

General Index

The general index is a guide to the written text only. To find the range for a particular script, see the table on the end-papers. To find the code for a specific character, see the preceding Character Names Index. For the definition of specific terms, see the Glossary.

- 7-bit transmission 625
- Adding characters 631–33
- Alternate spellings 17–18, 625
- ANSI C wide character type 624
- Arabic glyphs 121–22
- Architecture 7 ff.
- Backing store 16–17
- Bidirectionality 16–17, 77–78, 611–21
- Braille 3
- Burmese 636
- Byte Order Mark (BOM) 23, 123, 623
- Canonical forms and canonical mappings 411 ff., 625
- Case pairs and case mapping 33, 34, 37, 440 ff.
- Character order 16–21, 53–57
- Character properties 15, 389–409
- Character vs. Glyph 12–13
- Characters, Unknown or Missing 629–30
- Chinese/Japanese/Korean ideographs 110 ff.
 - Ordering 116
 - Selection 114–16
- Chinese/Japanese/Korean phonetics and symbols 98 ff.
- Compatibility blocks/zone 120–22
 - Mapping to canonical forms 420 ff.
- Codespace allocation 3, 13–14, 27–28, 74, 98, 110, 118, 120
- Composite characters *see* Precomposed characters
- Compression 624–25
- Conformance 2, 21–25, 620–21
- Consortium 5
- Control characters 15–16, 29, 32, 88, 624
- Conventions xv–xvii
- Conversion to other codes 12, 465 ff.
- Coptic 43
- Coverage 2–3
- Czech 33
- Dashes 76, 401
- Decimal digits 51, 391–94, 630
- Decompositions 412 ff.
- Diacritics 30, 32, 39, 81, 635; *see also* Non-spacing marks
- Digits 51–52, 57, 395–96, 630
- Digraphs
 - Mapping to canonical forms 419
- Directionality 9, 16–17, 407–9, 611–21, 629
- East Asian *see* Chinese/Japanese/Korean
- Ethiopian 635–36
- Escape characters 624
- Fancy text 9–10
- Future development 4–5, 631–36
- General scripts 28
- Glagolitic 45
- Glyphs 12–13, 121, 630
- Indic scripts 53–59
- Interchange (Conformance requirement) 21–23
- International standards
 - Mapping from Unicode characters 466 ff.
- Interpretation (Conformance requirement) 23
- Joiner 50, 77
- Kanbun marks 104
- Khmer 636
- Ligatures
 - Mapping to canonical forms 419
- Line breaking 402
- Line separator 16, 78, 624
- Logical order 16–17
- Mapping tables 411 ff.
- Missing characters 629–30
- Modification (Conformance requirement) 23
- Mongolian 636
- National standards
 - Mapping from Unicode characters 560 ff.
- New characters 631–33
- Non-conformance 24–25
- Non-Joiner 50, 55, 77
- Non-spacing marks 18–21, 37, 40, 42, 47, 51, 56–57, 403–6, 627–28
- Notations xv–xvii
- Numbers 51, 390–99, 630
- Paragraph separator 16, 78, 624
- Pinyin 34
- Plain text 9–10
- Polytonic Greek 20, 42
 - Mapping to canonical forms 436 ff.
- Precomposed characters
 - Mapping to canonical forms 412 ff.

Principles 10–12
Properties of characters 15, 389–409
Private use characters 118–19, 629–30, 632–33
Punctuation pairs 30
Rendering 9, 16, 23–24, 53–57, 630
Right-to-left scripts *see* Bidirectionality
Sanskrit 53, 68, 71
Semantics 22, 30
Serbo-Croatian 34
Sinhala 636
Sorting 9, 626–27
Sources 3, 645–49
Space characters 75–76, 400

Standards
 Mapping from Unicode characters 465 ff.
Statistics 4
Subsets of characters 24, 626
Symbols 74 ff., 99
Technical Committee (UTC) 5
Text processes 7–9
Tone marks, Chinese 39, 99, 102
Tone marks, Vietnamese 20, 41, 432
Transmission 21–24, 625
Unicode Consortium 5
Unicode Technical Committee (UTC) 5

