The Unicode® Standard
Version 15.0 – Core Specification

To learn about the latest version of the Unicode Standard, see https://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.

© 2022 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 https://www.unicode.org/reporting.html. For information about the Unicode terms of use, please see https://www.unicode.org/copyright.html.

The Unicode Standard / the Unicode Consortium; edited by the Unicode Consortium. — Version 15.0.
Includes index.
ISBN 978-1-936213-32-0 (https://www.unicode.org/versions/Unicode15.0.0/)
1. Unicode (Computer character set) I. Unicode Consortium.
QA268.U545 2022

ISBN 978-1-936213-32-0
Published in Mountain View, CA
September 2022

I Index

The index covers the contents of this core specification. To find topics in the Unicode Standard Annexes, Unicode Technical Standards, and Unicode Technical Reports, use the search feature on the Unicode website.

For definitions of terms used, see the glossary on the Unicode website. To find the code points for specific characters or the code ranges for particular scripts, use the Character Index on the Unicode website. (See *Appendix B.3, Other Unicode Online Resources.*)

A

abbreviation, Coptic 313
abjads 258, 365
abstract character sequences
    definition 88
abstract characters 29
    definition 88
abugidas 259, 260, 461, 661
accent marks *see* diacritics
accented characters
    encoding 12
    Latin 291
    normalization 206
accounting numbers, ideographic 176
acrophonic numerals 205, 310
Adlam 814–815
Aegean numbers 344
Africa
    scripts of 793–816
Afrikaans 296
Ahom 655–656
Ainu 768
Aiton 676
Alchemical Symbols 900
Algonquian 820
Ali Gali 555
aliases
    character name 86, 181
    informative 952
    normative 953
    property 162
    property value 162
allocation areas 44
allocation of encoded characters 43–51
Alphabetic (informative property) 189
alphabets 258
    European 289–339
    mathematical 853–858
alternate format characters (deprecated) ...193, 929–930
Americas
    scripts of 817–825
Amharic 794
Anatolian hieroglyphs 458–459
Ancient Symbols 904
angle brackets (U+2329 and U+232A)
    deprecated for technical publication 885
Annexes, Unicode Standard (UAX) .... xxiii, 977
    as components of Unicode Standard 77
    conformance 83
    list of 83
annotation characters 942–944
    use in plain text discouraged 943
ANSI/ISO C
    wchar_t and Unicode 200
apostrophe (U+0027) 275
Arabic 373–400
    digits 861
Arabic-Indic digits 377–378
    signs used with 380
ArabicShaping.txt 383, 388, 406
Aramaic 424, 461, 556, 587, 593
areas of the Unicode Standard 44
ARIB 895
Armenian 321–322
arrows 881–882
ASCII
    characters with multiple semantics 265
    transparency of UTF-8 37
    Unicode modeled on 1
    zero extension 200, 990
Assamese 489
assigned code points 11, 30
Athapascan 820
atomic character boundaries 218
Avestan 432–433

B

Balinese 715–720
Bamum 809–810
Bangla 489–495
base characters 329
  definition 105
  multiple 59
  ordered before combining marks 220, 329
Basic Multilingual Plane (BMP) 1, 43
  allocation areas 48
  representation in UTF-16 36
Basque 296
Bassa Vah 811
Batak 725–726
Baybayin 710
benefits of Unicode 1
Bengali 489–495
Bhaiksuki 599–600
Bidi Class (normative property) 171
Bidi Mirrored (normative property) 178
Bidi Mirroring Glyph (informative property) 179
BidiMirroring.txt 179
Bidirectional Algorithm, Unicode 52, 82
bidirectional ordering 20
  controls 926
bidirectional text 52, 82
  Middle Eastern scripts 365
  nonspacing marks in 223
  punctuation in 264
big-endian 39
  definition 81
Bihari 485
binary comparison and sort order
  caution for UTF-16 36
  UTF differences 231, 233
  UTF-8 38
block 44, 88, 257, 947
  headers 959
BMP *see* Basic Multilingual Plane
BNF (Backus-Naur Form) 971
BOCU-1 *see* UTN #6, BOCU-1
  MIME-Compatible Unicode Compression
Bodhi 544
Bodo 484
BOM (U+FEFF) 39, 66, 130–133, 940–942
Bopomofo 764–766
boundaries, text 60, 190, 217–218, 228
  see also UAX #14, Unicode Line Breaking Algorithm
  see also UAX #29, Unicode Text Segmentation
boustrophedon 52, 354
box drawing symbols 889
Brahmi 461, 587, 589–592, 593, 663
Braille 828–829
Breton 296
Buginese 713–714
Buhid 710
Bulgarian 315
bullets 279
  numeric 863
Burmese *see* Myanmar
Byelorussian 315
byte order mark (BOM) (U+FEFF) .39, 66, 130–133, 940–942
byte ordering
  changing 79
  conformance 81
byte serialization 39, 66
Byzantine Musical Symbols 836

C

C language
  wchar_t and Unicode 200
C0 and C1 control codes 31, 188, 914
Cambodian *see* Khmer
Canadian Aboriginal Syllabics 820–821
candrabindu 487, 627
canonical composite characters
  see canonical decomposable characters
canonical composition algorithm 138
canonical decomposable characters
  definition 117
canonical decomposition 62
  definition 116
  mappings 115
canonical equivalence
  definition 117
  nonspacing marks 225
canonical equivalent character sequences
  conformance 79
canonical mappings
  see canonical decomposition mappings
canonical ordering algorithm 137
canonical precomposed characters
  see canonical decomposable characters
Cantonese 747
capital letters 164, 236, 289
Carian 348
carriage return (U+000D) (CR) 209, 915
carriage return and line feed (CRLF) 209
case 297
  and text processes 12
  beyond ASCII 237
  camelcase 239
  case folding 240
  case operations (conformance) 83, 152–158
  case operations and normalization 242

case operations, reversibility 239
cased (definition) 153
case-insensitive comparison 157, 231, 240
casing context (definition) 153
conversion 154
detection 156
European alphabets 289
exceptional Latin pairs 293, 297
Georgian 324
lowercase 164, 236, 289
mapping tables 196
mappings 152, 166, 236–238
mappings noted in code charts 956
titlecase 164, 236
Turkish I 238, 293
uppercase 164, 236, 289
see also default case

