canada.

For information on obtaining permission for use of material from this work, please submit a written request to the Unicode Consortium, Post Office Box 39146, Mountain View, CA 94039-1476, USA, Fax +1 650 693 3010 or to Pearson Education, Inc., Rights and Contracts Department, 75 Arlington Street, Suite 300 Boston, MA 02116, USA, Fax: +1 617 848 7047.

ISBN 0-321-18578-1

Text printed on recycled paper

1 2 3 4 5 6 7 8 9 10—CRW—0706050403

First printing, August 2003

# Contents

<b>Acknowledgments</b> .....	iii
<b>Unicode Consortium Members and Directors</b> .....	ix
Full Members .....	ix
Current Associate Members .....	ix
Current Liaison Members .....	ix
Current Specialist Members .....	x
Current Individual Members .....	x
Current Members of the Board of Directors .....	x
Former Members of the Board of Directors .....	x
<b>Figures</b> .....	xxiii
<b>Tables</b> .....	xxv
<b>Preface</b> .....	xxxix
<b>0.1 About the Unicode Standard</b> .....	xxxix
Concepts, Architecture, Conformance, and Guidelines .....	xxxix
Character Block Descriptions .....	xxxix
Charts and Han Radical-Stroke Index .....	xxxix
Appendices .....	xxxix
<b>0.2 The Unicode Character Database and Technical Reports</b> ..	xxxix
On the CD-ROM .....	xxxix
<b>0.3 Notational Conventions</b> .....	xxxix
Code Points .....	xxxix
Character Names .....	xxxix
Sequences .....	xxxix
Miscellaneous .....	xxxix
Extended BNF .....	xxxix
Operators .....	xxxix
<b>0.4 Resources</b> .....	xxxix
Unicode Web Site .....	xxxix
Unicode Anonymous FTP Site .....	xxxix
Unicode E-mail Discussion List .....	xxxix
How to Contact the Unicode Consortium .....	xxxix
<b>1 Introduction</b> .....	1
<b>1.1 Coverage</b> .....	2
Standards Coverage .....	3
New Characters .....	3
<b>1.2 Design Goals</b> .....	3
<b>1.3 Text Handling</b> .....	4
Interpreting Characters .....	5
Text Elements .....	5
<b>1.4 The Unicode Standard and ISO/IEC 10646</b> .....	5
<b>1.5 The Unicode Consortium</b> .....	6
The Unicode Technical Committee .....	6
<b>1.6 Submitting New Characters</b> .....	7

<b>2</b>	<b>General Structure</b>	11
<b>2.1</b>	<b>Architectural Context</b>	11
	Basic Text Processes	12
	Text Elements, Characters, and Text Processes	12
	Text Processes and Encoding	13
<b>2.2</b>	<b>Unicode Design Principles</b>	14
	Universality	14
	Efficiency	15
	Characters, Not Glyphs	15
	Semantics	17
	Plain Text	18
	Logical Order	18
	Unification	19
	Dynamic Composition	20
	Equivalent Sequences	20
	Convertibility	22
<b>2.3</b>	<b>Compatibility Characters</b>	23
	Compatibility Characters	23
	Compatibility Decomposable Characters	23
	Mapping Compatibility Characters	23
<b>2.4</b>	<b>Code Points and Characters</b>	24
	Types of Code Points	25
<b>2.5</b>	<b>Encoding Forms</b>	26
	UTF-32	29
	UTF-16	29
	UTF-8	30
	Comparison of the Advantages of UTF-32, UTF-16, and UTF-8	31
<b>2.6</b>	<b>Encoding Schemes</b>	32
<b>2.7</b>	<b>Unicode Strings</b>	34
<b>2.8</b>	<b>Unicode Allocation</b>	35
	Planes	35
	Allocation Areas and Character Blocks	36
	Details of Allocation	36
	Assignment of Code Points	42
<b>2.9</b>	<b>Writing Direction</b>	42
<b>2.10</b>	<b>Combining Characters</b>	43
	Sequence of Base Characters and Diacritics	44
	Multiple Combining Characters	44
	Ligated Multiple Base Characters	46
	Spacing Clones of European Diacritical Marks	46
	“Characters” and Grapheme Clusters	46
<b>2.11</b>	<b>Special Characters and Noncharacters</b>	47
	Byte Order Mark (BOM)	47
	Special Noncharacter Code Points	48
	Layout and Format Control Characters	48
	The Replacement Character	48
	Control Codes	48
<b>2.12</b>	<b>Conforming to the Unicode Standard</b>	48
	Supported Subsets	50
<b>2.13</b>	<b>Related Publications</b>	50

