
I.2 General Index

The general index is a guide to the written text only. To find the range for a particular script, see the Contents. To find the code for a specific character, see the preceding Unicode Names Index. For the definition of specific terms, see the Glossary.

A

Allocation	21–23
Alphabetic property	102
Annotation characters	325
Apostrophes	153
Arabic	189–198
Dual-joining characters (table)	195
Joining classes	192
Ligatures	194
Other character joining classes (table)	197
Right-joining characters (table)	196
Architecture	9–11
Armenian	172
Arrows	301

B

Bengali	224
Bidirectional algorithm	55–69
Bidirectional character types	59
Bidirectional ordering codes	318
Bidirectional placement	121
Block elements	304
BNF	<i>see</i> Extended BNF
Bopomofo	278
Boundaries	116–117, 124–133
Grapheme	126–127
Line	129–132
Sentence	132–133
Word	127–128
Box drawing	304
Braille	308
Burmese	<i>see</i> Myanmar
Byte order mark	28, 324–325
Byte ordering	37

C

Canadian Aboriginal Syllabics	288
Canonical decomposition	44
Canonical ordering behavior	50–52
Cantillation marks	187
Caron	162
Case	75
Case mappings	75, 141–143
Case pairs	164
Cedilla	162
Changes from version 2.0	973–980
Character classes	xxviii
Character codes	12
Character encoding schemes	21
Character names list	331

Character pairs

see Mirrored property

Character sequences	38
Character shaping selectors	320
Characters	40–42
Concept	13
Duplicate encoding	17
Editing	116–117
Equivalence	136
Interpretation	5, 30–32, 38
Missing and unknown	108
Modification	39
Names	101, 972
New in 3.0	3, 975
Number of assigned	973
Properties	42
Reassigned	109
Reserved	334
Transcoding	105–107
Cherokee	287
CJK compatibility	307
CJK compatibility forms	156
CJK enclosed letters and months	307
CJK punctuation	155
CJK standards	258
Code charts	331–335
Glyphs	332
Code space assignment	23
Code value	41
Collation	114, 139
Combining characters	24–27, 134
Canonical ordering	50–52, 75
Definition	43
Input	118
Rendering	118, 120–123
Truncating	119–120
Combining classes	51, 75–85
Combining half marks	181
Combining marks	179
Compatibility characters	17
Compatibility decomposition	44
Compatibility ideographs	267
Compatibility jamo	275
Compression	113
Conformance	30–32, 37–69
Bidirectional	67
Examples	30–32
Requirements	37–39
Conjoining jamo	52–55
Contextual forms	
Positioning	123
Control characters	29
Control codes	314
Control pictures	302

Control sequences 30
 Convertibility 18
 Currency symbols 297
 Cursive connection 316
 Cyrillic 171

D

Dashes 150
 Dead keys 118
 Decimal digits 89, 110
 Decomposition 44
 Default properties 74
 Deprecated format characters 320–321
 Design 3–4
 Design principles 12–18
 Devanagari 211–223
 Diacritical marks 27
 Diacritics 25, 179–180
 Combining 179
 Double 50, 179
 For symbols 180
 Latin script 160
 On i and j 163
 Polytonic Greek 169
 Spacing clones 161
 Digit names 191
 Dingbats 305
 Directional formatting 56
 Directionality 24
 Directionality (property) 42, 85–86
 Dynamic composition 18

E

East Asian scripts 257–280
 Enclosed alphanumerics 307
 Enclosing marks 50
 Encoding 11
 Encoding forms 19–21
 In ISO/IEC 10646 969
 Equivalence 136
 Equivalent sequence 18
 Ethiopic 284–286
 Euro sign 297
 European alphabetic scripts 159–181
 Extended BNF xxviii

F

Fancy text 15
 Format characters 313–327
 Format control characters 29, 134
 Fullwidth forms 273

G

General category 87–88
 Geometric shapes 304
 Geometrical symbols 304
 Georgian 173
 Glyphs 162, 332
 Grapheme 10
 Greek 167–170
 Gujarati 226
 Gurmukhi 225

H

Halfwidth forms 273
 Han characters 261
 Han ideograph arrangement 266
 Han radical-stroke index 849
 Han unification 961–962
 Language tags 115
 Principles 262–266
 Sources 259
 Handwriting sequence 118
 Hangul 275–277
 Hangul syllables 53–55, 276
 Collation 275
 Hebrew 186–188
 Hebrew alphabetic presentation forms 188
 Hiragana 272
 Hyphenation 10
 Hyphens 150

I

IANA charset names 21
 Identifier syntax 133–134
 Ideographic accounting numbers 97
 Ideographic Description Sequences (IDS) 269
 Ideographic numbers 97, 111
 Ideographic property 102
 Ideographic Variation Mark 270
 Ideographs 258–266, 335
 Component structure 264
 General characteristics 260
 Radicals 267
 Sorting 262
 Unknown or unavailable 155
 see also Han unification
 Implementation guidelines 105–143
 Informative properties 73–74
 International Phonetic Alphabet
 see IPA extensions
 Invalid code values 38
 IPA extensions 164
 ISO/IEC 10646 5, 967–972

J

Jamo 52–55, 275
 Short name (property) 86–87
 Joiner 190
 Jungseong 276
 Justification 122

K

Kanbun 267
 Kannada 234
 Katakana 273
 Kerning 123
 Khmer 251–253

L

Language information 114–116
 Language tagging 114
 Lao 239
 Latin script 160–166
 Layout control characters 29, 134, 315–319
 Letter (informative property) 102

