

# Package ‘rebus.unicode’

January 3, 2017

**Type** Package

**Title** Unicode Extensions for the 'rebus' Package

**Version** 0.0-2

**Date** 2017-01-02

**Author** Richard Cotton [aut, cre]

**Maintainer** Richard Cotton <richierocks@gmail.com>

**Description** Build regular expressions piece by piece using human readable code.  
This package contains Unicode functionality, and is primarily intended to be used by package developers.

**Depends** R (>= 3.1.0)

**Imports** rebus.base (>= 0.0-2)

**Suggests** stringi

**License** Unlimited

**LazyLoad** yes

**LazyData** yes

**Acknowledgments** Development of this package was partially funded by the Proteomics Core at Weill Cornell Medical College in Qatar <<http://qatar-weill.cornell.edu>>. The Core is supported by 'Biomedical Research Program' funds, a program funded by Qatar Foundation.

**Collate** 'imports.R' 'unicode-classes.R' 'unicode-constants.R'  
'unicode-general-category-classes.R'  
'unicode-general-category-constants.R' 'unicode-operators.R'

**RoxygenNote** 5.0.1

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2017-01-03 07:42:23

**R topics documented:**

ugc_cased_letter . . . . .	2
Unicode . . . . .	6
UnicodeOperators . . . . .	29
up_alphabetic . . . . .	31

<b>Index</b>	<b>36</b>
--------------	-----------

---

ugc_cased_letter	<i>Unicode General Categories</i>
------------------	-----------------------------------

---

**Description**

Match a Unicode General Category.

**Usage**

```
ugc_cased_letter(lo, hi, char_class = TRUE)
ugc_close_punctuation(lo, hi, char_class = TRUE)
ugc_connector_punctuation(lo, hi, char_class = TRUE)
ugc_control(lo, hi, char_class = TRUE)
ugc_currency_symbol(lo, hi, char_class = TRUE)
ugc_dash_punctuation(lo, hi, char_class = TRUE)
ugc_decimal_number(lo, hi, char_class = TRUE)
ugc_enclosing_mark(lo, hi, char_class = TRUE)
ugc_final_punctuation(lo, hi, char_class = TRUE)
ugc_format_control(lo, hi, char_class = TRUE)
ugc_initial_punctuation(lo, hi, char_class = TRUE)
ugc_letter(lo, hi, char_class = TRUE)
ugc_letter_number(lo, hi, char_class = TRUE)
ugc_line_separator(lo, hi, char_class = TRUE)
ugc_lowercase_letter(lo, hi, char_class = TRUE)
```

ugc\_mark(lo, hi, char\_class = TRUE)  
ugc\_math\_symbol(lo, hi, char\_class = TRUE)  
ugc\_modifier\_letter(lo, hi, char\_class = TRUE)  
ugc\_modifier\_symbol(lo, hi, char\_class = TRUE)  
ugc\_nonspacing\_mark(lo, hi, char\_class = TRUE)  
ugc\_number(lo, hi, char\_class = TRUE)  
ugc\_open\_punctuation(lo, hi, char\_class = TRUE)  
ugc\_other(lo, hi, char\_class = TRUE)  
ugc\_other\_letter(lo, hi, char\_class = TRUE)  
ugc\_other\_number(lo, hi, char\_class = TRUE)  
ugc\_other\_punctuation(lo, hi, char\_class = TRUE)  
ugc\_other\_symbol(lo, hi, char\_class = TRUE)  
ugc\_paragraph\_separator(lo, hi, char\_class = TRUE)  
ugc\_private\_use\_control(lo, hi, char\_class = TRUE)  
ugc\_punctuation(lo, hi, char\_class = TRUE)  
ugc\_separator(lo, hi, char\_class = TRUE)  
ugc\_space\_separator(lo, hi, char\_class = TRUE)  
ugc\_spacing\_mark(lo, hi, char\_class = TRUE)  
ugc\_surrogate\_control(lo, hi, char\_class = TRUE)  
ugc\_symbol(lo, hi, char\_class = TRUE)  
ugc\_titlecase\_letter(lo, hi, char\_class = TRUE)  
ugc\_unassigned\_control(lo, hi, char\_class = TRUE)  
ugc\_uppercase\_letter(lo, hi, char\_class = TRUE)  
UGC\_UPPERCASE\_LETTER

UGC\_LOWERCASE\_LETTER

UGC\_TITLECASE\_LETTER

UGC\_CASED\_LETTER

UGC\_MODIFIER\_LETTER

UGC\_OTHER\_LETTER

UGC\_LETTER

UGC\_NONSPACING\_MARK

UGC\_SPACING\_MARK

UGC\_ENCLOSING\_MARK

UGC\_MARK

UGC\_DECIMAL\_NUMBER

UGC\_LETTER\_NUMBER

UGC\_OTHER\_NUMBER

UGC\_NUMBER

UGC\_CONNECTOR\_PUNCTUATION

UGC\_DASH\_PUNCTUATION

UGC\_OPEN\_PUNCTUATION

UGC\_CLOSE\_PUNCTUATION

UGC\_INITIAL\_PUNCTUATION

UGC\_FINAL\_PUNCTUATION

UGC\_OTHER\_PUNCTUATION

UGC\_PUNCTUATION

UGC\_MATH\_SYMBOL

UGC\_CURRENCY\_SYMBOL

UGC\_MODIFIER\_SYMBOL  
UGC\_OTHER\_SYMBOL  
UGC\_SYMBOL  
UGC\_SPACE\_SEPARATOR  
UGC\_LINE\_SEPARATOR  
UGC\_PARAGRAPH\_SEPARATOR  
UGC\_SEPARATOR  
UGC\_CONTROL  
UGC\_FORMAT\_CONTROL  
UGC\_SURROGATE\_CONTROL  
UGC\_PRIVATE\_USE\_CONTROL  
UGC\_UNASSIGNED\_CONTROL  
UGC\_OTHER

### **Arguments**

lo	A non-negative integer. Minimum number of repeats, when grouped.
hi	positive integer. Maximum number of repeats, when grouped.
char_class	TRUE or FALSE. Should the values be wrapped into a character class?

### **Format**

An object of class `regex` (inherits from `character`) of length 1.

### **Value**

A character vector representing part or all of a regular expression.

### **References**

Table 12 of the Unicode Standard Annex #44 defines the Unicode General Categories. <http://www.unicode.org/reports/tr44>

You can see which characters are contained in a category by visiting, e.g., <http://www.fileformat.info/info/unicode/category/Nd/list.htm>

**See Also**

[unicode\\_property](#), [Unicode](#)

**Examples**

```
# Classes
ugc_lowercase_letter()
ugc_decimal_number()
ugc_paragraph_separator()
ugc_currency_symbol()

# With repetition
ugc_nonspacing_mark(3, 6)
ugc_separator(1, Inf)
ugc_dash_punctuation(0, Inf)

# Without a class wrapper
ugc_titlecase_letter(char_class = FALSE)

# Constants
UGC_UPPERCASE_LETTER
UGC_LETTER_NUMBER
UGC_MATH_SYMBOL
UGC_FORMAT_CONTROL

## Not run:
# All the Unicode general categories.
# Not run, since it generates lots of output
ls("package:rebus.unicode", pattern = "^ugc")

## End(Not run)

# Usage
library(rebus.base)
x <- "I exchanged $1000 for \u20ac665.41 and \u00a3243.13."
(rx <- capture(ugc_currency_symbol()) %R%
  capture(
    ugc_decimal_number(1, Inf) %R%
    optional(group(".", %R% ugc_decimal_number(2)))
  )
)
stringi::stri_match_all_regex(x, rx)
```

---

Unicode

*Unicode classes*

---

**Description**

Match ranges of unicode characters. In particular, you can match characters from a particular language.

**Usage**

```
armenian(lo, hi, char_class = TRUE)
armenian_ligatures(lo, hi, char_class = TRUE)
caucasian_albanian(lo, hi, char_class = TRUE)
cypriot_syllabary(lo, hi, char_class = TRUE)
cyrillic(lo, hi, char_class = TRUE)
cyrillic_supplement(lo, hi, char_class = TRUE)
cyrillic_extended_a(lo, hi, char_class = TRUE)
cyrillic_extended_b(lo, hi, char_class = TRUE)
elbasan(lo, hi, char_class = TRUE)
georgian(lo, hi, char_class = TRUE)
georgian_supplement(lo, hi, char_class = TRUE)
glagolitic(lo, hi, char_class = TRUE)
gothic(lo, hi, char_class = TRUE)
greek_and_coptic(lo, hi, char_class = TRUE)
greek_extended(lo, hi, char_class = TRUE)
latin(lo, hi, char_class = TRUE)
latin_1_supplement(lo, hi, char_class = TRUE)
latin_extended_a(lo, hi, char_class = TRUE)
latin_extended_b(lo, hi, char_class = TRUE)
latin_extended_c(lo, hi, char_class = TRUE)
latin_extended_d(lo, hi, char_class = TRUE)
latin_extended_e(lo, hi, char_class = TRUE)
latin_extended_additional(lo, hi, char_class = TRUE)
latin_ligatures(lo, hi, char_class = TRUE)
```

linear\_a(lo, hi, char\_class = TRUE)  
linear\_b\_syllabary(lo, hi, char\_class = TRUE)  
linear\_b\_ideograms(lo, hi, char\_class = TRUE)  
ogham(lo, hi, char\_class = TRUE)  
old\_italic(lo, hi, char\_class = TRUE)  
old\_permic(lo, hi, char\_class = TRUE)  
phaistos\_disc(lo, hi, char\_class = TRUE)  
runic(lo, hi, char\_class = TRUE)  
shavian(lo, hi, char\_class = TRUE)  
duployan(lo, hi, char\_class = TRUE)  
shorthand\_format\_controls(lo, hi, char\_class = TRUE)  
ipa\_extensions(lo, hi, char\_class = TRUE)  
phonetic\_extensions(lo, hi, char\_class = TRUE)  
phonetic\_extensions\_supplement(lo, hi, char\_class = TRUE)  
modifier\_tone\_letters(lo, hi, char\_class = TRUE)  
spacing\_modifier\_letters(lo, hi, char\_class = TRUE)  
superscripts\_and\_subscripts(lo, hi, char\_class = TRUE)  
combining\_diacritic\_marks(lo, hi, char\_class = TRUE)  
combining\_diacritic\_supplement(lo, hi, char\_class = TRUE)  
combining\_diacritic\_extended(lo, hi, char\_class = TRUE)  
combining\_half\_marks(lo, hi, char\_class = TRUE)  
bamun(lo, hi, char\_class = TRUE)  
bamun\_supplement(lo, hi, char\_class = TRUE)  
bassa\_vah(lo, hi, char\_class = TRUE)