<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, Corrigenda, and Future Updates	56
	References to the Unicode Standard	57
	References to Unicode Character Properties	57
	References to Unicode Algorithms	57
<b>3.2</b>	<b>Conformance Requirements</b>	58
	Byte Ordering	58
	Unassigned Code Points	58
	Interpretation	59
	Modification	60
	Character Encoding Forms	60
	Character Encoding Schemes	61
	Bidirectional Text	61
	Normalization Forms	62
	Normative References	62
	Unicode Algorithms	62
	Default Casing Operations	62
	Unicode Standard Annexes	62
<b>3.3</b>	<b>Semantics</b>	63
	Definitions	63
	Character Identity and Semantics	63
<b>3.4</b>	<b>Characters and Encoding</b>	64
<b>3.5</b>	<b>Properties</b>	66
	Normative and Informative Properties	66
	Simple and Derived Properties	68
	Property Aliases	68
	Default Property Values	69
	Private Use	69
<b>3.6</b>	<b>Combination</b>	70
<b>3.7</b>	<b>Decomposition</b>	71
	Compatibility Decomposition	71
	Canonical Decomposition	72
<b>3.8</b>	<b>Surrogates</b>	72
<b>3.9</b>	<b>Unicode Encoding Forms</b>	73
	UTF-32	76
	UTF-16	76
	UTF-8	77
	Encoding Form Conversion	78
<b>3.10</b>	<b>Unicode Encoding Schemes</b>	78
<b>3.11</b>	<b>Canonical Ordering Behavior</b>	82
	Application of Combining Marks	83
	Combining Classes	83
	Canonical Ordering	84
	Canonical Ordering and Collation	85
<b>3.12</b>	<b>Conjoining Jamo Behavior</b>	85
	Hangul Syllable Boundaries	86
	Standard Korean Syllables	86
	Hangul Syllable Composition	87

	Hangul Syllable Decomposition .....	88
	Hangul Syllable Names .....	88
<b>3.13</b>	<b>Default Case Operations</b> .....	89
	Definitions .....	89
	Case Conversion of Strings .....	90
	Case Detection for Strings .....	90
	Caseless Matching .....	91
<b>4</b>	<b>Character Properties</b> .....	95
<b>4.1</b>	<b>Unicode Character Database</b> .....	96
<b>4.2</b>	<b>Case—Normative</b> .....	96
	Case Mapping .....	97
<b>4.3</b>	<b>Combining Classes—Normative</b> .....	97
<b>4.4</b>	<b>Directionality—Normative</b> .....	98
<b>4.5</b>	<b>General Category—Normative</b> .....	98
<b>4.6</b>	<b>Numeric Value—Normative</b> .....	100
	Ideographic Numeric Values .....	100
<b>4.7</b>	<b>Bidi Mirrored—Normative</b> .....	101
<b>4.8</b>	<b>Unicode 1.0 Names</b> .....	101
<b>4.9</b>	<b>Letters, Alphabetic, and Ideographic</b> .....	102
<b>4.10</b>	<b>Boundary Control</b> .....	102
<b>4.11</b>	<b>Characters with Unusual Properties</b> .....	103
<b>5</b>	<b>Implementation Guidelines</b> .....	107
<b>5.1</b>	<b>Transcoding to Other Standards</b> .....	107
	Issues .....	107
	Multistage Tables .....	108
<b>5.2</b>	<b>ANSI/ISO C <code>wchar_t</code></b> .....	109
<b>5.3</b>	<b>Unknown and Missing Characters</b> .....	110
	Reserved and Private-Use Character Codes .....	110
	Interpretable but Unrenderable Characters .....	110
	Default Property Values .....	110
	Default Ignorable Code Points .....	111
	Interacting with Downlevel Systems .....	111
<b>5.4</b>	<b>Handling Surrogate Pairs in UTF-16</b> .....	111
<b>5.5</b>	<b>Handling Numbers</b> .....	114
<b>5.6</b>	<b>Normalization</b> .....	114
<b>5.7</b>	<b>Compression</b> .....	115
<b>5.8</b>	<b>Newline Guidelines</b> .....	116
	Definitions .....	116
	Background .....	117
	Recommendations .....	118
<b>5.9</b>	<b>Regular Expressions</b> .....	119
<b>5.10</b>	<b>Language Information in Plain Text</b> .....	119
	Requirements for Language Tagging .....	119
	Language Tags and Han Unification .....	120
<b>5.11</b>	<b>Editing and Selection</b> .....	121
	Consistent Text Elements .....	121
<b>5.12</b>	<b>Strategies for Handling Nonspacing Marks</b> .....	122
	Keyboard Input .....	123
	Truncation .....	124