- Letter spacing 316
 Letterlike symbols 298
 Liaison members 6
 Ligatures 27, 318
 Arabic 194
 Devanagari 221
 Latin 166
 Positioning 122
 Syriac 205
 Line breaking 315
 Line handling 113
 Line separator 29
 Logical order 16
- M**
 Malayalam 235
 Mandarin tone marks 278
 Mapping tables 107
 Mathematical operators 300
 Mathematical property 101
 Matras 212
 Middle Eastern scripts 185–206
 Minnan and Hakka tone marks 279
 Mirrored property 42, 97–101
 Miscellaneous symbols 305
 Modifier letters 177
 Mongolian 289–291
 Multiple base characters 27
 Myanmar 249–250
- N**
 Names 86–87, 101, 972
 New character proposals 963–965
 Newlines 113
 Noncharacter values 28
 Noncharacters 327
 Nongraphic characters 23
 Non-joiner 190
 Nonspacing marks 117–120, 120–123
 Positioning 122
 see Combining characters
 Normalization 112–113
 Normative properties 40, 73–74
 Notational conventions xxviii–xxx
 Number forms 299
 Numbers 110
 Ideographic 111
 Naming 190
 Numeric shape selectors 321
 Numeric value 89–96
- O**
 Ogham 176
 Optical character recognition 303
 Oriya 227
- P**
 Paired punctuation 149
 Paragraph separator 29
 Plain text 15
 Private use area 323
 Properties 42, 73–102, 111–112
 see also individual properties
- Proposals for new scripts and characters 963–965
 Punctuation 147–156
- Q**
 Quotation marks 151
- R**
 Radicals 267
 Yi 280
 Random access 133
 Reassigned characters 109
 Regular expressions 114
 Rendering 108, 118
 Replacement characters 29, 326
 Reserved characters 334
 Reserved codes 23
 Rich text 15
 Runic 174–175
- S**
 Searching 140
 Semantics 15, 40
 Separators 29
 Shift-JIS index 923
 Signature for Unicode data 28, 325
 Sinhala 236
 Small form variants 156
 Sorting 135–140
 South and Southeast Asian scripts 209–253
 Space characters 149
 Special areas 313–327
 Special character properties 42, 47–50
 Subscripts 299
 Subsets 31–32
 Superscripts 299
 Surrogates 21, 45, 322
 Handling surrogate pairs 109–110
 Symbols 295–309
 Symmetric swapping 320
 Syriac 199–205
 Cursive joining 203
 Ligatures 205
 Shaping 203
 Syriac abbreviation mark 200
- T**
 Table optimization 106
 Tables
 Multistage 106
 Two-stage 106
 Tamil 228–232
 Technical symbols 302
 Telugu 233
 Text elements 5, 10, 116
 Text handling 4–5
 Text processes 9–11
 Thaana 206
 Thai 237–238
 Tibetan 240–248
 Tilde 149
 Tone letters 178
 Transcoding 105–107
 Transformations 39, 45–47

Transmission
 7-bit and 8-bit107
 Truncation119

U

UCS

see ISO/IEC 10646

Unassigned characters108
 Unassigned codes23
 Unicode 1.0 names101
 Unicode Character Databasexxvii
 Unicode Consortium6
 Addressesxxx
 ISO membership6
 Liaison members6
 Membership6
 Unicode Consortium Resourcesxxx
 Anonymous FTP sitexxx
 Public mailing listxxx
 Web sitexxx
 Unicode Standard
 About the Unicode Standardxxv–xxviii
 Changes from version 2.0973–980
 Coverage2–3
 Design3–4
 Design principles12–18
 Future expansion2
 New characters3, 975
 Synchronization with ISO/IEC 10646971
 Technical reportsxxvii
 Versions32, 973–974
 Unicode Technical Committee (UTC)6
 Unification17
 see also Han unification
 Universal Character Set
 see ISO/IEC 10646
 Unknown characters108
 Unrenderable characters108
 UTF-1619, 970
 UTF-820, 970

V

Versions of the Unicode Standard32, 973–974
 Vietnamese letters and tone marks166
 Virama213
 Visual ambiguity17

W

Wide character type107–108
 Word breaking315
 Writing direction24

Y

Yi280
 Yiddish digraphs187

This PDF file is an excerpt from *The Unicode Standard, Version 3.0*, 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 (see <http://www.unicode.org/unicode/uni2errata/UnicodeErrata.html>). More recent versions of the Unicode standard exist (see <http://www.unicode.org/unicode/standard/versions/>).

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 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.

ISBN 0-201-61633-5

Copyright © 1991-2000 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.

This book is set in Minion, designed by Rob Slimbach at Adobe Systems, Inc. It was typeset using FrameMaker 5.5 running under Windows NT. ASMUS, Inc. created custom software for chart layout. The Han radical-stroke index was typeset by Apple Computer, Inc. The following companies and organizations supplied fonts:

Apple Computer, Inc.
Atelier Fluxus Virus
Beijing Zhong Yi (Zheng Code) Electronics Company
DecoType, Inc.
IBM Corporation
Monotype Typography, Inc.
Microsoft Corporation
Peking University Founder Group Corporation
Production First Software

Additional fonts were supplied by individuals as listed in the *Acknowledgments*.

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.

All other company and product names are trademarks or registered trademarks of the company or manufacturer, respectively.

The publisher offers discounts on this book when ordered in quantity for special sales. For more information please contact:

Corporate, Government, and Special Sales
Addison Wesley Longman, Inc.
One Jacob Way
Reading, Massachusetts 01867

Visit A-W on the Web: <http://www.awl.com/cseng/>

First printing, January 2000.