coptic(lo, hi, char\_class = TRUE)  
coptic\_epact\_numbers(lo, hi, char\_class = TRUE)  
egyptian\_hieroglyphs(lo, hi, char\_class = TRUE)  
ethiopic(lo, hi, char\_class = TRUE)  
ethiopic\_supplement(lo, hi, char\_class = TRUE)  
ethiopic\_extended(lo, hi, char\_class = TRUE)  
ethiopic\_extended\_a(lo, hi, char\_class = TRUE)  
mende\_kikakui(lo, hi, char\_class = TRUE)  
meroitic\_cursive(lo, hi, char\_class = TRUE)  
meroitic\_hieroglyphs(lo, hi, char\_class = TRUE)  
nko(lo, hi, char\_class = TRUE)  
osmanya(lo, hi, char\_class = TRUE)  
tifinagh(lo, hi, char\_class = TRUE)  
vai(lo, hi, char\_class = TRUE)  
arabic(lo, hi, char\_class = TRUE)  
arabic\_supplement(lo, hi, char\_class = TRUE)  
arabic\_extended\_a(lo, hi, char\_class = TRUE)  
arabic\_presentation\_forms\_a(lo, hi, char\_class = TRUE)  
arabic\_presentation\_forms\_b(lo, hi, char\_class = TRUE)  
imperial\_aramaic(lo, hi, char\_class = TRUE)  
avestan(lo, hi, char\_class = TRUE)  
carian(lo, hi, char\_class = TRUE)  
cuneiform(lo, hi, char\_class = TRUE)  
cuneiform\_numbers\_and\_punctuation(lo, hi, char\_class = TRUE)

old\_persian(lo, hi, char\_class = TRUE)  
ugaritic(lo, hi, char\_class = TRUE)  
hebrew(lo, hi, char\_class = TRUE)  
lycian(lo, hi, char\_class = TRUE)  
lydian(lo, hi, char\_class = TRUE)  
mandaic(lo, hi, char\_class = TRUE)  
nabataean(lo, hi, char\_class = TRUE)  
old\_north\_arabian(lo, hi, char\_class = TRUE)  
old\_south\_arabian(lo, hi, char\_class = TRUE)  
pahlavi\_inscriptional(lo, hi, char\_class = TRUE)  
pahlavi\_psalter(lo, hi, char\_class = TRUE)  
palmyrene(lo, hi, char\_class = TRUE)  
phoenician(lo, hi, char\_class = TRUE)  
samaritan(lo, hi, char\_class = TRUE)  
syriac(lo, hi, char\_class = TRUE)  
manichaeen(lo, hi, char\_class = TRUE)  
mongolian(lo, hi, char\_class = TRUE)  
old\_turkic(lo, hi, char\_class = TRUE)  
phags\_pa(lo, hi, char\_class = TRUE)  
tibetan(lo, hi, char\_class = TRUE)  
bengali\_and\_assamese(lo, hi, char\_class = TRUE)  
brahmi(lo, hi, char\_class = TRUE)  
chakma(lo, hi, char\_class = TRUE)  
devanagari(lo, hi, char\_class = TRUE)

devanagari\_extended(lo, hi, char\_class = TRUE)  
grantha(lo, hi, char\_class = TRUE)  
gujarati(lo, hi, char\_class = TRUE)  
gurmukhi(lo, hi, char\_class = TRUE)  
kaithi(lo, hi, char\_class = TRUE)  
kannada(lo, hi, char\_class = TRUE)  
kharoshthi(lo, hi, char\_class = TRUE)  
khojki(lo, hi, char\_class = TRUE)  
khudawadi(lo, hi, char\_class = TRUE)  
lepcha(lo, hi, char\_class = TRUE)  
limbu(lo, hi, char\_class = TRUE)  
mahajani(lo, hi, char\_class = TRUE)  
malayalam(lo, hi, char\_class = TRUE)  
meetei\_mayek(lo, hi, char\_class = TRUE)  
meetei\_mayek\_extensions(lo, hi, char\_class = TRUE)  
modi(lo, hi, char\_class = TRUE)  
mro(lo, hi, char\_class = TRUE)  
ol\_chiki(lo, hi, char\_class = TRUE)  
oriya(lo, hi, char\_class = TRUE)  
saurashtra(lo, hi, char\_class = TRUE)  
sharada(lo, hi, char\_class = TRUE)  
siddham(lo, hi, char\_class = TRUE)  
sinhala(lo, hi, char\_class = TRUE)  
sinhala\_archaic\_numbers(lo, hi, char\_class = TRUE)

sora\_sompeng(lo, hi, char\_class = TRUE)  
syloiti\_nagri(lo, hi, char\_class = TRUE)  
takri(lo, hi, char\_class = TRUE)  
tamil(lo, hi, char\_class = TRUE)  
telugu(lo, hi, char\_class = TRUE)  
thaana(lo, hi, char\_class = TRUE)  
tirhuta(lo, hi, char\_class = TRUE)  
vedic\_extensions(lo, hi, char\_class = TRUE)  
warang\_citi(lo, hi, char\_class = TRUE)  
cham(lo, hi, char\_class = TRUE)  
kayah\_li(lo, hi, char\_class = TRUE)  
khmer(lo, hi, char\_class = TRUE)  
khmer\_symbols(lo, hi, char\_class = TRUE)  
lao(lo, hi, char\_class = TRUE)  
myanmar(lo, hi, char\_class = TRUE)  
myanmar\_extended\_a(lo, hi, char\_class = TRUE)  
myanmar\_extended\_b(lo, hi, char\_class = TRUE)  
new\_tai\_lue(lo, hi, char\_class = TRUE)  
pahawh\_hmong(lo, hi, char\_class = TRUE)  
pau\_cin\_hau(lo, hi, char\_class = TRUE)  
tai\_le(lo, hi, char\_class = TRUE)  
tai\_tham(lo, hi, char\_class = TRUE)  
tai\_viet(lo, hi, char\_class = TRUE)  
thai(lo, hi, char\_class = TRUE)

balinese(lo, hi, char\_class = TRUE)  
batak(lo, hi, char\_class = TRUE)  
buginese(lo, hi, char\_class = TRUE)  
buhid(lo, hi, char\_class = TRUE)  
hanunoo(lo, hi, char\_class = TRUE)  
javanese(lo, hi, char\_class = TRUE)  
rejang(lo, hi, char\_class = TRUE)  
sundanese(lo, hi, char\_class = TRUE)  
sundanese\_supplement(lo, hi, char\_class = TRUE)  
tagalog(lo, hi, char\_class = TRUE)  
tagbanwa(lo, hi, char\_class = TRUE)  
bopomofo(lo, hi, char\_class = TRUE)  
bopomofo\_extended(lo, hi, char\_class = TRUE)  
cjk\_unified\_ideographs(lo, hi, char\_class = TRUE)  
cjk\_unified\_ideographs\_extension\_a(lo, hi, char\_class = TRUE)  
cjk\_unified\_ideographs\_extension\_b(lo, hi, char\_class = TRUE)  
cjk\_unified\_ideographs\_extension\_c(lo, hi, char\_class = TRUE)  
cjk\_unified\_ideographs\_extension\_d(lo, hi, char\_class = TRUE)  
cjk\_compatibility\_ideographs(lo, hi, char\_class = TRUE)  
cjk\_compatibility\_ideographs\_supplement(lo, hi, char\_class = TRUE)  
kangxi\_radicals(lo, hi, char\_class = TRUE)  
kangxi\_radicals\_supplement(lo, hi, char\_class = TRUE)  
cjk\_strokes(lo, hi, char\_class = TRUE)  
cjk\_ideographic\_description\_characters(lo, hi, char\_class = TRUE)

hangul\_jamo(lo, hi, char\_class = TRUE)  
hangul\_jamo\_extended\_a(lo, hi, char\_class = TRUE)  
hangul\_jamo\_extended\_b(lo, hi, char\_class = TRUE)  
hangul\_compatibility\_jamo(lo, hi, char\_class = TRUE)  
hangul\_syllables(lo, hi, char\_class = TRUE)  
hiragana(lo, hi, char\_class = TRUE)  
katakana(lo, hi, char\_class = TRUE)  
katakana\_phonetic\_extensions(lo, hi, char\_class = TRUE)  
kana\_supplement(lo, hi, char\_class = TRUE)  
kanbun(lo, hi, char\_class = TRUE)  
lisu(lo, hi, char\_class = TRUE)  
miao(lo, hi, char\_class = TRUE)  
yi\_syllables(lo, hi, char\_class = TRUE)  
yi\_radicals(lo, hi, char\_class = TRUE)  
cherokee(lo, hi, char\_class = TRUE)  
deseret(lo, hi, char\_class = TRUE)  
unified\_canadian\_aboriginal\_syllabics(lo, hi, char\_class = TRUE)  
unified\_canadian\_aboriginal\_syllabics\_extended(lo, hi, char\_class = TRUE)  
alphabetic\_presentation\_forms(lo, hi, char\_class = TRUE)  
halfwidth\_and\_fullwidth\_forms(lo, hi, char\_class = TRUE)  
general\_punctuation(lo, hi, char\_class = TRUE)  
latin\_1\_punctuation(lo, hi, char\_class = TRUE)  
small\_form\_variants(lo, hi, char\_class = TRUE)  
supplemental\_punctuation(lo, hi, char\_class = TRUE)