5.13	<b>Rendering Nonspacing Marks</b> .....	125
	Canonical Equivalence .....	127
	Positioning Methods .....	128
5.14	<b>Locating Text Element Boundaries</b> .....	130
5.15	<b>Identifiers</b> .....	130
	Property-Based Identifier Syntax .....	130
	Syntactic Rule .....	131
	Alternative Recommendation .....	132
5.16	<b>Sorting and Searching</b> .....	132
	Culturally Expected Sorting and Searching .....	133
	Language-Insensitive Sorting .....	133
	Searching .....	133
	Sublinear Searching .....	134
5.17	<b>Binary Order</b> .....	135
	UTF-8 in UTF-16 Order .....	135
	UTF-16 in UTF-8 Order .....	136
5.18	<b>Case Mappings</b> .....	136
	Complications for Case Mapping .....	137
	Reversibility .....	138
	Caseless Matching .....	138
	Normalization .....	139
5.19	<b>Unicode Security</b> .....	140
5.20	<b>Default Ignorable Code Points</b> .....	142
6	<b>Writing Systems and Punctuation</b> .....	147
6.1	<b>Writing Systems</b> .....	148
6.2	<b>General Punctuation</b> .....	152
	Punctuation: U+0020–U+00BF .....	152
	General Punctuation: U+2000–U+206F .....	154
	CJK Symbols and Punctuation: U+3000–U+303F .....	160
	CJK Compatibility Forms: U+FE30–U+FE4F .....	161
	Small Form Variants: U+FE50–U+FE6F .....	162
7	<b>European Alphabetic Scripts</b> .....	165
7.1	<b>Latin</b> .....	166
	Letters of Basic Latin: U+0041–U+007A .....	166
	Letters of the Latin-1 Supplement: U+00C0–U+00FF .....	167
	Latin Extended-A: U+0100–U+017F .....	167
	Latin Extended-B: U+0180–U+024F .....	169
	IPA Extensions: U+0250–U+02AF .....	170
	Phonetic Extensions: U+1D00–U+1D6A .....	171
	Latin Extended Additional: U+1E00–U+1EFF .....	172
	Latin Ligatures: FB00–FB06 .....	173
7.2	<b>Greek</b> .....	174
	Greek: U+0370–U+03FF .....	174
	Greek Extended: U+1F00–U+1FFF .....	177
7.3	<b>Cyrillic</b> .....	179
	Cyrillic: U+0400–U+04FF .....	179
	Cyrillic Supplement: U+0500–U+052F .....	179
7.4	<b>Armenian</b> .....	180
	Armenian: U+0530–U+058F .....	180
7.5	<b>Georgian</b> .....	182
	Georgian: U+10A0–U+10FF .....	182