Case (normative property) 164, 236
CaseFolding.txt 166, 240
caseless letters 297
Catalan 295
Caucasian Albanian 359
cedilla 292
CEF *see* character encoding forms
CES *see* character encoding schemes
Chakma 575
Cham 701–702
character encoding forms (CEF) 33–38, 990
see also Unicode encoding forms
character encoding model 33, 41
see also UTR #17, Unicode Character Encoding Model
character encoding schemes (CES) 39–42
see also Unicode encoding schemes
character encoding standards
coverage by Unicode 3
Character Index 978
character names 86, 180–187, 994
aliases 86, 181
conventions 968
for CJK ideographs 961
for control codes 185, 188
in code charts 952
matching 181
character properties
see properties
see also individual properties, e.g. Combining Class
character semantics 1, 78, 85–86, 995
as Unicode design principle 18
ASCII 265
definition 85
character sequences
abstract *see* abstract character sequences
canonical equivalent *see* canonical equivalent character sequences
compatibility equivalent *see* compatibility equivalent character sequences
conformance 79
named 181
character sequences, combining 105
character shaping selectors (deprecated) 929
character statistics 978
character tabulation (U+0009) 915
characters
abstract *see* abstract characters
arrangement in Unicode 45
assigned 11, 30
boundaries 217
canonical decomposable *see* canonical decomposable characters
classes 971
code charts 947–965
coded *see* encoded characters
combining *see* combining characters
compatibility decomposable *see* compatibility decomposable characters
composite *see* decomposable characters
concept of 15, 60
conformance definitions 88–91
confusable 245
conversion 196–197
decomposable *see* decomposable characters
deprecated *see* deprecated characters
encoded *see* encoded characters
encoding forms *see* encoding forms
encoding schemes *see* encoding schemes
end-user perceived 60
format control 30, 67, 266, 913–930
glyphs, relationship to 15
graphic 30
identity (definition) 85
ignored in processing 248–254
interpretation 78
layout control 67, 917–927
modification 79
names list 948–960
names *see* character names
not encoded in Unicode 3
number encoded in Version 14.0 3
precomposed *see* decomposable characters
properties *see* properties
semantics *see* character semantics
special 66, 913–945
supplementary *see* supplementary characters
transcoding 196–197
unsupported 201

characters, not glyphs
in spoofing 246
Unicode principle 15
charsets
IANA registered names 40
Cherokee 818
Chinese 746–747
Cantonese 747
Hakka 765
Mandarin 747
Minnan (Hokkien/Fujian, incl. Taiwanese) 765
simplified and traditional 746
Chorasmian 434
Chu hán 745
Chu Nôm 1002
citations for
properties 75
Unicode algorithms 76
Unicode Standard 74
CJK ideographs 261, 740–756
accounting numbers 176
CJK Compatibility Ideographs 755
CJK Compatibility Supplement 756
CJK Strokes 758, 1005
CJK Unified Ideographs 740–755
CJK Unified Ideographs Extension B 755
code charts 961
compatibility ideographs in Plane 2 51
component structure 750
encoding blocks 741
ideographic description sequences 760–763
ideographic variation mark (U+303E) 762
Kangxi radicals 754, 757–758
names 961
numbers 861
numeric values 176, 205
order of encoding 752
radicals 757–758
source standards 1004
unknown or unavailable 287
Vietnamese 738
CJK Miscellaneous Area 49
CJK punctuation and symbols 286
compatibility forms 288
overscores and underscores 288
quotation marks 273
sesame dots 287
vertical forms 288
CJK Unified Ideographs extensions 742–743
CJK-JRG (Chinese/Japanese/Korean Joint Research Group) 1000
CJKV Ideographs Area 49
cluster boundaries 217
code charts 947–965
representative glyphs 948
code point sequences
notation 969
code points 7, 29
assigned 11, 30
assignment 45
categories 30
default ignorable 201, 253
definition 88
designated 30
notation 968
number in Unicode Standard 1
private-use *see* private-use code points
reserved *see* reserved code points
semantics 32
surrogate *see* surrogates
unassigned *see* unassigned code points
undesignated 30
code positions *see* code points
code set independence 18
code unit sequences
definition 119
ill-formed (definition) 121
notation 969
well-formed (definition) 121
code units
definition 119
isolated 118
code values *see* code units
coded character representations
see coded character sequences
coded character sequences
definition 90
coded characters *see* encoded characters
codespace *see* Unicode codespace
coeng 677, 680
Collation Algorithm, Unicode (UCA) 12
collation *see* sorting
collation tables 196
combining character sequences 55, 105
defective 223
definition 107
Latin 291
line breaking 219
matching 219
order of base character and marks 220, 329
rendering 219
selection 217
truncation 220–221
combining characters 54–60, 109–114, 219–227
blocking reordering 924
canonical ordering 61, 137, 168

combining marks 329–330
definition 105
dependence 329
display order 57
keyboard input 220
ligatures 59
multiple 56
multiple base characters 59
normalization of 206
ordering conventions 55
rendering of marks 222–227
reordrant 169
script-specific 55
split 169
strikethrough 170
subjoined 170
typographical interaction 57, 168
vertical stacking 57
see also diacritics
Combining Class (normative property) 168
combining classes 135, 168, 225–226
class zero characters 168
definition 135
combining grapheme joiner (U+034F) 924
combining half marks 191, 337
combining marks *see* combining characters
comma below 292
Compatibility and Specials Area 26, 49
compatibility characters 22
compatibility composite characters 27
see compatibility decomposable characters
compatibility decomposable characters 26
definition 115
compatibility decomposition 62
definition 115
compatibility decomposition mappings 115
compatibility equivalence
definition 116
compatibility equivalent character sequences
conformance 79
compatibility mappings
see compatibility decomposition mappings
compatibility precomposed characters
see compatibility decomposable characters
compatibility variants 26
mapping 243
composite characters
see decomposable characters
Composition Exclusion (normative property) 99
compression 208
see also UTS #6, A Standard Compression Scheme for Unicode (SCSU)
conferences 978
conformance 71–158
definitions 85–91
examples 68
ISO/IEC 10646 implementations 995
requirements 77–82
confusables 245
conjunct consonants
Indic 217, 469
Myanmar 671
selection of clusters 217
contextual shaping
apostrophe 275
Arabic 373
not used for Hebrew final forms 368
Syriac 405
contour tones 327
control codes 31, 67, 914
graphics for 884
names 188
properties 915
semantics 32, 915
specified in Unicode 915
control sequences 914
conversion of characters 196–197
convertibility
as Unicode design principle 24
Coptic 309, 312–314
Coptic Epact numbers 866
corporate use subarea 935
corrigenda 74
CR (U+000D carriage return) 209, 915
Creative Commons 910
CRLF (carriage return and line feed) 209
Croatian 296
digraphs 296
culturally expected sorting 12, 230
Cuneiform
Old Persian 445
Sumero-Akkadian 440–443
Ugaritic 444
Cuneiform and Hieroglyphic Area 50
Cuneiform and Hieroglyphs 439–459
currency symbols block 847–850
currency symbols encoded in other blocks . . 848
currency symbols, other 849
dollar sign, form and usage 848
euro sign 849
lari sign 849
lira sign, compatibility usage 848
lira sign, Turkish 849
peso signs, usage 848
ruble sign 849
rupee signs, Indian, usage 849
yen and yuan signs, usage 848