CJK\_symbols\_and\_punctuation(lo, hi, char\_class = TRUE)  
CJK\_compatibility\_forms(lo, hi, char\_class = TRUE)  
fullwidth\_ascii\_punctuation(lo, hi, char\_class = TRUE)  
vertical\_forms(lo, hi, char\_class = TRUE)  
letterlike\_symbols(lo, hi, char\_class = TRUE)  
ancient\_symbols(lo, hi, char\_class = TRUE)  
mathematical\_alphanumeric\_symbols(lo, hi, char\_class = TRUE)  
arabic\_mathematical\_alphanumeric\_symbols(lo, hi, char\_class = TRUE)  
enclosed\_alphanumerics(lo, hi, char\_class = TRUE)  
enclosed\_alphanumeric\_supplement(lo, hi, char\_class = TRUE)  
enclosed\_CJK\_letters\_and\_months(lo, hi, char\_class = TRUE)  
enclosed\_ideographic\_supplement(lo, hi, char\_class = TRUE)  
CJK\_compatibility(lo, hi, char\_class = TRUE)  
miscellaneous\_technical(lo, hi, char\_class = TRUE)  
control\_pictures(lo, hi, char\_class = TRUE)  
optical\_character\_recognition(lo, hi, char\_class = TRUE)  
combining\_diacritic\_marks\_for\_symbols(lo, hi, char\_class = TRUE)  
aegean\_numbers(lo, hi, char\_class = TRUE)  
ancient\_greek\_numbers(lo, hi, char\_class = TRUE)  
fullwidth\_ascii\_digits(lo, hi, char\_class = TRUE)  
common\_indic\_number\_forms(lo, hi, char\_class = TRUE)  
coptic\_epact\_numbers(lo, hi, char\_class = TRUE)  
counting\_rod\_numerals(lo, hi, char\_class = TRUE)  
number\_forms(lo, hi, char\_class = TRUE)

```
rumi_numerals(lo, hi, char_class = TRUE)
sinhala_archaic_numbers(lo, hi, char_class = TRUE)
math_arrows(lo, hi, char_class = TRUE)
supplemental_arrows_a(lo, hi, char_class = TRUE)
supplemental_arrows_b(lo, hi, char_class = TRUE)
supplemental_arrows_c(lo, hi, char_class = TRUE)
additional_arrows(lo, hi, char_class = TRUE)
supplemental_mathematical_operators(lo, hi, char_class = TRUE)
miscellaneous_mathematical_symbols_a(lo, hi, char_class = TRUE)
miscellaneous_mathematical_symbols_b(lo, hi, char_class = TRUE)
floors_and_ceilings(lo, hi, char_class = TRUE)
invisible_operators(lo, hi, char_class = TRUE)
geometric_shapes(lo, hi, char_class = TRUE)
box_drawing(lo, hi, char_class = TRUE)
block_elements(lo, hi, char_class = TRUE)
geometric_shapes_extended(lo, hi, char_class = TRUE)
alchemical_symbols(lo, hi, char_class = TRUE)
braille_patterns(lo, hi, char_class = TRUE)
currency_symbols(lo, hi, char_class = TRUE)
dingbats(lo, hi, char_class = TRUE)
ornamental_dingbats(lo, hi, char_class = TRUE)
emoticons(lo, hi, char_class = TRUE)
chess_checkers draughts(lo, hi, char_class = TRUE)
domino_tiles(lo, hi, char_class = TRUE)
```



```
japanese_chess(lo, hi, char_class = TRUE)
mahjong_tiles(lo, hi, char_class = TRUE)
playing_cards(lo, hi, char_class = TRUE)
card_suits(lo, hi, char_class = TRUE)
miscellaneous_symbols_and_pictographs(lo, hi, char_class = TRUE)
musical_symbols(lo, hi, char_class = TRUE)
ancient_greek_musical_notation(lo, hi, char_class = TRUE)
byzantine_musical_symbols(lo, hi, char_class = TRUE)
transport_and_map_symbols(lo, hi, char_class = TRUE)
yijing_mono_di_and_trigrams(lo, hi, char_class = TRUE)
yijing_hexagram_symbols(lo, hi, char_class = TRUE)
tai_xuan_jing_symbols(lo, hi, char_class = TRUE)
specials(lo, hi, char_class = TRUE)
tags(lo, hi, char_class = TRUE)
variation_selectors(lo, hi, char_class = TRUE)
variation_selectors_supplement(lo, hi, char_class = TRUE)
private_use_area(lo, hi, char_class = TRUE)
supplementary_private_use_area_a(lo, hi, char_class = TRUE)
supplementary_private_use_area_b(lo, hi, char_class = TRUE)
ARMENIAN
ARMENIAN_LIGATURES
CAUCASIAN_ALBANIAN
CYPRIOT_SYLLABARY
CYRILLIC
```

CYRILLIC\_SUPPLEMENT

CYRILLIC\_EXTENDED\_A

CYRILLIC\_EXTENDED\_B

ELBASAN

GEORGIAN

GEORGIAN\_SUPPLEMENT

GLAGOLITIC

GOTHIC

GREEK\_AND\_COPTIC

GREEK\_EXTENDED

LATIN

LATIN\_1\_SUPPLEMENT

LATIN\_EXTENDED\_A

LATIN\_EXTENDED\_B

LATIN\_EXTENDED\_C

LATIN\_EXTENDED\_D

LATIN\_EXTENDED\_E

LATIN\_EXTENDED\_ADDITIONAL

LATIN\_LIGATURES

LINEAR\_A

LINEAR\_B\_SYLLABARY

LINEAR\_B\_IDEOGRAMS

OGHAM

OLD\_ITALIC

OLD\_PERMIC  
PHAISTOS\_DISC  
RUNIC  
SHAVIAN  
DUPLOYAN  
SHORTHAND\_FORMAT\_CONTROLS  
IPA\_EXTENSIONS  
PHONETIC\_EXTENSIONS  
PHONETIC\_EXTENSIONS\_SUPPLEMENT  
MODIFIER\_TONE\_LETTERS  
SPACING\_MODIFIER\_LETTERS  
SUPERSCRIPTS\_AND\_SUBSCRIPTS  
COMBINING\_DIACRITIC\_MARKS  
COMBINING\_DIACRITIC\_SUPPLEMENT  
COMBINING\_DIACRITIC\_EXTENDED  
COMBINING\_HALF\_MARKS  
BAMUN  
BAMUN\_SUPPLEMENT  
BASSA\_VAH  
COPTIC  
COPTIC\_EPACT\_NUMBERS  
EGYPTIAN\_HIEROGLYPHS  
ETHIOPIA  
ETHIOPIA\_SUPPLEMENT

ETHIOPIC\_EXTENDED  
ETHIOPIC\_EXTENDED\_A  
MENDE\_KIKAKUI  
MEROITIC\_CURSIVE  
MEROITIC\_HIEROGLYPHS  
NKO  
OSMANYA  
TIFINAGH  
VAI  
ARABIC  
ARABIC\_SUPPLEMENT  
ARABIC\_EXTENDED\_A  
ARABIC\_PRESENTATION\_FORMS\_A  
ARABIC\_PRESENTATION\_FORMS\_B  
IMPERIAL\_ARAMAIC  
AVESTAN  
CARIAN  
CUNEIFORM  
CUNEIFORM\_NUMBERS\_AND\_PUNCTUATION  
OLD\_PERSIAN  
UGARITIC  
HEBREW  
LYCIAN  
LYDIAN

MANDAIC  
NABATAEAN  
OLD\_NORTH\_ARABIAN  
OLD\_SOUTH\_ARABIAN  
PAHLAVI\_INSCRIPTIONAL  
PAHLAVI\_PALTER  
PALMYRENE  
PHOENICIAN  
SAMARITAN  
SYRIAC  
MANICHAEAN  
MONGOLIAN  
OLD\_TURKIC  
PHAGS\_PA  
TIBETAN  
BENGLI\_AND\_ASSAMESE  
BRAHMI  
CHAKMA  
DEVANAGARI  
DEVANAGARI\_EXTENDED  
GRANTHA  
GUJARATI  
GURMUKHI  
KAITHI

KANNADA

KHAROSHTHI

KHOJKI

KHUDAWADI

LEPCHA

LIMBU

MAHAJANI

MALAYALAM

MEETEI\_MAYEK

MEETEI\_MAYEK\_EXTENSIONS

MODI

MRO

OL\_CHIKI

ORIYA

SAURASHTRA

SHARADA

SIDDHAM

SINHALA

SINHALA\_ARCHAIC\_NUMBERS

SORA\_SOMPENG

SYLOTI\_NAGRI

TAKRI

TAMIL

TELUGU

THAANA

TIRHUTA

VEDIC\_EXTENSIONS

WARANG\_CITI

CHAM

KAYAH\_LI

KHMER

KHMER\_SYMBOLS

LAO

MYANMAR

MYANMAR\_EXTENDED\_A

MYANMAR\_EXTENDED\_B

NEW\_TAI\_LUE

PAHAWH\_HMONG

PAU\_CIN\_HAU

TAI\_LE

TAI\_THAM

TAI\_VIET

THAI

BALINESE

BATAK

BUGINESE

BUHID

HANUNOO

JAVANESE  
REJANG  
SUNDANESE  
SUNDANESE\_SUPPLEMENT  
TAGALOG  
TAGBANWA  
BOPOMOFO  
BOPOMOFO\_EXTENDED  
CJK\_UNIFIED\_IDEOGRAPHS  
CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_A  
CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_B  
CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_C  
CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_D  
CJK\_COMPATIBILITY\_IDEOGRAPHS  
CJK\_COMPATIBILITY\_IDEOGRAPHS\_SUPPLEMENT  
KANGXI\_RADICALS  
KANGXI\_RADICALS\_SUPPLEMENT  
CJK\_STROKES  
CJK\_IDEOGRAPHIC\_DESCRIPTION\_CHARACTERS  
HANGUL\_JAMO  
HANGUL\_JAMO\_EXTENDED\_A  
HANGUL\_JAMO\_EXTENDED\_B  
HANGUL\_COMPATIBILITY\_JAMO  
HANGUL\_SYLLABLES