<b>7.6</b>	<b>Modifier Letters</b> .....	184
	Spacing Modifier Letters: U+02B0–U+02FF .....	184
<b>7.7</b>	<b>Combining Marks</b> .....	186
	Combining Diacritical Marks: U+0300–U+036F .....	186
	Combining Marks for Symbols: U+20D0–U+20FF .....	188
	Combining Half Marks: U+FE20–U+FE2F .....	188
<b>8</b>	<b>Middle Eastern Scripts</b> .....	191
<b>8.1</b>	<b>Hebrew</b> .....	192
	Hebrew: U+0590–U+05FF .....	192
	Alphabetic Presentation Forms: U+FB1D–U+FB4F .....	194
<b>8.2</b>	<b>Arabic</b> .....	195
	Arabic: U+0600–U+06FF .....	195
	Cursive Joining .....	199
	Ligatures .....	201
	Arabic Presentation Forms-A: U+FB50–U+FDFF .....	204
	Arabic Presentation Forms-B: U+FE70–U+FEFF .....	205
<b>8.3</b>	<b>Syriac</b> .....	206
	Syriac: U+0700–U+074F .....	206
	Syriac Shaping .....	210
	Syriac Cursive Joining .....	210
	Ligatures .....	212
<b>8.4</b>	<b>Thaana</b> .....	213
	Thaana: U+0780–U+07BF .....	213
<b>9</b>	<b>South Asian Scripts</b> .....	217
<b>9.1</b>	<b>Devanagari</b> .....	219
	Devanagari: U+0900–U+097F .....	219
<b>9.2</b>	<b>Bengali</b> .....	232
	Bengali: U+0980–U+09FF .....	232
<b>9.3</b>	<b>Gurmukhi</b> .....	234
	Gurmukhi: U+0A00–U+0A7F .....	234
<b>9.4</b>	<b>Gujarati</b> .....	236
	Gujarati: U+0A80–U+0AFF .....	236
<b>9.5</b>	<b>Oriya</b> .....	237
	Oriya: U+0B00–U+0B7F .....	237
<b>9.6</b>	<b>Tamil</b> .....	239
	Tamil: U+0B80–U+0BFF .....	239
<b>9.7</b>	<b>Telugu</b> .....	244
	Telugu: U+0C00–U+0C7F .....	244
<b>9.8</b>	<b>Kannada</b> .....	245
	Kannada: U+0C80–U+0CFF .....	245
<b>9.9</b>	<b>Malayalam</b> .....	248
	Malayalam: U+0D00–U+0D7F .....	248
<b>9.10</b>	<b>Sinhala</b> .....	250
	Sinhala: U+0D80–U+0DFF .....	250
<b>9.11</b>	<b>Tibetan</b> .....	251
	Tibetan: U+0F00–U+0FFF .....	251
<b>9.12</b>	<b>Limbu</b> .....	260
	Limbu: U+1900–U+194F .....	260