cursive joining 919–923
  Arabic 382–390
  control characters for 192, 375–376, 559, 918
  Mandaic 414
  Mongolian 558–559
  N’Ko 805
  Phags-pa 606
  Syriac 405–409
  transparency 922
cursive scripts 365
Cypriot 346
  see also Linear B
Cypro-Minoan 347
Cyrillic 315–318
Czech 296

D

danda, in Devanagari block 483
Danish 295
dashes 268
Database, Unicode Character
  see Unicode Character Database (UCD)
dead consonants, Indic 466
dead keys 220
decomposable characters 62
  definition 115
  normalization of 206
decomposition 62, 115–117
  canonical *see* canonical decomposition
  compatibility *see* compatibility decomposition
  definition 115
  in normalization 206
  mapping, definition 115
  mappings noted in code charts 956
default case
  algorithms 83, 152–158
  conversion 154
  detection 156
  folding 155
default caseless matching 157
default grapheme clusters 217
  see also UAX #29, Unicode Text Segmentation
Default Ignorable Code Point (property) 253
default ignorable code points 201, 253
default property values
  definition 95
defective combining character sequences 223
  definition 108
dependent vowel signs
  Indic 465
  Khmer 682
  Philippine scripts 710
deprecated characters 72, 951
  alternate format 193, 929–930
  definition 90
Derived Age (property) 202
derived properties
  definition 102
DerivedCoreProperties.txt 153, 164, 253
DerivedNormalizationProps.txt 242
Deseret 823–825
design goals of Unicode 4
design principles of Unicode 14–24
designated code points 30
Devanagari 463–488
Dhivehi 535
diacritics 54, 329
  alternative glyphs 291, 329
  Czech 292
  display in isolation 60, 268, 330
  double 113, 191, 331
  German dialectology 335
  Greek 305–306, 309
  Latin 291–294
  Latvian 292
  mathematical 857
  on i and j 293
  rendering 222–227
  Slovak 292
  spacing clones of 327, 331
  symbol 54, 336
  see also combining characters
dictionary symbols 896
digit form names 378
digits 205
  Arabic 861
  Arabic-Indic 377–378
  compatibility 861
  decimal 175
  glyph variants 863
  hexadecimal 861
  Myanmar 861
  national shapes 930
  Shan 861
  superscript and subscript 862
  Tai Laing 861
  Tai Tham 861
digraphs 296, 299, 301
dingbats 898–899
directionality 20, 52
  East Asian scripts 738
  Middle Eastern scripts 365
  Mongolian 557
  musical symbols 831
  normative property 171
  Ogham 362

Old Italic ... 351
Philippine scripts ... 711
Runic ... 354
discussion list for Unicode ... 978
Dives Akuru ... 653–654
Dogra ... 658–659
Dogri ... 484
Domino Tiles ... 901
dotless i ... 238, 293
dotted circle
in code charts ... 106, 330
in fallback rendering ... 222
to indicate diacritic ... 54
to indicate vowel sign placement ... 56
double diacritics ... 113, 191, 331
Duployan ... 841–842
Dutch ... 295, 296
dynamic composition
as Unicode design principle ... 23
Dzongkha ... 544

E

East Asian scripts ... 737–789
writing direction ... 52
see also CJK ideographs
Eastern Arabic-Indic digits ... 377
EBCDIC
newline function ... 210
editing, text boundaries for ... 217–218
efficiency
as Unicode design principle ... 15
Egyptian hieroglyphs ... 446–455
format controls ... 448–455
Elbasan ... 358
ellipsis ... 277–278
Elymaic ... 435
e-mail discussion list for Unicode ... 978
emoji ... 893, 894, 978
animal symbols ... 897
charts ... 978
cultural symbols ... 897
zodiacal symbols ... 897
emoji modifiers ... 897
emoticons ... 898
Enclosed Alphanumerics ... 908
enclosing marks ... 337
definition ... 106
encoded characters ... 7, 29
allocation ... 43–51
definition ... 90
encoding form conversion
definition ... 126
encoding forms ... 33–38
ISO/IEC 10646 definitions ... 990
encoding forms, Unicode
see Unicode encoding forms
encoding model for Unicode characters ... 33, 41
see also UTR #17, Unicode Character Encoding Model
encoding schemes ... 39–42
encoding schemes, Unicode
see Unicode encoding schemes
endian ordering
see byte order mark (BOM) (U+FEFF)
end-user subarea ... 936
English ... 295
equivalent sequences ... 206
as Unicode design principle ... 23
case-insensitivity ... 231, 240
combining characters in matching ... 219
conformance ... 80
Hangul syllables ... 776
in sorting and searching ... 230
language-specific ... 117
security implications ... 245
see also canonical equivalence
see also compatibility equivalence
see also encoding forms, encoding schemes
errata ... xxv, 74, 980
escape sequences ... 914
not used in Unicode ... 1, 4
Esperanto ... 296
Estonian ... 296
Ethiopic ... 794–797
Etruscan ... 350
European scripts ... 289–339
ancient ... 341–363
eyelash-RA ... 475