HIRAGANA  
KATAKANA  
KATAKANA\_PHONETIC\_EXTENSIONS  
KANA\_SUPPLEMENT  
KANBUN  
LISU  
MIAO  
YI\_SYLLABLES  
YI\_RADICALS  
CHEROKEE  
DESERET  
UNIFIED\_CANADIAN\_ABORIGINAL\_SYLLABICS  
UNIFIED\_CANADIAN\_ABORIGINAL\_SYLLABICS\_EXTENDED  
ALPHABETIC\_PRESENTATION\_FORMS  
HALFWIDTH\_AND\_FULLWIDTH\_FORMS  
GENERAL\_PUNCTUATION  
LATIN\_1\_PUNCTUATION  
SMALL\_FORM\_VARIANTS  
SUPPLEMENTAL\_PUNCTUATION  
CJK\_SYMBOLS\_AND\_PUNCTUATION  
CJK\_COMPATIBILITY\_FORMS  
FULLWIDTH\_ASCII\_PUNCTUATION  
VERTICAL\_FORMS  
LETTERLIKE\_SYMBOLS

ANCIENT\_SYMBOLS  
MATHEMATICAL\_ALPHANUMERIC\_SYMBOLS  
ARABIC\_MATHEMATICAL\_ALPHANUMERIC\_SYMBOLS  
ENCLOSED\_ALPHANUMERICS  
ENCLOSED\_ALPHANUMERIC\_SUPPLEMENT  
ENCLOSED\_CJK\_LETTERS\_AND\_MONTHS  
ENCLOSED\_IDEOGRAPHIC\_SUPPLEMENT  
CJK\_COMPATIBILITY  
MISCELLANEOUS\_TECHNICAL  
CONTROL\_PICTURES  
OPTICAL\_CHARACTER\_RECOGNITION  
COMBINING\_DIACRITIC\_MARKS\_FOR\_SYMBOLS  
AEGEAN\_NUMBERS  
ANCIENT\_GREEK\_NUMBERS  
FULLWIDTH\_ASCII\_DIGITS  
COMMON\_INDIC\_NUMBER\_FORMS  
COPTIC\_EPACT\_NUMBERS  
COUNTING\_ROD\_NUMERALS  
NUMBER\_FORMS  
RUMI\_NUMERAL\_SYMBOLS  
SINHALA\_ARCHAIC\_NUMBERS  
MATH\_ARROWS  
SUPPLEMENTAL\_ARROWS\_A  
SUPPLEMENTAL\_ARROWS\_A

SUPPLEMENTAL\_ARROWS\_A  
ADDITIONAL\_ARROWS  
SUPPLEMENTAL\_MATHEMATICAL\_OPERATORS  
MISCELLANEOUS\_MATHEMATICAL\_SYMBOLS\_A  
MISCELLANEOUS\_MATHEMATICAL\_SYMBOLS\_B  
FLOORS\_AND\_CEILINGS  
INVISIBLE\_OPERATORS  
GEOMETRIC\_SHAPES  
BOX\_DRAWING  
BLOCK\_ELEMENTS  
GEOMETRIC\_SHAPES\_EXTENDED  
ALCHEMICAL\_SYMBOLS  
BRAILLE\_PATTERNS  
CURRENCY\_SYMBOLS  
DINGBATS  
ORNAMENTAL\_DINGBATS  
EMOTICONS  
CHESS\_CHECKERS\_DRAUGHTS  
DOMINO\_TILES  
JAPANESE\_CHESS  
MAHJONG\_TILES  
PLAYING\_CARDS  
CARD\_SUITS  
MISCELLANEOUS\_SYMBOLS\_AND\_PICTOGRAPHS

MUSICAL\_SYMBOLS  
ANCIENT\_GREEK\_MUSICAL\_NOTATION  
BYZANTINE\_MUSICAL\_SYMBOLS  
TRANSPORT\_AND\_MAP\_SYMBOLS  
YIJING\_MONO\_DI\_AND\_TRIGRAMS  
YIJING\_HEXAGRAM\_SYMBOLS  
TAI\_XUAN\_JING\_SYMBOLS  
SPECIALS  
TAGS  
VARIATION\_SELECTORS  
VARIATION\_SELECTORS\_SUPPLEMENT  
PRIVATE\_USE\_AREA  
SUPPLEMENTARY\_PRIVATE\_USE\_AREA\_A  
SUPPLEMENTARY\_PRIVATE\_USE\_AREA\_B

**Arguments**

lo	A non-negative integer. Minimum number of repeats, when grouped.
hi	positive integer. Maximum number of repeats, when grouped.
char_class	TRUE or FALSE. Should the values be wrapped into a character class?

**Format**

An object of class `regex` (inherits from `character`) of length 1.

**Value**

A character vector representing part or all of a regular expression.

**Note**

Windows currently doesn't handle Unicode points with more than four digits correctly. See [https://bugs.r-project.org/bugzilla3/show\\_bug.cgi?id=16098](https://bugs.r-project.org/bugzilla3/show_bug.cgi?id=16098)

**References**

<http://www.unicode.org/charts>

**See Also**

[ClassGroups](#)

**Examples**

```
# Classes
latin()
greek_and_coptic()
cyrillic()
arabic()

# With repetition
hebrew(3, 6)
hiragana(1, Inf)
katakana(0, Inf)

# Without a class wrapper
cjk_unified_ideographs(char_class = FALSE)

# Constants
ARMENIAN
LINEAR_B_IDEOGRAMS
DUPLOYAN
OSMANYA

## Not run:
# All the Unicode character classes
# Not run, since it generates lots of output
setdiff(
  ls("package:rebus.unicode", pattern = lower()),
  ls(
    "package:rebus.unicode",
    pattern = START %R% case_insensitive(or("up", "ugc", "unicode")))
)

## End(Not run)

# Usage
pythag <- "\u03b1^2 + \u03b2^2 = \u03b3^2"
stringi::stri_extract_all_regex(pythag, greek_and_coptic())
```

**Description**

Manipulate and combine Unicode Properties.

**Usage**

```
unicode_inverse(x, char_class = TRUE)
unicode_union(..., char_class = TRUE)
unicode_intersect(x, y, char_class = TRUE)
unicode_setdiff(x, y, char_class = TRUE)
```

**Arguments**

x	A character vector containing Unicode General Category or Unicode Properties. Use the functional forms (ugc_*) not the constants.
char_class	TRUE or FALSE. Should the values be wrapped into a character class?
...	Character vectors containing Unicode General Category or Unicode Properties. Use the functional forms (ugc_*) not the constants.
y	A character vector containing Unicode General Category or Unicode Properties. Use the functional forms (ugc_*) not the constants.

**Note**

Use these with ICU-based regular expression engines (stringi and stringr).

**References**

<http://userguide.icu-project.org/strings/unicodeset>

**Examples**

```
# POSIX [:punct:] is more or less equivalent to the union of
# Unicode punctuation and symbol general categories
unicode_union(ugc_punctuation(), ugc_symbol())

# Everything except "A" to "Z" (including punctuation, control chars etc.)
unicode_inverse("[A-Z]")

# Uppercase letters, except "A" to "Z"
unicode_setdiff(ugc_uppercase_letter(), "[A-Z]")

# "A" to "F" (in upper or lower case)
unicode_intersect(ugc_letter(), up_ascii_hex_digit())

# Usage
x <- c(letters, LETTERS)
rx <- unicode_intersect(ugc_letter(), up_ascii_hex_digit())
stringi::stri_extract_first_regex(x, rx)
```

---

up_alphabetic	<i>Unicode Properties</i>
---------------	---------------------------

---

**Description**

Match a Unicode Property.

**Usage**

```
up_alphabetic(lo, hi, char_class = TRUE)
up_ascii_hex_digit(lo, hi, char_class = TRUE)
up_bidi_control(lo, hi, char_class = TRUE)
up_bidi_mirrored(lo, hi, char_class = TRUE)
up_case_ignorable(lo, hi, char_class = TRUE)
up_case_sensitive(lo, hi, char_class = TRUE)
up_cased(lo, hi, char_class = TRUE)
up_changes_when_casefolded(lo, hi, char_class = TRUE)
up_changes_when_casemapped(lo, hi, char_class = TRUE)
up_changes_when_lowercased(lo, hi, char_class = TRUE)
up_changes_when_nfkc_casefolded(lo, hi, char_class = TRUE)
up_changes_when_titlecased(lo, hi, char_class = TRUE)
up_changes_when_uppercased(lo, hi, char_class = TRUE)
up_dash(lo, hi, char_class = TRUE)
up_default_ignorable_code_point(lo, hi, char_class = TRUE)
up_deprecated(lo, hi, char_class = TRUE)
up_diacritic(lo, hi, char_class = TRUE)
up_extender(lo, hi, char_class = TRUE)
up_hex_digit(lo, hi, char_class = TRUE)
```

```
up_hyphen(lo, hi, char_class = TRUE)
up_id_continue(lo, hi, char_class = TRUE)
up_id_start(lo, hi, char_class = TRUE)
up_ideographic(lo, hi, char_class = TRUE)
up_lowercase(lo, hi, char_class = TRUE)
up_math(lo, hi, char_class = TRUE)
up_noncharacter_code_point(lo, hi, char_class = TRUE)
up_posix_alnum(lo, hi, char_class = TRUE)
up_posix_blank(lo, hi, char_class = TRUE)
up_posix_graph(lo, hi, char_class = TRUE)
up_posix_print(lo, hi, char_class = TRUE)
up_posix_xdigit(lo, hi, char_class = TRUE)
up_quotation_mark(lo, hi, char_class = TRUE)
up_soft_dotted(lo, hi, char_class = TRUE)
up_terminal_punctuation(lo, hi, char_class = TRUE)
up_uppercase(lo, hi, char_class = TRUE)
up_white_space(lo, hi, char_class = TRUE)
UP_ALPHABETIC
UP_ASCII_HEX_DIGIT
UP_BIDI_CONTROL
UP_BIDI_MIRRORED
UP_DASH
UP_DEFAULT_IGNOREABLE_CODE_POINT
UP_DEPRECATED
```



UP\_DIACRITIC  
UP\_EXTENDER  
UP\_HEX\_DIGIT  
UP\_HYPHEN  
UP\_ID\_CONTINUE  
UP\_ID\_START  
UP\_IDEOGRAPHIC  
UP\_LOWERCASE  
UP\_MATH  
UP\_NONCHARACTER\_CODE\_POINT  
UP\_QUOTATION\_MARK  
UP\_SOFT\_DOTTED  
UP\_TERMINAL\_PUNCTUATION  
UP\_UPPERCASE  
UP\_WHITE\_SPACE  
UP\_CASE\_SENSITIVE  
UP\_POSIX\_ALNUM  
UP\_POSIX\_BLANK  
UP\_POSIX\_GRAPH  
UP\_POSIX\_PRINT  
UP\_POSIX\_XDIGIT  
UP\_CASED  
UP\_CASE\_IGNOREABLE  
UP\_CHANGES\_WHEN\_LOWERCASED

UP\_CHANGES\_WHEN\_UPPERCASED  
 UP\_CHANGES\_WHEN\_TITLECASED  
 UP\_CHANGES\_WHEN\_CASEFOLDED  
 UP\_CHANGES\_WHEN\_CASEMAPPED  
 UP\_CHANGES\_WHEN\_NFKC\_CASEFOLDED

### Arguments

lo                    A non-negative integer. Minimum number of repeats, when grouped.  
 hi                    positive integer. Maximum number of repeats, when grouped.  
 char\_class          TRUE or FALSE. Should the values be wrapped into a character class?