<b>10 Southeast Asian Scripts</b> .....	265
<b>10.1 Thai</b> .....	266
Thai: U+0E00–U+0E7F .....	266
<b>10.2 Lao</b> .....	269
Lao: U+0E80–U+0EFF .....	269
<b>10.3 Myanmar</b> .....	271
Myanmar: U+1000–U+109F .....	271
<b>10.4 Khmer</b> .....	274
Khmer: U+1780–U+17FF .....	274
Khmer Symbols: U+19E0–U+19FF .....	283
<b>10.5 Tai Le</b> .....	284
Tai Le: U+1950–U+197F .....	284
<b>10.6 Philippine Scripts</b> .....	286
Tagalog: U+1700–U+171F .....	286
Hanunóo: U+1720–U+173F .....	286
Buhid: U+1740–U+175F .....	286
Tagbanwa: U+1760–U+177F .....	286
<b>11 East Asian Scripts</b> .....	291
<b>11.1 Han</b> .....	293
CJK Unified Ideographs .....	293
CJK Unified Ideographs Ext. B: U+20000–U+2A6D6 .....	304
CJK Compatibility Ideographs: U+F900–U+FAFF .....	305
CJK Compatibility Supplement: U+2F800–U+2FA1D .....	305
Kanbun: U+3190–U+319F .....	305
CJK and KangXi Radicals: U+2E80–U+2FD5 .....	306
Ideographic Description: U+2FF0–U+2FFB .....	307
<b>11.2 Bopomofo</b> .....	310
Bopomofo: U+3100–U+312F .....	310
<b>11.3 Hiragana and Katakana</b> .....	312
Hiragana: U+3040–U+309F .....	312
Katakana: U+30A0–U+30FF .....	312
Katakana Phonetic Extensions: U+31F0–U+31FF .....	313
Halfwidth and Fullwidth Forms: U+FF00–U+FFEF .....	313
<b>11.4 Hangul</b> .....	314
Hangul Jamo: U+1100–U+11FF .....	314
Hangul Compatibility Jamo: U+3130–U+318F .....	314
Hangul Syllables: U+AC00–U+D7A3 .....	315
<b>11.5 Yi</b> .....	317
Yi: U+A000–U+A4CF .....	317
<b>12 Additional Modern Scripts</b> .....	321
<b>12.1 Ethiopic</b> .....	322
Ethiopic: U+1200–U+137F .....	322
<b>12.2 Mongolian</b> .....	325
Mongolian: U+1800–U+18AF .....	325
<b>12.3 Osmanya</b> .....	329
Osmanya: U+10480–U+104AF .....	329
<b>12.4 Cherokee</b> .....	330
Cherokee: U+13A0–U+13FF .....	330
<b>12.5 Canadian Aboriginal Syllabics</b> .....	331
Canadian Aboriginal Syllabics: U+1400–U+167F .....	331



12.6	<b>Deseret</b> .....	332
	Deseret: U+10400–U+1044F .....	332
12.7	<b>Shavian</b> .....	334
	Shavian: U+10450–U+1047F .....	334
13	<b>Archaic Scripts</b> .....	337
13.1	<b>Ogham</b> .....	338
	Ogham: U+1680–U+169F .....	338
13.2	<b>Old Italic</b> .....	339
	Old Italic: U+10300–U+1032F .....	339
13.3	<b>Runic</b> .....	341
	Runic: U+16A0–U+16F0 .....	341
13.4	<b>Gothic</b> .....	343
	Gothic: U+10330–U+1034F .....	343
13.5	<b>Ugaritic</b> .....	344
	Ugaritic: U+10380–U+1039F .....	344
13.6	<b>Linear B</b> .....	345
	Linear B Syllabary: U+10000–U+1007F .....	345
	Linear B Ideograms: U+10080–U+108FF .....	345
	Aegean Numbers: U+10100–U+1013F .....	345
13.7	<b>Cypriot Syllabary</b> .....	346
	Cypriot Syllabary: U+10800–U+1083F .....	346
14	<b>Symbols</b> .....	349
14.1	<b>Currency Symbols</b> .....	351
	Currency Symbols: U+20A0–U+20CF .....	351
14.2	<b>Letterlike Symbols</b> .....	353
	Letterlike Symbols: U+2100–U+214F .....	353
	Math Alphanumeric Symbols: U+1D400–U+1D7FF .....	354
	Mathematical Alphabets .....	354
	Fonts Used for Mathematical Alphabets .....	356
14.3	<b>Number Forms</b> .....	358
	Number Forms: U+2150–U+218F .....	358
	Superscripts and Subscripts: U+2070–U+209F .....	359
14.4	<b>Mathematical Symbols</b> .....	360
	Mathematical Operators: U+2200–U+22FF .....	360
	Supplements to Mathematical Symbols and Arrows .....	362
	Supplemental Math Operators: U+2A00–U+2AFF .....	362
	Miscellaneous Math Symbols-A: U+27C0–U+27EF .....	362
	Miscellaneous Math Symbols-B: U+2980–U+29FF .....	362
	Arrows: U+2190–U+21FF .....	363
	Supplemental Arrows .....	363
	Standardized Variants of Mathematical Symbols .....	363
14.5	<b>Technical Symbols</b> .....	365
	Control Pictures: U+2400–U+243F .....	365
	Miscellaneous Technical: U+2300–U+23FF .....	365
	Optical Character Recognition: U+2440–U+245F .....	367
14.6	<b>Geometrical Symbols</b> .....	368
	Box Drawing: U+2500–U+257F .....	368
	Block Elements: U+2580–U+259F .....	368
	Geometric Shapes: U+25A0–U+25FF .....	368