F

fallback rendering ... 252
of nonspacing marks ... 222
FAQ (Frequently Asked Questions) ... 979
Faroese ... 295
Farsi ... 373, 375
featural syllabaries ... 259
FF (U+000C form feed) ... 209, 915
file separator (U+001C) ... 915
Finnish ... 295
Finno-Ugric Transcription (FUT)
see Uralic Phonetic Alphabet (UPA)
fixed-width Unicode encoding form (UTF-32) ..35, 123
flat tables ... 196
Flemish ... 295
fleurons ... 900

fonts
  and Unicode characters 16
  for mathematical alphabets 856–858
  style variation for symbols 845
form feed (U+000C) (FF) 209, 915
format control characters 30, 67, 266, 913–930
  deprecated 929–930
  prefixed 193, 333
  stateful 927
fraction characters 875
fraction slash (U+2044) 276, 871
French 296
Frisian 296
fullwidth forms in East Asian encodings 773
futhark 353

G

Garshuni 401
Ge’ez 794
General Category (normative property) 172
  list of values 172
general punctuation 264–288
General Scripts Area 49
geometrical symbols 889–892
Georgian 323–324
German 295
geta mark (U+3013) 287
Glagolitic 320
Glossary 979
glyph selection tables 196
glyphs 6, 15
  characters, relationship to 15
  diacritics alternative 291, 329
  Greek alternative 306–308
  Latin alternative 291
  mathematical alternative 877
  missing 253
  representative in code charts 948
  standardized variants 931
  symbols alternative 845
golden numbers 355
Gothic 357
Grantha 649–652
grapheme base 329
  definition 108
grapheme clusters 11, 60
  see also UAX #29, Unicode Text Segmentation
  default 217
  definition 108
grapheme extender
  definition 108
grapheme joiner, combining (U+034F) 924
graphic characters 30
Greek 305–310
  acrophonic numerals 205, 310
  alternative glyphs 306–308
  ancient musical notation 838–840
  editorial marks 282
  letters as symbols 306–308, 878
  see also Cypriot, Linear B
Greenlandic 296
group separator (U+001D) 915
guillemets 272
Gujarati 501–502
Gunjala Gondi 582–583
Gurmukhi 496–500

H

Hakka 765
halant 461
  see also virama
half marks, combining 191, 337
half-consonants, Indic 470
halfwidth forms in East Asian encodings 773
hamza 395–397
Han ideographs *see* CJK ideographs
Han unification 748–755
  and language tags 215
  history 999–1004
  language usage 745
  source separation rule 743, 749
  source standards 1004
hand symbols 897
Hangul Area 49
Hangul syllables 737, 774–777
  and combining marks 114
  canonical decomposition 144
  collation 776
  composition 146
  conjoining jamo 142–151
  equivalent sequences 776
  Hangul Compatibility Jamo 775
  Hangul Jamo 774–777
  Hangul Syllables block 776–777
  Johab set 776
  name generation 147
  normalization 775
  standard 143
Hangzhou numerals 870
Hanifi Rohingya 708
Hanja *see* CJK ideographs
Hanunóo 710
Hanzi *see* CJK ideographs
harakat 374
hasant 489
hash tables 197

Hatran 438
Hebrew 367–372
hentaigana 768–770
hieroglyphs
  Anatolian 458–459
  Egyptian 446–455
  Meroitic 456–457
high surrogate
  definition 118
  high-surrogate code points 77, 937
  high-surrogate code units 118
higher-level protocols
  definition 91
Hindi 463
Hiragana 767
horizontal tab (U+0009) 915
HTML newline function 210
Hungarian 296
hyphenation 918
  as a text process 10
hyphens 268, 918

I

I Ching symbols 903
IANA charset names 40
Icelandic 295
identifiers 229
  see also UAX #31, Unicode Identifier and Pattern Syntax
Ideographic (informative property) 189
ideographic description sequences 761
Ideographic Rapporteur Group (IRG) 1002
Ideographic Research Group (IRG) 1003
ideographs *see also* CJK ideographs
IICore 743, 1002
ill-formed
  definition 121
Imperial Aramaic 424–425
implementation guidelines 195–255
in a Unicode encoding form
  definition 122
India
  Official scripts 461–532
Indian rupee signs, usage 849
Indic scripts 461–532
  principles, in terms of Devanagari 464–474
  relation to ISCII standard 463
Indic Siyaq 868
Indonesian 295
industry character sets
  covered in Unicode 3
information separators (U+001C..U+001F) 915
informative properties
  definition 99
Inscriptional Pahlavi 430
Inscriptional Parthian 430
inside-out rule 222
interchange restrictions 31
International Phonetic Alphabet (IPA) 258, 298–299
  Spacing Modifier Letters 326
  see also phonetic alphabets
internationalization 18
Internationalization & Unicode Conference 978
Internet protocols
  UTF-8 as preferred encoding 37
Inuktitut 820
invisible operators 883
iota subscript 306
IPA *see* International Phonetic Alphabet
IRG (Ideographic Research Group) 1003
Irish 295, 362
ISCII standard and Unicode 463
ISO/IEC 10646 981–995
  conformance of Unicode implementations . . 995
  encoding forms 990
  synchrony with Unicode Standard 992
  timeline compared to Unicode versions 984
Italian 295
ITC Zapf Dingbats 898
IUC *see* Internationalization & Unicode Conference

J

jamos *see* Hangul syllables
Japanese 737
Japanese era names 909
Javanese 721–723
Jawi 397
jihvamuliya 488, 627
Johab 776
joiners 375
  combining grapheme joiner (U+034F) 924
  word joiner (U+2060) 917
  zero width joiner (U+200D) 375–376, 920
justification 224

K

Kaithi 624–626
Kana (Hiragana and Katakana) 767–768
Kanbun 756
Kangxi radicals 754, 757–758
Kanji *see* CJK ideographs
Kannada 520–523
Kashmiri 485
Katakana 767–768
Kawi 715, 732–735

Kayah Li 700

KC (normalization form)
  see Normalization Form KC

KD (normalization form)
  see Normalization Form KD

keytop labels 884

Khamti Shan 674

Kharoshthi 593–594

Khitan Small Script 790–791

Khmer 677–688
  characters not recommended 685
  syllable components, order of 686

Khojki 635–636

Khudawadi 638–639

killer 260
  Batak 725
  Brahmi 589
  Meetei Mayek 569
  Myanmar (asat) 672
  see also virama

Konkani 483

Kurdish 397

L

Ladino 367

language tags 215, 945
  and Han unification 215
  use strongly discouraged 945

Lanna 691

Lao 667–669

last-resort glyphs 253

Latin 291–304
  alternative glyphs 291
  Basic Latin 295
  encoding blocks 44
  IPA Extensions 298–299
  Latin Extended Additional 301–304
  Latin Extended-A 295
  Latin Extended-B 296–298
  Latin Extended-C 301
  Latin Extended-D 302
  Latin Extended-E 303
  Latin Ligatures 301
  Latin-1 Supplement 295
  Phonetic Extensions 300–304