### Format

An object of class `regex` (inherits from `character`) of length 1.

### Value

A character vector representing part or all of a regular expression.

### References

Table 12 of the Unicode Standard Annex #44 defines the Unicode General Categories. <http://www.unicode.org/reports/tr44/>

You can see which characters are contained in a category by visiting, e.g., <http://www.fileformat.info/info/unicode/category/Nd/list.htm>

### See Also

[unicode\\_general\\_category](#), [Unicode](#), [stringi-search-charclass](#)

### Examples

```
# Classes
up_math()
up_posix_alnum()
up_changes_when_uppercased()
up_diacritic()

# With repetition
ugc_nonspacing_mark(3, 6)
up_quotation_mark(1, Inf)
up_posix_xdigit(0, Inf)

# Without a class wrapper
up_hyphen(char_class = FALSE)
```

```
# Constants
UP_ALPHABETIC
UP_DASH
UP_POSIX_ALNUM
UP_CHANGES_WHEN_LOWERCASED

## Not run:
# All the Unicode properties.
# Not run, since it generates lots of output
ls("package:rebus.unicode", pattern = "^up")

## End(Not run)

# Usage
# Hello in Samoan, Serbian, Persian, Simplified Chinese
hello <- "t\u101lofa, \u437\u434\u440\u430\u432\u43e, \u633\u644\u627\u645, \u4f60\u597d"
stringi::stri_extract_all_regex(hello, up_alphabetic(1, Inf))
stringi::stri_extract_all_regex(hello, up_case_sensitive(1, Inf))
```

# Index

## \*Topic **datasets**

- ugc\_cased\_letter, 2
  - Unicode, 6
  - up\_alphabetic, 31
- ADDITIONAL\_ARROWS (Unicode), 6
- additional\_arrows (Unicode), 6
- AEGEAN\_NUMBERS (Unicode), 6
- aegean\_numbers (Unicode), 6
- ALCHEMICAL\_SYMBOLS (Unicode), 6
- alchemical\_symbols (Unicode), 6
- ALPHABETIC\_PRESENTATION\_FORMS (Unicode), 6
- alphabetic\_presentation\_forms (Unicode), 6
- ANCIENT\_GREEK\_MUSICAL\_NOTATION (Unicode), 6
- ancient\_greek\_musical\_notation (Unicode), 6
- ANCIENT\_GREEK\_NUMBERS (Unicode), 6
- ancient\_greek\_numbers (Unicode), 6
- ANCIENT\_SYMBOLS (Unicode), 6
- ancient\_symbols (Unicode), 6
- ARABIC (Unicode), 6
- arabic (Unicode), 6
- ARABIC\_EXTENDED\_A (Unicode), 6
- arabic\_extended\_a (Unicode), 6
- ARABIC\_MATHEMATICAL\_ALPHANUMERIC\_SYMBOLS (Unicode), 6
- arabic\_mathematical\_alphanumeric\_symbols (Unicode), 6
- ARABIC\_PRESENTATION\_FORMS\_A (Unicode), 6
- arabic\_presentation\_forms\_a (Unicode), 6
- ARABIC\_PRESENTATION\_FORMS\_B (Unicode), 6
- arabic\_presentation\_forms\_b (Unicode), 6
- ARABIC\_SUPPLEMENT (Unicode), 6
- arabic\_supplement (Unicode), 6
- ARMENIAN (Unicode), 6
- armenian (Unicode), 6
- ARMENIAN\_LIGATURES (Unicode), 6
- armenian\_ligatures (Unicode), 6
- AVESTAN (Unicode), 6
- avestan (Unicode), 6
- BALINESE (Unicode), 6
- balinese (Unicode), 6
- BAMUN (Unicode), 6
- bamun (Unicode), 6
- BAMUN\_SUPPLEMENT (Unicode), 6
- bamun\_supplement (Unicode), 6
- BASSA\_VAH (Unicode), 6
- bassa\_vah (Unicode), 6
- BATAK (Unicode), 6
- batak (Unicode), 6
- BENGALI\_AND\_ASSAMESE (Unicode), 6
- bengali\_and\_assamese (Unicode), 6
- BLOCK\_ELEMENTS (Unicode), 6
- block\_elements (Unicode), 6
- BOPOMOFO (Unicode), 6
- bopomofo (Unicode), 6
- BOPOMOFO\_EXTENDED (Unicode), 6
- bopomofo\_extended (Unicode), 6
- BOX\_DRAWING (Unicode), 6
- box\_drawing (Unicode), 6
- BRAHMI (Unicode), 6
- brahmi (Unicode), 6
- BRAILLE\_PATTERNS (Unicode), 6
- braille\_patterns (Unicode), 6
- BUGINESE (Unicode), 6
- buginese (Unicode), 6
- BUHID (Unicode), 6
- buhid (Unicode), 6
- BYZANTINE\_MUSICAL\_SYMBOLS (Unicode), 6
- byzantine\_musical\_symbols (Unicode), 6
- CARD\_SUITS (Unicode), 6
- card\_suits (Unicode), 6
- CARIAN (Unicode), 6
- carian (Unicode), 6
- CAUCASIAN\_ALBANIAN (Unicode), 6

- caucasian\_albanian (Unicode), 6
- CHAKMA (Unicode), 6
- chakma (Unicode), 6
- CHAM (Unicode), 6
- cham (Unicode), 6
- CHEROKEE (Unicode), 6
- cherokee (Unicode), 6
- CHESS\_CHECKERS\_DRAUGHTS (Unicode), 6
- chess\_checkers draughts (Unicode), 6
- CJK\_COMPATIBILITY (Unicode), 6
- chk\_compatibility (Unicode), 6
- CJK\_COMPATIBILITY\_FORMS (Unicode), 6
- chk\_compatibility\_forms (Unicode), 6
- CJK\_COMPATIBILITY\_IDEOGRAPHS (Unicode), 6
- chk\_compatibility\_ideographs (Unicode), 6
- CJK\_COMPATIBILITY\_IDEOGRAPHS\_SUPPLEMENT (Unicode), 6
- chk\_compatibility\_ideographs\_supplement (Unicode), 6
- CJK\_IDEOGRAPHIC\_DESCRIPTION\_CHARACTERS (Unicode), 6
- chk\_ideographic\_description\_characters (Unicode), 6
- CJK\_STROKES (Unicode), 6
- chk\_strokes (Unicode), 6
- CJK\_SYMBOLS\_AND\_PUNCTUATION (Unicode), 6
- chk\_symbols\_and\_punctuation (Unicode), 6
- CJK\_UNIFIED\_IDEOGRAPHS (Unicode), 6
- chk\_unified\_ideographs (Unicode), 6
- CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_A (Unicode), 6
- chk\_unified\_ideographs\_extension\_a (Unicode), 6
- CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_B (Unicode), 6
- chk\_unified\_ideographs\_extension\_b (Unicode), 6
- CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_C (Unicode), 6
- chk\_unified\_ideographs\_extension\_c (Unicode), 6
- CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_D (Unicode), 6
- chk\_unified\_ideographs\_extension\_d (Unicode), 6
- ClassGroups, 29
- COMBINING\_DIACRITIC\_EXTENDED (Unicode), 6
- combining\_diacritic\_extended (Unicode), 6
- COMBINING\_DIACRITIC\_MARKS (Unicode), 6
- combining\_diacritic\_marks (Unicode), 6
- COMBINING\_DIACRITIC\_MARKS\_FOR\_SYMBOLS (Unicode), 6
- combining\_diacritic\_marks\_for\_symbols (Unicode), 6
- COMBINING\_DIACRITIC\_SUPPLEMENT (Unicode), 6
- combining\_diacritic\_supplement (Unicode), 6
- COMBINING\_HALF\_MARKS (Unicode), 6
- combining\_half\_marks (Unicode), 6
- COMMON\_INDIC\_NUMBER\_FORMS (Unicode), 6
- common\_indic\_number\_forms (Unicode), 6
- CONTROL\_PICTURES (Unicode), 6
- control\_pictures (Unicode), 6
- COPTIC (Unicode), 6
- coptic (Unicode), 6
- COPTIC\_EPACT\_NUMBERS (Unicode), 6
- coptic\_epact\_numbers (Unicode), 6
- COUNTING\_ROD\_NUMERALS (Unicode), 6
- counting\_rod\_numerals (Unicode), 6
- CUNEIFORM (Unicode), 6
- cuneiform (Unicode), 6
- CUNEIFORM\_NUMBERS\_AND\_PUNCTUATION (Unicode), 6
- cuneiform\_numbers\_and\_punctuation (Unicode), 6
- CURRENCY\_SYMBOLS (Unicode), 6
- currency\_symbols (Unicode), 6
- CYPRIOT\_SYLLABARY (Unicode), 6
- cyriot\_syllabary (Unicode), 6
- CYRILLIC (Unicode), 6
- cyrillic (Unicode), 6
- CYRILLIC\_EXTENDED\_A (Unicode), 6
- cyrillic\_extended\_a (Unicode), 6
- CYRILLIC\_EXTENDED\_B (Unicode), 6
- cyrillic\_extended\_b (Unicode), 6
- CYRILLIC\_SUPPLEMENT (Unicode), 6
- cyrillic\_supplement (Unicode), 6
- DESERET (Unicode), 6
- deseret (Unicode), 6
- DEVANAGARI (Unicode), 6
- devanagari (Unicode), 6