<b>14.7</b>	<b>Miscellaneous Symbols and Dingbats</b> .....	370
	Miscellaneous Symbols: U+2600–U+26FF .....	370
	Dingbats: U+2700–U+27BF .....	371
	Yijing Hexagram Symbols: U+4DC0–U+4DFF .....	372
	Tai Xuan Jing Symbols: U+1D300–U+1D356 .....	372
<b>14.8</b>	<b>Enclosed and Square</b> .....	373
	Enclosed Alphanumerics: U+2460–U+24FF .....	373
	Enclosed CJK Letters and Months: U+3200–U+32FF .....	373
	CJK Compatibility: U+3300–U+33FF .....	373
<b>14.9</b>	<b>Braille</b> .....	374
	Braille Patterns: U+2800–U+28FF .....	374
<b>14.10</b>	<b>Byzantine Musical Symbols</b> .....	376
	Byzantine Musical Symbols: U+1D000–U+1D0FF .....	376
<b>14.11</b>	<b>Western Musical Symbols</b> .....	377
	Musical Symbols: U+1D100–U+1D1FF .....	377
<b>15</b>	<b>Special Areas and Format Characters</b> .....	383
<b>15.1</b>	<b>Control Codes</b> .....	385
<b>15.2</b>	<b>Layout Controls</b> .....	387
<b>15.3</b>	<b>Invisible Operators</b> .....	393
<b>15.4</b>	<b>Deprecated Format Characters</b> .....	394
	Deprecated Format Characters: U+206A–U+206F .....	394
<b>15.5</b>	<b>Surrogates Area</b> .....	396
	Surrogates Area: U+D800–U+DFFF .....	396
<b>15.6</b>	<b>Variation Selectors</b> .....	397
<b>15.7</b>	<b>Private-Use Characters</b> .....	398
	Private Use Area: U+E000–U+F8FF .....	398
	Supplementary Private Use Areas .....	399
<b>15.8</b>	<b>Noncharacters</b> .....	400
	Noncharacters: U+FFFE, U+FFFF, and Others .....	400
<b>15.9</b>	<b>Specials</b> .....	401
	Specials: U+FEFF, U+FFF0–U+FFFD .....	401
<b>15.10</b>	<b>Tag Characters</b> .....	405
	Tag Characters: U+E0000–U+E007F .....	405
<b>16</b>	<b>Code Charts</b> .....	413
<b>16.1</b>	<b>Character Names List</b> .....	413
	Images in the Code Charts and Character Lists .....	414
	Character Names .....	414
	Aliases .....	415
	Cross References .....	415
	Information About Languages .....	415
	Case Mappings .....	415
	Decompositions .....	416
	Reserved Characters .....	417
	Noncharacters .....	417
	Subheads .....	417
<b>16.2</b>	<b>CJK Unified Ideographs</b> .....	417
<b>16.3</b>	<b>Hangul Syllables</b> .....	418
<b>17</b>	<b>Han Radical-Stroke Index</b> .....	1189