Latin Extended-F 304

Latin Extended-G 304

Latvian 296, 303
  cedilla 292

layout control characters 67, 917–927

leading surrogates
  see high-surrogate code units

legibility criterion for plain text 19

Lepcha 576–578

letter spacing 918

letterlike symbols 851–858

LF (U+000A line feed) 209, 915

ligatures 919–923
  Arabic 385–387
  combining characters on 59
  control characters for 192
  for nonspacing marks 226
  Latin 301
  selection 218
  Syriac 409

Limbu 565–568

line breaking 209–213, 917–919
  control characters 191
  in South Asian scripts 665, 673, 688
  recommendations 211
  see also UAX #14, Unicode Line Breaking Algorithm

line feed (U+000A) (LF) 209, 915

line separator (U+2028) (LS) 209, 919

line tabulation (U+000B) (VT) 915

Linear A 343

Linear B 344–345
  see also Cypriot

linear boundaries 218

Lisu 782–784

Lithuanian 296

little-endian 39
  definition 81

logical order
  as Unicode design principle 19
  exceptions to 169

logograph 261

logosyllabaries 260

low surrogate
  definition 118
  low-surrogate code points 77, 937
  low-surrogate code units 118

lowercase 164, 236, 289

LS (U+2028 line separator) 209, 919

Lycian 348

Lydian 348

M

MacOS newline function 210

Mahajani 633–634

Mahjong Tiles 900

mail discussion list for Unicode 978

Maithili 484

major version 73

Makasar 730–731

Malay 295

Malay, Patani 666
Malayalam 524–532
  Suriyani 410, 525
Maltese 296
Manchu 556
Mandaic 413–415
Mandarin 747
Manden 802
Manichaean 426–429
map symbols 896
mapping tables *see* tables of character data
Marathi 463, 475, 482
Marchen 608
markup languages
  line breaking 209
Masaram Gondi 580–581
Mathematical (informative property) 875
mathematical expression format characters 193
  see also UTR #25, Unicode Support for Mathematics
mathematical symbols 875–882
  alphabets 853–858
  alphanumeric 852–858
  fonts 856–858
  format characters 883
  fragments for typesetting 885
  invisible operators 883
  operators 876–879
  standardized variants 882
MathML 879
matras 168, 465
Medefaidrin 816
Meetei Mayek 569–570
Mende Kikakui 812–813
Meroitic
  cursive 456–457
  hieroglyphs 456–457
Miao 785–786
Middle Eastern scripts 365–536
  ancient 417–438
Min 747
Minnan (Hokkien/Fujian, incl. Taiwanese) 765
minor version 73
minus sign 878
  commercial (U+2052) 280
mirrored property
  see Bidi Mirrored (normative property)
mirroring of paired punctuation 270
Miscellaneous Symbols 895
missing glyphs 253
Modi 644–646
modifier letters 325–328
Modifier Letters, Spacing 301
Mongolian 555–564, 601
  writing direction 557
moon symbols 895
Mro 571
Multani 640
multibyte encodings
  compared to UTF-8 37
multistage tables 196
musical symbols 830–840
  ancient Greek 838–840
  Balinese 719
  Byzantine 836
  directionality 831
  Gregorian 835
  Kievan 835
  Persian 835
  Western 830–835
Myanmar 670–676
  digits 861
  Myanmar Extended-A 674
  Myanmar Extended-B 674

N

N’Ko 802–806
Nabataean 436
Nag Mundari 637
named character sequences 181
names, character *see* character names
namespace 87
Nandinagari 647–648
NEL (U+0085 next line) 209, 915
Nepali 463
neutral directional characters 171
New Tai Lue 691–693
Newa 541–543
newline function (NLF) 210, 916
newline guidelines 209–213
next line (U+0085) (NEL) 209, 915
NFC (Normalization Form C) 61
NFD (Normalization Form D) 61
NFKC (Normalization Form KC) 61
NFKD (Normalization Form KD) 61
NLF (newline function) 210, 916
no-break space (U+00A0) 917
  base for diacritic in isolation 60, 268, 330
no-break space, narrow (U+202F) 562
noncharacter code points *see* noncharacters
noncharacters 31, 938
  conformance 77
  definition 91
  handling 80
  in code charts 951
  interchange restrictions 31
  semantics 32
  U+10FFFF (not a character code) 938

U+FDD0..U+FDEF 31, 938
U+FFFE (not a character code) 66, 939
U+FFFF (not a character code) 31, 938
nondecomposable characters 63
non-joiner, zero width (U+200C) 375–376, 921
nonlinear boundaries 218
non-overlap principle in Unicode encoding forms 33
nonspacing marks 329
definition 106
display in isolation 60, 268, 330
positioning 226
rendering 222–227
see also combining characters
see also diacritics
normalization 61, 206–207
and case operations 242
canonical ordering algorithm 61, 137, 168
conformance 82
of private-use characters 935
see also UAX #15, Unicode Normalization Forms
stability 134
Normalization Form C (NFC) 61
Normalization Form D (NFD) 61
Normalization Form KC (NFKC) 61
Normalization Form KD (NFKD) 61
normalization forms 134–141
definition 140
specification 136
normative behaviors
definition 85
normative properties
definition 97
list 98
may change 98
Norwegian 295
notational conventions 967–974
notational systems 262, 827–844
nukta 374, 399, 476
null (U+0000)
as Unicode string terminator 916
number forms
CJK ideographs 205
numbers
Coptic Epact 866
handling 205
ideographic accounting 176
numerals 859–872
acrophonic 310
Chinese counting rods 873
Coptic 314
Cuneiform 443
Ethiopic 796
Greek acrophonic 205
Hangzhou 870
Meroitic cursive 457
old-style 277
Roman 205, 875
Rumi 867
Suzhou-style 870
numeric separators 279
numeric shape selectors (deprecated) 930
Numeric Type (normative property) 175
Numeric Value (normative property) 175
numero sign (U+2116) 851
Nüshu 781
Nyiakeng Puachue Hmong 705–706