- DEVANAGARI\_EXTENDED (Unicode), 6
- devanagari\_extended (Unicode), 6
- DINGBATS (Unicode), 6
- dingbats (Unicode), 6
- DOMINO\_TILES (Unicode), 6
- domino\_tiles (Unicode), 6
- DUPLOYAN (Unicode), 6
- duployan (Unicode), 6
  
- EGYPTIAN\_HIEROGLYPHS (Unicode), 6
- egyptian\_hieroglyphs (Unicode), 6
- ELBASAN (Unicode), 6
- elbasan (Unicode), 6
- EMOTICONS (Unicode), 6
- emoticons (Unicode), 6
- ENCLOSED\_ALPHANUMERIC\_SUPPLEMENT (Unicode), 6
- enclosed\_alphanumeric\_supplement (Unicode), 6
- ENCLOSED\_ALPHANUMERICS (Unicode), 6
- enclosed\_alphanumerics (Unicode), 6
- ENCLOSED\_CJK\_LETTERS\_AND\_MONTHS (Unicode), 6
- enclosed\_cjk\_letters\_and\_months (Unicode), 6
- ENCLOSED\_IDEOGRAPHIC\_SUPPLEMENT (Unicode), 6
- enclosed\_ideographic\_supplement (Unicode), 6
- ETHIOPIIC (Unicode), 6
- ethiopic (Unicode), 6
- ETHIOPIIC\_EXTENDED (Unicode), 6
- ethiopic\_extended (Unicode), 6
- ETHIOPIIC\_EXTENDED\_A (Unicode), 6
- ethiopic\_extended\_a (Unicode), 6
- ETHIOPIIC\_SUPPLEMENT (Unicode), 6
- ethiopic\_supplement (Unicode), 6
  
- FLOORS\_AND\_CEILINGS (Unicode), 6
- floors\_and\_ceilings (Unicode), 6
- FULLWIDTH\_ASCII\_DIGITS (Unicode), 6
- fullwidth\_ascii\_digits (Unicode), 6
- FULLWIDTH\_ASCII\_PUNCTUATION (Unicode), 6
- fullwidth\_ascii\_punctuation (Unicode), 6
  
- GENERAL\_PUNCTUATION (Unicode), 6
- general\_punctuation (Unicode), 6
- GEOMETRIC\_SHAPES (Unicode), 6
- geometric\_shapes (Unicode), 6
  
- GEOMETRIC\_SHAPES\_EXTENDED (Unicode), 6
- geometric\_shapes\_extended (Unicode), 6
- GEORGIAN (Unicode), 6
- georgian (Unicode), 6
- GEORGIAN\_SUPPLEMENT (Unicode), 6
- georgian\_supplement (Unicode), 6
- GLAGOLITIC (Unicode), 6
- glagolitic (Unicode), 6
- GOTHIC (Unicode), 6
- gothic (Unicode), 6
- GRANTHA (Unicode), 6
- grantha (Unicode), 6
- GREEK\_AND\_COPTIC (Unicode), 6
- greek\_and\_coptic (Unicode), 6
- GREEK\_EXTENDED (Unicode), 6
- greek\_extended (Unicode), 6
- GUJARATI (Unicode), 6
- gujarati (Unicode), 6
- GURMUKHI (Unicode), 6
- gurmukhi (Unicode), 6
  
- HALFWIDTH\_AND\_FULLWIDTH\_FORMS (Unicode), 6
- halfwidth\_and\_fullwidth\_forms (Unicode), 6
- HANGUL\_COMPATIBILITY\_JAMO (Unicode), 6
- hangul\_compatibility\_jamo (Unicode), 6
- HANGUL\_JAMO (Unicode), 6
- hangul\_jamo (Unicode), 6
- HANGUL\_JAMO\_EXTENDED\_A (Unicode), 6
- hangul\_jamo\_extended\_a (Unicode), 6
- HANGUL\_JAMO\_EXTENDED\_B (Unicode), 6
- hangul\_jamo\_extended\_b (Unicode), 6
- HANGUL\_SYLLABLES (Unicode), 6
- hangul\_syllables (Unicode), 6
- HANUNOO (Unicode), 6
- hanunoo (Unicode), 6
- HEBREW (Unicode), 6
- hebrew (Unicode), 6
- HIRAGANA (Unicode), 6
- hiragana (Unicode), 6
  
- IMPERIAL\_ARAMAIC (Unicode), 6
- imperial\_aramaic (Unicode), 6
- INVISIBLE\_OPERATORS (Unicode), 6
- invisible\_operators (Unicode), 6
- IPA\_EXTENSIONS (Unicode), 6
- ipa\_extensions (Unicode), 6

- JAPANESE\_CHESS (Unicode), 6  
 japanese\_chess (Unicode), 6  
 JAVANESE (Unicode), 6  
 javanese (Unicode), 6  
  
 KAITHI (Unicode), 6  
 kaithi (Unicode), 6  
 KANA\_SUPPLEMENT (Unicode), 6  
 kana\_supplement (Unicode), 6  
 KANBUN (Unicode), 6  
 kanbun (Unicode), 6  
 KANGXI\_RADICALS (Unicode), 6  
 kangxi\_radicals (Unicode), 6  
 KANGXI\_RADICALS\_SUPPLEMENT (Unicode), 6  
 kangxi\_radicals\_supplement (Unicode), 6  
 KANNADA (Unicode), 6  
 kannada (Unicode), 6  
 KATAKANA (Unicode), 6  
 katakana (Unicode), 6  
 KATAKANA\_PHONETIC\_EXTENSIONS (Unicode),  
     6  
 katakana\_phonetic\_extensions (Unicode),  
     6  
 KAYAH\_LI (Unicode), 6  
 kayah\_li (Unicode), 6  
 KHAROSHTHI (Unicode), 6  
 kharoshthi (Unicode), 6  
 KHMER (Unicode), 6  
 khmer (Unicode), 6  
 KHMER\_SYMBOLS (Unicode), 6  
 khmer\_symbols (Unicode), 6  
 KHOJKI (Unicode), 6  
 khojki (Unicode), 6  
 KHUDAWADI (Unicode), 6  
 khudawadi (Unicode), 6  
  
 LAO (Unicode), 6  
 lao (Unicode), 6  
 LATIN (Unicode), 6  
 latin (Unicode), 6  
 LATIN\_1\_PUNCTUATION (Unicode), 6  
 latin\_1\_punctuation (Unicode), 6  
 LATIN\_1\_SUPPLEMENT (Unicode), 6  
 latin\_1\_supplement (Unicode), 6  
 LATIN\_EXTENDED\_A (Unicode), 6  
 latin\_extended\_a (Unicode), 6  
 LATIN\_EXTENDED\_ADDITIONAL (Unicode), 6  
 latin\_extended\_additional (Unicode), 6  
 LATIN\_EXTENDED\_B (Unicode), 6  
 latin\_extended\_b (Unicode), 6  
 LATIN\_EXTENDED\_C (Unicode), 6  
 latin\_extended\_c (Unicode), 6  
 LATIN\_EXTENDED\_D (Unicode), 6  
 latin\_extended\_d (Unicode), 6  
 LATIN\_EXTENDED\_E (Unicode), 6  
 latin\_extended\_e (Unicode), 6  
 LATIN\_LIGATURES (Unicode), 6  
 latin\_ligatures (Unicode), 6  
 LEPCHA (Unicode), 6  
 lepcha (Unicode), 6  
 LETTERLIKE\_SYMBOLS (Unicode), 6  
 letterlike\_symbols (Unicode), 6  
 LIMBU (Unicode), 6  
 limbu (Unicode), 6  
 LINEAR\_A (Unicode), 6  
 linear\_a (Unicode), 6  
 LINEAR\_B\_IDEOGRAMS (Unicode), 6  
 linear\_b\_ideograms (Unicode), 6  
 LINEAR\_B\_SYLLABARY (Unicode), 6  
 linear\_b\_syllabary (Unicode), 6  
 LISU (Unicode), 6  
 lisu (Unicode), 6  
 LYCIAN (Unicode), 6  
 lycian (Unicode), 6  
 LYDIAN (Unicode), 6  
 lydian (Unicode), 6  
  
 MAHAJANI (Unicode), 6  
 mahajani (Unicode), 6  
 MAHJONG\_TILES (Unicode), 6  
 mahjong\_tiles (Unicode), 6  
 MALAYALAM (Unicode), 6  
 malayalam (Unicode), 6  
 MANDAIC (Unicode), 6  
 mandaic (Unicode), 6  
 MANICHAEAN (Unicode), 6  
 manichaeen (Unicode), 6  
 MATH\_ARROWS (Unicode), 6  
 math\_arrows (Unicode), 6  
 MATHEMATICAL\_ALPHANUMERIC\_SYMBOLS  
     (Unicode), 6  
 mathematical\_alphanumeric\_symbols  
     (Unicode), 6  
 MEETEI\_MAYEK (Unicode), 6  
 meetei\_mayek (Unicode), 6  
 MEETEI\_MAYEK\_EXTENSIONS (Unicode), 6  
 meetei\_mayek\_extensions (Unicode), 6  
 MENDE\_KIKAKUI (Unicode), 6

- mende\_kikakui (Unicode), 6
- MEROITIC\_CURSIVE (Unicode), 6
- meroitic\_cursive (Unicode), 6
- MEROITIC\_HIEROGLYPHS (Unicode), 6
- meroitic\_hieroglyphs (Unicode), 6
- MIAO (Unicode), 6
- miao (Unicode), 6
- MISCELLANEOUS\_MATHEMATICAL\_SYMBOLS\_A (Unicode), 6
- miscellaneous\_mathematical\_symbols\_a (Unicode), 6
- MISCELLANEOUS\_MATHEMATICAL\_SYMBOLS\_B (Unicode), 6
- miscellaneous\_mathematical\_symbols\_b (Unicode), 6
- MISCELLANEOUS\_SYMBOLS\_AND\_PICTOGRAPHS (Unicode), 6
- miscellaneous\_symbols\_and\_pictographs (Unicode), 6
- MISCELLANEOUS\_TECHNICAL (Unicode), 6
- miscellaneous\_technical (Unicode), 6
- MODI (Unicode), 6
- modi (Unicode), 6
- MODIFIER\_TONE\_LETTERS (Unicode), 6
- modifier\_tone\_letters (Unicode), 6
- MONGOLIAN (Unicode), 6
- mongolian (Unicode), 6
- MRO (Unicode), 6
- mro (Unicode), 6
- MUSICAL\_SYMBOLS (Unicode), 6
- musical\_symbols (Unicode), 6
- MYANMAR (Unicode), 6
- myanmar (Unicode), 6
- MYANMAR\_EXTENDED\_A (Unicode), 6
- myanmar\_extended\_a (Unicode), 6
- MYANMAR\_EXTENDED\_B (Unicode), 6
- myanmar\_extended\_b (Unicode), 6
  