Unification 11, 36, 114–16
Unknown characters 629–30
Vendor standards
 Mapping from Unicode characters 494 ff.
Vietnamese 20, 41
 Mapping to canonical forms 432 ff.
Wide character type (ANSI C) 624
Width variants 122
Zero-width characters 76–77
Zhuyin–Fuhao 102
Zones 3, 13–14

Quick Guide to Unicode Character Blocks

<i>Character Block Name</i>	<i>Unicode Range</i>	<i>Description</i>	<i>Chart</i>
ASCII	U+0000 → U+007F	29	172
Latin1	U+0080 → U+00FF	32	176
European Latin	U+0100 → U+017F	33	180
Extended Latin	U+0180 → U+01FF	34	184
Standard Phonetic	U+0250 → U+02AF	36	188
Modifier Letters	U+02B0 → U+02FF	38	192
Generic Diacritical Marks	U+0300 → U+036F	40	196
Greek	U+0370 → U+03FF	42	200
Cyrillic	U+0400 → U+04FF	44	204
Armenian	U+0530 → U+058F	46	210
Hebrew	U+0590 → U+05FF	47	214
Arabic	U+0600 → U+06FF	49	218
Devanagari	U+0900 → U+097F	53	226
Bengali	U+0980 → U+09FF	60	230
Gurmukhi	U+0A00 → U+0A7F	61	234
Gujarati	U+0A80 → U+0AFF	62	238
Oriya	U+0B00 → U+0B7F	63	242
Tamil	U+0B80 → U+0BFF	64	246
Telugu	U+0C00 → U+0C7F	65	250
Kannada	U+0C80 → U+0CFF	66	254
Malayalam	U+0D00 → U+0D7F	67	258
Thai	U+0E00 → U+0E7F	68	262
Lao	U+0E80 → U+0EFF	70	266
Tibetan	U+1000 → U+105F	71	270
Georgian	U+10A0 → U+10FF	73	274
General Punctuation	U+2000 → U+206F	75	278
Superscripts and Subscripts	U+2070 → U+209F	79	282
Currency Symbols	U+20A0 → U+20CF	80	284
Diacritical Marks for Symbols	U+20D0 → U+20FF	81	286
Letterlike Symbols	U+2100 → U+214F	82	288
Number Forms	U+2150 → U+218F	83	292
Arrows	U+2190 → U+21FF	84	294
Mathematical Operators	U+2200 → U+22FF	85	298

<i>Character Block Name</i>	<i>Unicode Range</i>	<i>Description</i>	<i>Chart</i>
Miscellaneous Technical	U+2300 → U+23FF	87	304
Pictures for Control Codes	U+2400 → U+243F	88	306
Optical Character Recognition	U+2440 → U+245F	89	308
Enclosed Alphanumerics	U+2460 → U+24FF	90	310
Form and Chart Components	U+2500 → U+257F	91	314
Blocks	U+2580 → U+259F	92	318
Geometric Shapes	U+25A0 → U+25FF	93	320
Miscellaneous Dingbats	U+2600 → U+26FF	94	324
Zapf Dingbats	U+2700 → U+27BF	96	328
CJK Symbols and Punctuation	U+3000 → U+303F	99	332
Hiragana	U+3040 → U+309F	100	336
Katakana	U+30A0 → U+30FF	101	340
Bopomofo	U+3100 → U+312F	102	344
Hangul Elements	U+3130 → U+318F	103	346
CJK Miscellaneous	U+3190 → U+31FF	104	350
Enclosed CJK Letters and Ideographs	U+3200 → U+32FF	105	352
CJK Squared Words	U+3300 → U+337F	106	358
CJK Squared Abbreviations	U+3380 → U+33FF	107	362
Korean Hangul Syllables	U+3400 → U+3D2F	108	157
Chinese/Japanese/Korean Ideographs	U+4000 → U+8BFF	111	Vol. 2
Private Use Area	U+E800 → U+FDFE	119	—
Compatibility Zone	U+FE00 → U+FFEF	121	
CNS 11643 Compatibility	U+FE30 → U+FE4F	121	372
Small Variants	U+FE50 → U+FE6F	121	374
Basic Glyphs for Arabic Language	U+FE70 → U+FEFE	121	376
Halfwidth and Fullwidth Variants	U+FF00 → U+FFEF	121	380
Special	U+FFF0 → U+FFFD	123	386