O

object replacement character (U+FFFC) 944
octet 970
Ogham 362
Ol Chiki 573–574
Old Church Slavonic 315
Old Hungarian 356
Old Italic 350–352
Old North Arabian 419
Old Permic 361
Old Persian 445
Old Sogdian 615
Old South Arabian 420–421
Old Turkic 614
Old Uyghur 617–618
old-style numerals 277
Oriya 503–506
ornamental dingbats 899
Oromo 794
Osage 822
Osmanya 798
Ottoman Siyaq 868
overlapping encodings 33
overscores 276

P

Pahawh Hmong 703–704
Pahlavi, Inscriptional 430
Pahlavi, Psalter 431
Palmyrene 437
Panjabi 496
paragraph or section marks 279
paragraph separator (U+2029) (PS) 209, 919
Parthian, Inscriptional 430
Pashto 373
Patani Malay 666
Pau Cin Hau 707
Persian 373, 375
Phags-pa 601–607
Phaistos Disc symbols 904

Phake 676
Philippine scripts 710–712
Phoenician 422
phonemes 261
phonetic alphabets 258
- IPA Extensions 298–299
- Phonetic Extensions 300–304
- Spacing Modifier Letters 326–328
- Uralic Phonetic Alphabet (UPA) 280, 300
- *see also* International Phonetic Alphabet (IPA)

Pinyin 295
pipeline table
- proposed new characters 979

pivot code, Unicode as 196
plain text
- as Unicode design principle 18
- legibility criterion 19

planes of Unicode codespace 43
- Plane 0 (BMP) 43
- Plane 1 (SMP) 43, 50
- Plane 14 (SSP) 44
- Plane 2 (SIP) 43, 51
- Plane 3 (TIP) 43, 51
- Planes 15-16 (Private Use) 51, 936

Playing Cards 901
points, Hebrew pronunciation marks 367
policies of the Unicode Consortium 979
Polish 296
Portuguese 295
precomposed characters
- *see* decomposable characters
- compatibility *see* compatibility decomposable characters

prefixed format control characters 193
prepended concatenation marks 254, 333
Private Use Area (PUA) 49, 935
Private Use planes 44, 51, 936
private-use characters
- properties 934
- semantics 32

private-use code points 31, 201
- conformance 78
- definition 104
- high surrogates 937

properties 18, 93–104, 159–194
- aliases 162
- aliases (definition) 103
- and Unicode algorithms 98
- data tables 196
- derived *see* derived properties
- in Unicode Character Database (UCD) 45
- informative *see* informative properties
- normative references to 75, 82
- normative *see* normative properties
- of control codes 915
- provisional *see* provisional properties
- simple *see* simple properties
- *see also individual properties, e.g.* combining classes

property values
- aliases 162
- aliases (definition) 103
- default 95
- default (definition) 95
- normative references to 82

PropertyAliases.txt 103, 972
PropertyValueAliases.txt 103, 972
PropList.txt 166
Provençal 296
provisional properties
- definition 100

PS (U+2029 paragraph separator) 209, 919
Psalter Pahlavi 431
PUA (Private Use Area) 49, 935
punctuation 264–288
- blocks containing 257
- CJK 286
- doubled 277
- ideographic 759
- in bidirectional text 264
- paired 270
- small form variants 288
- typographic forms 264
- vertical forms 288

Punctuation and Symbols Area 49
Punjabi 496

Q

quotation marks 271–275
- East Asian 274
- European 272

R

radicals, Kangxi and other CJK 757–758
radical-stroke index 754
record separator (U+001E) 915
recycling symbols 896
references 979
referencing 82
- properties 75
- Unicode algorithms 76
- Unicode Standard 74

regional indicator symbols 909
regular expressions 214
- and line breaking 209
- *see also* UTS #18, Unicode Regular Expressions

Rejang 724

rendering of text 6, 10, 17
  fallback 252
  unsupported characters 201
repertoire of abstract characters 29
reph 474, 478, 518
replacement character (U+FFFD) 42, 67, 81, 944
reserved code points 30, 201
  definition 91
  in code charts 951
  preservation in interchange 31
  see also unassigned code points
Rhaeto-Romanic 296
rich text 18
right single quotation mark (U+2019)
  preferred for apostrophe 275
right-to-left text 52
  East Asian scripts 738
  Middle Eastern scripts 365
roadmap for script additions 45, 979
Roman numerals 205, 875
Romanian 296
  comma below 293
Romany 296
Rong 576–578
Rumi numeral symbols 867
Runic 353–355
Russian 315

S

Samaritan 411–412
Sami 296
Sanskrit 463
Saurashtra 579
scalar values, Unicode
  see Unicode scalar values
scripts
  in Unicode Standard 3
  roadmap for future additions 45, 979
  types of 263
  see also UAX #24, Unicode Script Property
SCSU
  see UTS #6, A Standard Compression Scheme for Unicode
searching 230–232
  as a text process 10
  case-insensitive 231, 240
section or paragraph marks 279
security issues 245
self-synchronization of encoding forms 34
semantics
  see character semantics
sequences
  notation 969
Serbian
  corresponding digraphs in Croatian 296
Shan 689
  digits 861
Sharada 627–628
Shavian 363, 782
Show Hidden 79, 222, 252, 932
SHY (U+00AD soft hyphen) 918
Sibe 557
Siddham 631–632
signature for Unicode data 66, 940–942
simple properties
  definition 102
simplified Chinese 746
Sindhi 373, 483
Sinhala 537–540
Sinological dot 303
SIP (Supplementary Ideographic Plane) 43, 51
Siyaq Numbers 868
  Indic 868
slash, fraction (U+2044) 276
Slovak 296
Slovenian 296
small letters 164, 236, 289
SMP (Supplementary Multilingual Plane) 43, 50
soft hyphen (U+00AD) (SHY) 918
Sogdian 616
Somali 798
Sora Sompeng 657
Sorbian 296
sorting 12, 230
  and combining grapheme joiner 925
  as a text process 10
  case-insensitive 231
  culturally expected 12, 230
  language-insensitive 230
  see also Unicode Collation Algorithm (UCA)
source separation rule 743, 749
South and Central Asian scripts
  Ancient 587–616
  Other historic 619–659
  Other modern 533–583
South Asian scripts 461–568
Southeast Asian scripts
  Indonesia and the Philippines 709–735
  Mainland 661–708
Soyombo 612–613
space (U+0020)
  base for diacritic in isolation 60, 268, 330
space characters 267, 917–919
  graphics for 884
space, zero width (U+200B) 267
spacing clones of diacritics 327, 331
spacing marks 329