- NABATAEAN (Unicode), 6
- nabataean (Unicode), 6
- NEW\_TAI\_LUE (Unicode), 6
- new\_tai\_lue (Unicode), 6
- NKO (Unicode), 6
- nko (Unicode), 6
- NUMBER\_FORMS (Unicode), 6
- number\_forms (Unicode), 6
  
- OGHAM (Unicode), 6
- ogham (Unicode), 6
  
- OL\_CHIKI (Unicode), 6
- ol\_chiki (Unicode), 6
- OLD\_ITALIC (Unicode), 6
- old\_italic (Unicode), 6
- OLD\_NORTH\_ARABIAN (Unicode), 6
- old\_north\_arabian (Unicode), 6
- OLD\_PERMIC (Unicode), 6
- old\_permic (Unicode), 6
- OLD\_PERSIAN (Unicode), 6
- old\_persian (Unicode), 6
- OLD\_SOUTH\_ARABIAN (Unicode), 6
- old\_south\_arabian (Unicode), 6
- OLD\_TURKIC (Unicode), 6
- old\_turkic (Unicode), 6
- OPTICAL\_CHARACTER\_RECOGNITION (Unicode), 6
- optical\_character\_recognition (Unicode), 6
- ORIYA (Unicode), 6
- oriya (Unicode), 6
- ORNAMENTAL\_DINGBATS (Unicode), 6
- ornamental\_dingbats (Unicode), 6
- OSMANYA (Unicode), 6
- osmanya (Unicode), 6
  
- PAHAHW\_HMONG (Unicode), 6
- pahawh\_hmong (Unicode), 6
- PAHLAVI\_INSCRIPTIONAL (Unicode), 6
- pahlavi\_inscriptional (Unicode), 6
- PAHLAVI\_PSALTER (Unicode), 6
- pahlavi\_psalter (Unicode), 6
- PALMYRENE (Unicode), 6
- palmyrene (Unicode), 6
- PAU\_CIN\_HAU (Unicode), 6
- pau\_cin\_hau (Unicode), 6
- PHAGS\_PA (Unicode), 6
- phags\_pa (Unicode), 6
- PHAISTOS\_DISC (Unicode), 6
- phaistos\_disc (Unicode), 6
- PHOENICIAN (Unicode), 6
- phoenician (Unicode), 6
- PHONETIC\_EXTENSIONS (Unicode), 6
- phonetic\_extensions (Unicode), 6
- PHONETIC\_EXTENSIONS\_SUPPLEMENT (Unicode), 6
- phonetic\_extensions\_supplement (Unicode), 6
- PLAYING\_CARDS (Unicode), 6
- playing\_cards (Unicode), 6



- PRIVATE\_USE\_AREA (Unicode), 6  
private\_use\_area (Unicode), 6
- REJANG (Unicode), 6  
rejang (Unicode), 6
- RUMI\_NUMERAL\_SYMBOLS (Unicode), 6  
rumi\_numeral\_symbols (Unicode), 6
- RUNIC (Unicode), 6  
runic (Unicode), 6
- SAMARITAN (Unicode), 6  
samaritan (Unicode), 6
- SAURASHTRA (Unicode), 6  
saurashtra (Unicode), 6
- SHARADA (Unicode), 6  
sharada (Unicode), 6
- SHAVIAN (Unicode), 6  
shavian (Unicode), 6
- SHORTHAND\_FORMAT\_CONTROLS (Unicode), 6  
shorthand\_format\_controls (Unicode), 6
- SIDDHAM (Unicode), 6  
siddham (Unicode), 6
- SINHALA (Unicode), 6  
sinhala (Unicode), 6
- SINHALA\_ARCHAIC\_NUMBERS (Unicode), 6  
sinhala\_archaic\_numbers (Unicode), 6
- SMALL\_FORM\_VARIANTS (Unicode), 6  
small\_form\_variants (Unicode), 6
- SORA\_SOMPENG (Unicode), 6  
sora\_sompeng (Unicode), 6
- SPACING\_MODIFIER\_LETTERS (Unicode), 6  
spacing\_modifier\_letters (Unicode), 6
- SPECIALS (Unicode), 6  
specials (Unicode), 6
- SUNDANESE (Unicode), 6  
sundanese (Unicode), 6
- SUNDANESE\_SUPPLEMENT (Unicode), 6  
sundanese\_supplement (Unicode), 6
- SUPERSCRIPTS\_AND\_SUBSCRIPTS (Unicode), 6  
superscripts\_and\_subscripts (Unicode), 6
- SUPPLEMENTAL\_ARROWS\_A (Unicode), 6  
supplemental\_arrows\_a (Unicode), 6
- SUPPLEMENTAL\_MATHEMATICAL\_OPERATORS (Unicode), 6  
supplemental\_mathematical\_operators (Unicode), 6
- SUPPLEMENTAL\_PUNCTUATION (Unicode), 6  
supplemental\_punctuation (Unicode), 6
- SUPPLEMENTARY\_PRIVATE\_USE\_AREA\_A (Unicode), 6  
supplementary\_private\_use\_area\_a (Unicode), 6
- SUPPLEMENTARY\_PRIVATE\_USE\_AREA\_B (Unicode), 6  
supplementary\_private\_use\_area\_b (Unicode), 6
- SYLOTI\_NAGRI (Unicode), 6  
syloiti\_nagri (Unicode), 6
- SYRIAC (Unicode), 6  
syriac (Unicode), 6
- TAGALOG (Unicode), 6  
tagalog (Unicode), 6
- TAGBANWA (Unicode), 6  
tagbanwa (Unicode), 6
- TAGS (Unicode), 6  
tags (Unicode), 6
- TAI\_LE (Unicode), 6  
tai\_le (Unicode), 6
- TAI\_THAM (Unicode), 6  
tai\_tham (Unicode), 6
- TAI\_VIET (Unicode), 6  
tai\_viet (Unicode), 6
- TAI\_XUAN\_JING\_SYMBOLS (Unicode), 6  
tai\_xuan\_jing\_symbols (Unicode), 6
- TAKRI (Unicode), 6  
takri (Unicode), 6
- TAMIL (Unicode), 6  
tamil (Unicode), 6
- TELUGU (Unicode), 6  
telugu (Unicode), 6
- THAANA (Unicode), 6  
thaana (Unicode), 6
- THAI (Unicode), 6  
thai (Unicode), 6
- TIBETAN (Unicode), 6  
tibetan (Unicode), 6
- TIFINAGH (Unicode), 6  
tifinagh (Unicode), 6
- TIRHUTA (Unicode), 6  
tirhuta (Unicode), 6
- TRANSPORT\_AND\_MAP\_SYMBOLS (Unicode), 6  
transport\_and\_map\_symbols (Unicode), 6
- UGARITIC (Unicode), 6  
ugaritic (Unicode), 6
- UGC\_CASED\_LETTER (ugc\_cased\_letter), 2

ugc\_cased\_letter, 2  
 UGC\_CLOSE\_PUNCTUATION  
     (ugc\_cased\_letter), 2  
 ugc\_close\_punctuation  
     (ugc\_cased\_letter), 2  
 UGC\_CONNECTOR\_PUNCTUATION  
     (ugc\_cased\_letter), 2  
 ugc\_connector\_punctuation  
     (ugc\_cased\_letter), 2  
 UGC\_CONTROL (ugc\_cased\_letter), 2  
 ugc\_control (ugc\_cased\_letter), 2  
 UGC\_CURRENCY\_SYMBOL (ugc\_cased\_letter),  
     2  
 ugc\_currency\_symbol (ugc\_cased\_letter),  
     2  
 UGC\_DASH\_PUNCTUATION  
     (ugc\_cased\_letter), 2  
 ugc\_dash\_punctuation  
     (ugc\_cased\_letter), 2  
 UGC\_DECIMAL\_NUMBER (ugc\_cased\_letter), 2  
 ugc\_decimal\_number (ugc\_cased\_letter), 2  
 UGC\_ENCLOSING\_MARK (ugc\_cased\_letter), 2  
 ugc\_enclosing\_mark (ugc\_cased\_letter), 2  
 UGC\_FINAL\_PUNCTUATION  
     (ugc\_cased\_letter), 2  
 ugc\_final\_punctuation  
     (ugc\_cased\_letter), 2  
 UGC\_FORMAT\_CONTROL (ugc\_cased\_letter), 2  
 ugc\_format\_control (ugc\_cased\_letter), 2  
 UGC\_INITIAL\_PUNCTUATION  
     (ugc\_cased\_letter), 2  
 ugc\_initial\_punctuation  
     (ugc\_cased\_letter), 2  
 UGC\_LETTER (ugc\_cased\_letter), 2  
 ugc\_letter (ugc\_cased\_letter), 2  
 UGC\_LETTER\_NUMBER (ugc\_cased\_letter), 2  
 ugc\_letter\_number (ugc\_cased\_letter), 2  
 UGC\_LINE\_SEPARATOR (ugc\_cased\_letter), 2  
 ugc\_line\_separator (ugc\_cased\_letter), 2  
 UGC\_LOWERCASE\_LETTER  
     (ugc\_cased\_letter), 2  
 ugc\_lowercase\_letter  
     (ugc\_cased\_letter), 2  
 UGC\_MARK (ugc\_cased\_letter), 2  
 ugc\_mark (ugc\_cased\_letter), 2  
 UGC\_MATH\_SYMBOL (ugc\_cased\_letter), 2  
 ugc\_math\_symbol (ugc\_cased\_letter), 2  
 UGC\_MODIFIER\_LETTER (ugc\_cased\_letter),  
     2  
 ugc\_modifier\_letter (ugc\_cased\_letter),  
     2  
 UGC\_MODIFIER\_SYMBOL (ugc\_cased\_letter),  
     2  
 ugc\_modifier\_symbol (ugc\_cased\_letter),  
     2  
 UGC\_NONSPACING\_MARK (ugc\_cased\_letter),  
     2  
 ugc\_nonspacing\_mark (ugc\_cased\_letter),  
     2  
 UGC\_NUMBER (ugc\_cased\_letter), 2  
 ugc\_number (ugc\_cased\_letter), 2  
 UGC\_OPEN\_PUNCTUATION  
     (ugc\_cased\_letter), 2  
 ugc\_open\_punctuation  
     (ugc\_cased\_letter), 2  
 UGC\_OTHER (ugc\_cased\_letter), 2  
 ugc\_other (ugc\_cased\_letter), 2  
 UGC\_OTHER\_LETTER (ugc\_cased\_letter), 2  
 ugc\_other\_letter (ugc\_cased\_letter), 2  
 UGC\_OTHER\_NUMBER (ugc\_cased\_letter), 2  
 ugc\_other\_number (ugc\_cased\_letter), 2  
 UGC\_OTHER\_PUNCTUATION  
     (ugc\_cased\_letter), 2  
 ugc\_other\_punctuation  
     (ugc\_cased\_letter), 2  
 UGC\_OTHER\_SYMBOL (ugc\_cased\_letter), 2  
 ugc\_other\_symbol (ugc\_cased\_letter), 2  
 UGC\_PARAGRAPH\_SEPARATOR  
     (ugc\_cased\_letter), 2  
 ugc\_paragraph\_separator  
     (ugc\_cased\_letter), 2  
 UGC\_PRIVATE\_USE\_CONTROL  
     (ugc\_cased\_letter), 2  
 ugc\_private\_use\_control  
     (ugc\_cased\_letter), 2  
 UGC\_PUNCTUATION (ugc\_cased\_letter), 2  
 ugc\_punctuation (ugc\_cased\_letter), 2  
 UGC\_SEPARATOR (ugc\_cased\_letter), 2  
 ugc\_separator (ugc\_cased\_letter), 2  
 UGC\_SPACE\_SEPARATOR (ugc\_cased\_letter),  
     2  
 ugc\_space\_separator (ugc\_cased\_letter),  
     2  
 UGC\_SPACING\_MARK (ugc\_cased\_letter), 2  
 ugc\_spacing\_mark (ugc\_cased\_letter), 2  
 UGC\_SURROGATE\_CONTROL