<b>A</b>	<b>Han Unification History</b> .....	1341
<b>B</b>	<b>Abstracts of Unicode Technical Reports</b> .....	1343
	<b>B.1 Unicode Standard Annexes</b> .....	1343
	UAX #9: The Bidirectional Algorithm .....	1343
	UAX #11: East Asian Width .....	1343
	UAX #14: Line Breaking Properties .....	1343
	UAX #15: Unicode Normalization Forms .....	1344
	UAX #24: Script Names .....	1344
	UAX #29: Text Boundaries .....	1344
	<b>B.2 Unicode Technical Standards</b> .....	1344
	UTS #6: A Standard Compression Scheme for Unicode .....	1344
	UTS #10: Unicode Collation Algorithm .....	1344
	<b>B.3 Unicode Technical Reports</b> .....	1344
	UTR #16: UTF-EBCDIC .....	1344
	UTR #17: Character Encoding Model .....	1344
	UTR #18: Unicode Regular Expression Guidelines .....	1344
	UTR #20: Unicode in XML and Other Markup Languages .....	1345
	UTR #22: Character Mapping Markup Language (CharMapML) .....	1345
	UTR #26: Compatibility Encoding Scheme for UTF-16: 8-Bit (CESU-8) ..	1345
	<b>B.4 Other Unicode References</b> .....	1345
	Unicode Technical Notes .....	1345
	FAQ (Frequently Asked Questions) .....	1346
	Charts .....	1346
	Conferences .....	1346
	Policies .....	1346
	Updates and Errata .....	1346
	Versions .....	1346
	Where Is My Character? .....	1346
<b>C</b>	<b>Relationship to ISO/IEC 10646</b> .....	1347
	<b>C.1 History</b> .....	1347
	Unicode 1.0 .....	1348
	Unicode 2.0 .....	1348
	Unicode 3.0 .....	1349
	Unicode 4.0 .....	1350
	<b>C.2 Encoding Forms in ISO/IEC 10646</b> .....	1350
	Zero Extending .....	1350
	<b>C.3 UCS Transformation Formats</b> .....	1351
	UTF-8 .....	1351
	UTF-16 .....	1351
	<b>C.4 Synchronization of the Standards</b> .....	1352
	<b>C.5 Identification of Features for the Unicode Standard</b> .....	1352
	<b>C.6 Character Names</b> .....	1353
	<b>C.7 Character Functional Specifications</b> .....	1353
<b>D</b>	<b>Changes from Unicode Version 3.0</b> .....	1355
	<b>D.1 Versions of the Unicode Standard</b> .....	1355
	<b>D.2 Changes from Unicode Version 3.0 to Version 3.1</b> .....	1357
	New Characters Added .....	1357
	Unicode Character Database Changes .....	1357
	Changes Affecting Conformance .....	1357
	Unicode Standard Annexes .....	1358

<b>D.3</b>	<b>Changes from Unicode Version 3.1 to Version 3.2</b> .....	1358
	New Characters Added .....	1358
	Unicode Character Database Changes .....	1358
	Changes Affecting Conformance .....	1358
	Unicode Standard Annexes .....	1359
<b>D.4</b>	<b>Changes from Unicode Version 3.2 to Version 4.0</b> .....	1359
	New Characters Added .....	1359
	Unicode Character Database Changes .....	1359
	Changes Affecting Conformance .....	1360
	Unicode Standard Annexes .....	1360
	Errata .....	1360
<b>G</b>	<b>Glossary</b> .....	1363
<b>R</b>	<b>References</b> .....	1385
	<b>R.1 Source Standards and Specifications</b> .....	1385
	<b>R.2 Source Dictionaries for Han Unification</b> .....	1391
	<b>R.3 Other Sources for the Unicode Standard</b> .....	1391
	<b>R.4 Selected Resources: Technical</b> .....	1397
	<b>R.5 Selected Resources: Scripts and Languages</b> .....	1400
<b>I</b>	<b>Indices</b> .....	1407
	<b>I.1 Unicode Names Index</b> .....	1407
	<b>I.2 General Index</b> .....	1449