definition 107
Spacing Modifier Letters 326–328
Spanish 295
special characters 66, 913–945
SpecialCasing.txt 152, 166
Specials 940–944
spell-checking
as a text process 11
spellings, alternative
see equivalent sequences
spoofing 245
SSP (Supplementary Special-purpose Plane) 44
stability 101, 161
as Unicode design principle 23
stacked boundaries 217
stacking sequences
nondefault 57
standardized variants 560, 931
in the code charts 958
mathematical symbols 882
StandardizedVariants.txt 560, 882
standards coverage 3
starters 136
stateful encoding
not used in Unicode 4
paired format controls 927
string comparison 12
strings, Unicode 42, 120
null termination 916
strong directional characters 171
styled text 18
sublinear searching 232
subsets, supported 70
conformance 78
ISO/IEC 10646 specification for 993
substitution character
see replacement character
Sumero-Akkadian 440–443
Sundanese 727–729
superscripts 327
and subscripts 873
supplementary characters
in UTF-16 strings 42
tables for 197
Supplementary General Scripts Area 49
Supplementary Ideographic Plane (SIP) 43, 51
Supplementary Multilingual Plane (SMP) 43, 50
supplementary planes
representation in UTF-16 36
representation in UTF-8 37
Supplementary Private Use Areas 51, 936
Supplementary Special-purpose Plane (SSP) 44
supported subsets 70
conformance 78
supralineation 313
surrogate code points
see surrogates
surrogate pairs 36, 124
definition 118
processing 203–204
surrogates 31, 118, 937
interchange restrictions 31
isolated surrogates, handling 42
isolated surrogates, ill-formed 124
isolated surrogates, uninterpreted 118
support levels 203
Surrogates Area 49, 937
Sutton SignWriting 843–844
Suzhou-style numerals 870
svasti signs 551
Swahili 295
Swedish 295
syllabaries 259
alphabetic property 189
featural 259
Syloti Nagri 622–623
symbols 845–911
animal 897
appearance variation 845
arrows 881–882
box drawing 889
cultural 897
currency symbols block 847–850
dictionary 896
dingbats 898–899
emoji 893, 894, 909
Enclosed Alphanumerics 908
fragments for mathematical typesetting 885
game 897
gender 896
genealogical 897
geometrical 889–892
hand 897
Khmer lunar calendar 688
letterlike 851–858
map 896
mathematical 875–882
mathematical alphanumeric 852–858
miscellaneous 895
musical 830–840
numerals 859–872
recycling 896
regional indicator 909
technical 884–888
weather 895
zodiacal 897
symmetric swapping format characters 929
Syriac 401–410

T

tab (U+0009 character tabulation) 915
tab, vertical (U+000B) 209, 915
tables of character data 196–197
  optimization 197
  supplementary characters 197
tag characters 945
Tagalog 710
Tagbanwa 710
tags, language 215, 945
Tai Laing
  digits 861
Tai Le 689–690
Tai Tham 694–696
  digits 861
Tai Viet 697–699
Tai Xuan Jing symbols 903
Takri 629–630
Tamil 507–516
Tangsa 586
Tangut 787–789
  code charts 963
  components 788–789
  radicals 788
tashkil 374
tashkil, harakat, points 376
TCHAR in Win32 API 200
Technical Reports (UTR) 977
Technical Standards (UTS) xxiv, 977
  abstracts 978
technical symbols 884–888
Telugu 517–519
terminal emulation 846
Tertiary Ideographic Plane (TIP) 43, 51
text boundaries 60, 190, 217–218, 228
  see also UAX #14, Unicode Line Breaking Algorithm
  see also UAX #29, Unicode Text Boundaries
text elements 6, 10, 217
  boundaries 228
  for sorting 230
text processes 6, 10–13
text rendering 6, 10, 17
text selection, boundaries for 217–218
Thaana 535–536
Thai 663–666
Tibetan 544–554
Tifinagh 799
Tigre 794
tilde (U+007E) 280
TIP (Tertiary Ideographic Plane) 43, 51
Tirhuta 641–643
titlecase 164, 236
Todo 556
tone letters 327–328
tone marks
  Bopomofo spacing 764, 765
  Chinantec 328
  Chinese 328
  Tai Le 689
  Thai 663
  Vietnamese 294
Toto 585
traditional Chinese 746
traffic signs 896
trailing surrogates
  see low-surrogate code units
transcoding 196–197
  tables 196
Transport and Map Symbols 898
triangulation in transcoding 196
tries 196
truncation
  combining character sequences 220–221
  surrogates and 204
Turkish 296
  case mapping of I 238, 293
  cedilla 293
  lira sign 849
two-stage tables 197

U

U+10FFFF (not a character code) 938
U+FEFF (BOM) 940–942
U+FFFE (not a character code) 939
U+FFFF (not a character code) 938
UAX (Unicode Standard Annex) xxiii, 977
  as component of Unicode Standard 77
  conformance 83
  list of 83
UCA *see* Unicode Collation Algorithm and *see also* UTS #10, Unicode Collation Algorithm
UCD *see* Unicode Character Database
UCS (Universal Character Set)
  see ISO/IEC 10646
UCS-2 990
UCS-4 990
Ugaritic 444
Ukrainian 315
unassigned code points 30, 77, 201
  defined as reserved code points 91
  handling 72
  properties of 95
  semantics 77
  see also reserved code points
underscores 276
undesignated code points 30