- (ugc\_cased\_letter), 2
- ugc\_surrogate\_control
  - (ugc\_cased\_letter), 2
- UGC\_SYMBOL (ugc\_cased\_letter), 2
- ugc\_symbol (ugc\_cased\_letter), 2
- UGC\_TITLECASE\_LETTER
  - (ugc\_cased\_letter), 2
- ugc\_titlecase\_letter
  - (ugc\_cased\_letter), 2
- UGC\_UNASSIGNED\_CONTROL
  - (ugc\_cased\_letter), 2
- ugc\_unassigned\_control
  - (ugc\_cased\_letter), 2
- UGC\_UPPERCASE\_LETTER
  - (ugc\_cased\_letter), 2
- ugc\_uppercase\_letter
  - (ugc\_cased\_letter), 2
- Unicode, 6, 6, 34
- unicode\_general\_category, 34
- unicode\_general\_category
  - (ugc\_cased\_letter), 2
- unicode\_intersect (UnicodeOperators), 29
- unicode\_inverse (UnicodeOperators), 29
- unicode\_operators (UnicodeOperators), 29
- unicode\_property, 6
- unicode\_property (up\_alphabetic), 31
- unicode\_setdiff (UnicodeOperators), 29
- unicode\_union (UnicodeOperators), 29
- UnicodeGeneralCategory
  - (ugc\_cased\_letter), 2
- UnicodeOperators, 29
- UnicodeProperty (up\_alphabetic), 31
- UNIFIED\_CANADIAN\_ABORIGINAL\_SYLLABICS
  - (Unicode), 6
- unified\_canadian\_aboriginal\_syllabics
  - (Unicode), 6
- UNIFIED\_CANADIAN\_ABORIGINAL\_SYLLABICS\_EXTENDED
  - (Unicode), 6
- unified\_canadian\_aboriginal\_syllabics\_extended
  - (Unicode), 6
- UP\_ALPHABETIC (up\_alphabetic), 31
- up\_alphabetic, 31
- UP\_ASCII\_HEX\_DIGIT (up\_alphabetic), 31
- up\_ascii\_hex\_digit (up\_alphabetic), 31
- UP\_BIDI\_CONTROL (up\_alphabetic), 31
- up\_bidi\_control (up\_alphabetic), 31
- UP\_BIDI\_MIRRORED (up\_alphabetic), 31
- up\_bidi\_mirrored (up\_alphabetic), 31
- UP\_CASE\_IGNOREABLE (up\_alphabetic), 31
- up\_case\_ignorable (up\_alphabetic), 31
- UP\_CASE\_SENSITIVE (up\_alphabetic), 31
- up\_case\_sensitive (up\_alphabetic), 31
- UP\_CASED (up\_alphabetic), 31
- up\_cased (up\_alphabetic), 31
- UP\_CHANGES\_WHEN\_CASEFOLDED
  - (up\_alphabetic), 31
- up\_changes\_when\_casefolded
  - (up\_alphabetic), 31
- UP\_CHANGES\_WHEN\_CASEMAPPED
  - (up\_alphabetic), 31
- up\_changes\_when\_casemapped
  - (up\_alphabetic), 31
- UP\_CHANGES\_WHEN\_LOWERCASED
  - (up\_alphabetic), 31
- up\_changes\_when\_lowercased
  - (up\_alphabetic), 31
- UP\_CHANGES\_WHEN\_NFKC\_CASEFOLDED
  - (up\_alphabetic), 31
- up\_changes\_when\_nfkc\_casefolded
  - (up\_alphabetic), 31
- UP\_CHANGES\_WHEN\_TITLECASED
  - (up\_alphabetic), 31
- up\_changes\_when\_titlecased
  - (up\_alphabetic), 31
- UP\_CHANGES\_WHEN\_UPPERCASED
  - (up\_alphabetic), 31
- up\_changes\_when\_uppercased
  - (up\_alphabetic), 31
- UP\_DASH (up\_alphabetic), 31
- up\_dash (up\_alphabetic), 31
- UP\_DEFAULT\_IGNOREABLE\_CODE\_POINT
  - (up\_alphabetic), 31
- up\_default\_ignorable\_code\_point
  - (up\_alphabetic), 31
- UP\_DEPRECATED (up\_alphabetic), 31
- up\_deprecated (up\_alphabetic), 31
- UP\_DIACRITIC (up\_alphabetic), 31
- up\_diacritic (up\_alphabetic), 31
- UP\_EXTENDER (up\_alphabetic), 31
- up\_extender (up\_alphabetic), 31
- UP\_HEX\_DIGIT (up\_alphabetic), 31
- up\_hex\_digit (up\_alphabetic), 31
- UP\_HYPHEN (up\_alphabetic), 31
- up\_hyphen (up\_alphabetic), 31
- UP\_ID\_CONTINUE (up\_alphabetic), 31
- up\_id\_continue (up\_alphabetic), 31

- UP\_ID\_START (up\_alphabetic), 31
- up\_id\_start (up\_alphabetic), 31
- UP\_IDEOGRAPHIC (up\_alphabetic), 31
- up\_ideographic (up\_alphabetic), 31
- UP\_LOWERCASE (up\_alphabetic), 31
- up\_lowercase (up\_alphabetic), 31
- UP\_MATH (up\_alphabetic), 31
- up\_math (up\_alphabetic), 31
- UP\_NONCHARACTER\_CODE\_POINT  
    (up\_alphabetic), 31
- up\_noncharacter\_code\_point  
    (up\_alphabetic), 31
- UP\_POSIX\_ALNUM (up\_alphabetic), 31
- up\_posix\_alnum (up\_alphabetic), 31
- UP\_POSIX\_BLANK (up\_alphabetic), 31
- up\_posix\_blank (up\_alphabetic), 31
- UP\_POSIX\_GRAPH (up\_alphabetic), 31
- up\_posix\_graph (up\_alphabetic), 31
- UP\_POSIX\_PRINT (up\_alphabetic), 31
- up\_posix\_print (up\_alphabetic), 31
- UP\_POSIX\_XDIGIT (up\_alphabetic), 31
- up\_posix\_xdigit (up\_alphabetic), 31
- UP\_QUOTATION\_MARK (up\_alphabetic), 31
- up\_quotation\_mark (up\_alphabetic), 31
- UP\_SOFT\_DOTTED (up\_alphabetic), 31
- up\_soft\_dotted (up\_alphabetic), 31
- UP\_TERMINAL\_PUNCTUATION  
    (up\_alphabetic), 31
- up\_terminal\_punctuation  
    (up\_alphabetic), 31
- UP\_UPPERCASE (up\_alphabetic), 31
- up\_uppercase (up\_alphabetic), 31
- UP\_WHITE\_SPACE (up\_alphabetic), 31
- up\_white\_space (up\_alphabetic), 31
  
- VAI (Unicode), 6
- vai (Unicode), 6
- VARIATION\_SELECTORS (Unicode), 6
- variation\_selectors (Unicode), 6
- VARIATION\_SELECTORS\_SUPPLEMENT  
    (Unicode), 6
- variation\_selectors\_supplement  
    (Unicode), 6
- VEDIC\_EXTENSIONS (Unicode), 6
- vedic\_extensions (Unicode), 6
- VERTICAL\_FORMS (Unicode), 6
- vertical\_forms (Unicode), 6
  
- WARANG\_CITI (Unicode), 6
  
- warang\_citi (Unicode), 6
  
- YI\_RADICALS (Unicode), 6
- yi\_radicals (Unicode), 6
- YI\_SYLLABLES (Unicode), 6
- yi\_syllables (Unicode), 6
- YIJING\_HEXAGRAM\_SYMBOLS (Unicode), 6
- yijing\_hexagram\_symbols (Unicode), 6
- YIJING\_MONO\_DI\_AND\_TRIGRAMS (Unicode), 6
- yijing\_mono\_di\_and\_trigrams (Unicode), 6