Unicode 1.0 Name (informative property) 188
Unicode algorithms
  and properties 98
  conformance 82
  definition 91
  normative references to 76, 82
Unicode Bidirectional Algorithm 21, 52
  see also UAX #9, Unicode Bidirectional Algorithm
Unicode Character Database (UCD) . . xxii, 161, 979
  as component of Unicode Standard 77
  changes 72
  properties in 45
Unicode character encoding model 33, 41
  see also UTR #17, Unicode Character Encoding Model
Unicode codespace
  definition 88
  planes 43
  size 1, 29
Unicode Collation Algorithm (UCA) 12
Unicode conferences 978
Unicode Consortium 976
  addresses 980
  Consortium membership in standards bodies 976
  e-mail discussion list 978
  membership 976
  policies 979
  website 978
Unicode data signature 66, 940–942
Unicode data types 199–200
  for C 199–200
Unicode encoding forms 119–126
  conformance 34, 80
  definition 120
  fixed-width (UTF-32) 35, 123
  signatures 941, 942
  variable-width 36, 37, 124
  see also encoding forms
Unicode encoding schemes
  conformance 130–133
  definition 130
  endian ordering 39
  see also encoding schemes
Unicode scalar values
  definition 119
Unicode security 245
  see also UTS #39, Unicode Security Mechanisms
Unicode Standard
  allocation of encoded characters 43–51
  architecture 10–13
  areas 44
  benefits 1
  blocks 257
  code charts 947–965
  components 77
  conformance 71–158
  conformance of ISO/IEC 10646 implementations 995
  corrections 74
  definitions for conformance 85–91
  design goals 4
  design principles 14–24
  errata 74, 980
  normative references to 74, 82
  number of characters 3
  number of code points 1, 29
  script coverage 3
  security issues 245
  synchrony with ISO/IEC 10646 992
  updates 980
  versions *see* versions of the Unicode Standard
Unicode Standard Annexes (UAX) . . . . xxiii, 977
  as components of Unicode Standard 77
  conformance 83
  list of 83
Unicode strings 42
  definition 120
Unicode Technical Committee (UTC) 976
Unicode Technical Reports (UTR) 977
Unicode Technical Standards (UTS) . . . . . . .xxiv, 977
  abstracts 978
UnicodeData.txt 152, 166
unification
  as Unicode design principle 21
  see also Han unification
Unified Repertoire and Ordering (URO) . .749, 1001
  see also Han unification
Unihan Database 161, 753, 754, 979, 1002
Unihan.zip 100, 161
UnihanCore2020 . . . . .744, 1002
unit separator (U+001F) 915
Universal Character Set (UCS)
  see ISO/IEC 10646
universality
  as Unicode design principle 14
Unix
  newline function 210
  UTF-8 in 18
unsupported characters 201
upadhmaniya 488, 627
update version 73
uppercase 164, 236, 289
Uralic Phonetic Alphabet (UPA) 280, 300
Urdu 373
URO (Unified Repertoire and Ordering) . .749, 1001
  see also Han unification

UTF, Unicode Transformation Formats 33, 120
as encoding form or scheme 133
binary comparison and sort order differences ...
231, ...233
in APIs 200
UTF-16 .... 36, 124, 991
binary comparison and sort order caution 36
bit distribution (table) 124
BOM in 131, 940
encoding form (definition) 124
encoding scheme (definition) 131
encoding schemes 39
in ISO/IEC 10646 991
in UTF-8 order 234
surrogates and string handling 42, 203
UTF-16BE (Big-endian) 941
encoding scheme 40
encoding scheme (definition) 130
UTF-16LE (Little-endian) 941
encoding scheme 40
encoding scheme (definition) 130
UTF-32 ....35, 123
BOM in 132
encoding form (definition) 123
encoding scheme (definition) 132
encoding schemes 39
UTF-32BE (Big-endian)
encoding scheme 40
encoding scheme (definition) 132
UTF-32LE (Little-endian)
encoding scheme 40
encoding scheme (definition) 132
UTF-8 .... 37, 124, 991
ASCII transparency 37
binary comparison and sort order 38
bit distribution (table) 125
BOM in 130, 133, 941
byte ranges 125
compared to multibyte encodings 37
encoding form (definition) 124
encoding scheme 39
encoding scheme (definition) 130
in Unix 18
in UTF-16 order 233
non-shortest form is invalid 124, 245
preferred encoding for Internet protocols 37
security and 245
signature 130, 133, 941
UTR (Unicode Technical Report) 977
UTS (Unicode Technical Standard) .... xxiv, 977
abstracts 978
Uyghur 373, 601

V

Vai 807–808
valid (synonym for well-formed) 122
variable-width Unicode encoding form ...36, 37, 124
variants
compatibility 26
fullwidth and halfwidth 288
mathematical symbols 882
small form 288
standardized 931
variation selectors 194, 931
ideographic variation mark (U+303E) 762
Mongolian free variation selectors 560
variation sequences 931
for Phags-pa 605–607
Version 14.0 77
number of characters 3
versions of the Unicode Standard ..xxiii, 72, 980, 997–998
backward compatibility 72
compared to ISO/IEC 10646 editions 997
content 73
interaction in implementations 202
numbering 73
property changes 72
stability 72
updates 980
vertical tab (U+000B) 209, 915
vertical text 52, 265, 288
East Asian scripts 738
Mongolian 557
Vietnamese 294, 301
ideographs 738
virama 260, 461
definition 466
Kharoshthi 597
Khmer 680
Myanmar 671
Philippine scripts 711
virama-like characters 192
visual order used for Thai and Lao 21
Vithkuqi 360
vowel harmony
Mongolian 561
vowel marks, Middle Eastern scripts 366
vowel separator
Mongolian 562
vowel signs
Indic 56, 465
Khmer 682
Philippine scripts 710

W

Wancho 584
Warang Citi 572
wchar_t
    in C language 200
weak directional characters 171
weather symbols 895
website, Unicode Consortium 978
Weierstrass elliptic function symbol 852
well-formed
    definition 121
Welsh 296
Where Is My Character? 980
wide characters
    data type in C 200
wiggly fence (U+29DB) 880
Windows newline function 210
word breaks 219, 917–919
    in South Asian scripts 665, 673, 688
word joiner (U+2060) 917
writing direction *see* directionality
writing systems 258–263
Wu (Shanghainese) 747

X

Xibe 557
Xishuangbanna Dai 691

Y

Yezidi 416
Yi 778–780
Yiddish 367
Yijing Hexagram Symbols 903
ypogegrammeni 306

Z

Zanabazar Square 609–611
Zapf Dingbats 898
zero extension relation among encodings 990
zero width joiner (U+200D) 375–376, 920
zero width no-break space (U+FEFF) ... 66, 82, 917
    initial 133, 941
zero width non-joiner (U+200C) 375–376, 921
zero width space (U+200B) 918
    for word breaks in South Asian scripts . . 665, 673, 688
zero-width space characters 918
Znamenny Musical Notation 837
ZWJ *see* zero width joiner (U+200D)
ZWNBSP *see* zero width no-break space (U+FEFF)
ZWNJ *see* zero width non-joiner (U+200C)
ZWSP *see* zero width space (U+200B)