

# Package ‘rebus.unicode’

September 23, 2025

**Type** Package

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

**Version** 0.0-2.1

**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

**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** 7.3.3

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2025-09-23 12:02:20 UTC

## Contents

ugc_cased_letter . . . . .	2
Unicode . . . . .	7
UnicodeOperators . . . . .	37
up_alphabetic . . . . .	39

<b>Index</b>	<b>45</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



An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
An object of class regex (inherits from character) of length 1.  
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)  
syloti\_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\_PSALTER

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

















An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
An object of class `regex` (inherits from `character`) of length 1.  
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_*( <i>)</i> ) 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_*( <i>)</i> ) not the constants.
y	A character vector containing Unicode General Category or Unicode Properties. Use the functional forms (ugc_*( <i>)</i> ) 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



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

## \* datasets

- ugc\_cased\_letter, 2
  - Unicode, 7
  - up\_alphabetic, 39
- ADDITIONAL\_ARROWS (Unicode), 7
- additional\_arrows (Unicode), 7
- AEGEAN\_NUMBERS (Unicode), 7
- aegean\_numbers (Unicode), 7
- ALCHEMICAL\_SYMBOLS (Unicode), 7
- alchemical\_symbols (Unicode), 7
- ALPHABETIC\_PRESENTATION\_FORMS (Unicode), 7
- alphabetic\_presentation\_forms (Unicode), 7
- ANCIENT\_GREEK\_MUSICAL\_NOTATION (Unicode), 7
- ancient\_greek\_musical\_notation (Unicode), 7
- ANCIENT\_GREEK\_NUMBERS (Unicode), 7
- ancient\_greek\_numbers (Unicode), 7
- ANCIENT\_SYMBOLS (Unicode), 7
- ancient\_symbols (Unicode), 7
- ARABIC (Unicode), 7
- arabic (Unicode), 7
- ARABIC\_EXTENDED\_A (Unicode), 7
- arabic\_extended\_a (Unicode), 7
- ARABIC\_MATHEMATICAL\_ALPHANUMERIC\_SYMBOLS (Unicode), 7
- arabic\_mathematical\_alphanumeric\_symbols (Unicode), 7
- ARABIC\_PRESENTATION\_FORMS\_A (Unicode), 7
- arabic\_presentation\_forms\_a (Unicode), 7
- ARABIC\_PRESENTATION\_FORMS\_B (Unicode), 7
- arabic\_presentation\_forms\_b (Unicode), 7
- ARABIC\_SUPPLEMENT (Unicode), 7
- arabic\_supplement (Unicode), 7
- ARMENIAN (Unicode), 7
- armenian (Unicode), 7
- ARMENIAN\_LIGATURES (Unicode), 7
- armenian\_ligatures (Unicode), 7
- AVESTAN (Unicode), 7
- avestan (Unicode), 7
- BALINESE (Unicode), 7
- balinese (Unicode), 7
- BAMUN (Unicode), 7
- bamun (Unicode), 7
- BAMUN\_SUPPLEMENT (Unicode), 7
- bamun\_supplement (Unicode), 7
- BASSA\_VAH (Unicode), 7
- bassa\_vah (Unicode), 7
- BATAK (Unicode), 7
- batak (Unicode), 7
- BENGALI\_AND\_ASSAMESE (Unicode), 7
- bengali\_and\_assamese (Unicode), 7
- BLOCK\_ELEMENTS (Unicode), 7
- block\_elements (Unicode), 7
- BOPOMOFO (Unicode), 7
- bopomofo (Unicode), 7
- BOPOMOFO\_EXTENDED (Unicode), 7
- bopomofo\_extended (Unicode), 7
- BOX\_DRAWING (Unicode), 7
- box\_drawing (Unicode), 7
- BRAHMI (Unicode), 7
- brahmi (Unicode), 7
- BRAILLE\_PATTERNS (Unicode), 7
- braille\_patterns (Unicode), 7
- BUGINESE (Unicode), 7
- buginese (Unicode), 7
- BUHID (Unicode), 7
- buhid (Unicode), 7
- BYZANTINE\_MUSICAL\_SYMBOLS (Unicode), 7
- byzantine\_musical\_symbols (Unicode), 7
- CARD\_SUITS (Unicode), 7
- card\_suits (Unicode), 7
- CARIAN (Unicode), 7
- carian (Unicode), 7
- CAUCASIAN\_ALBANIAN (Unicode), 7

- caucasian\_albanian (Unicode), 7
- CHAKMA (Unicode), 7
- chakma (Unicode), 7
- CHAM (Unicode), 7
- cham (Unicode), 7
- CHEROKEE (Unicode), 7
- cherokee (Unicode), 7
- CHESS\_CHECKERS\_DRAUGHTS (Unicode), 7
- chess\_checkers draughts (Unicode), 7
- CJK\_COMPATIBILITY (Unicode), 7
- chk\_compatibility (Unicode), 7
- CJK\_COMPATIBILITY\_FORMS (Unicode), 7
- chk\_compatibility\_forms (Unicode), 7
- CJK\_COMPATIBILITY\_IDEOGRAPHS (Unicode), 7
- chk\_compatibility\_ideographs (Unicode), 7
- CJK\_COMPATIBILITY\_IDEOGRAPHS\_SUPPLEMENT (Unicode), 7
- chk\_compatibility\_ideographs\_supplement (Unicode), 7
- CJK\_IDEOGRAPHIC\_DESCRIPTION\_CHARACTERS (Unicode), 7
- chk\_ideographic\_description\_characters (Unicode), 7
- CJK\_STROKES (Unicode), 7
- chk\_strokes (Unicode), 7
- CJK\_SYMBOLS\_AND\_PUNCTUATION (Unicode), 7
- chk\_symbols\_and\_punctuation (Unicode), 7
- CJK\_UNIFIED\_IDEOGRAPHS (Unicode), 7
- chk\_unified\_ideographs (Unicode), 7
- CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_A (Unicode), 7
- chk\_unified\_ideographs\_extension\_a (Unicode), 7
- CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_B (Unicode), 7
- chk\_unified\_ideographs\_extension\_b (Unicode), 7
- CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_C (Unicode), 7
- chk\_unified\_ideographs\_extension\_c (Unicode), 7
- CJK\_UNIFIED\_IDEOGRAPHS\_EXTENSION\_D (Unicode), 7
- chk\_unified\_ideographs\_extension\_d (Unicode), 7
- ClassGroups, 37
- COMBINING\_DIACRITIC\_EXTENDED (Unicode), 7
- combining\_diacritic\_extended (Unicode), 7
- COMBINING\_DIACRITIC\_MARKS (Unicode), 7
- combining\_diacritic\_marks (Unicode), 7
- COMBINING\_DIACRITIC\_MARKS\_FOR\_SYMBOLS (Unicode), 7
- combining\_diacritic\_marks\_for\_symbols (Unicode), 7
- COMBINING\_DIACRITIC\_SUPPLEMENT (Unicode), 7
- combining\_diacritic\_supplement (Unicode), 7
- COMBINING\_HALF\_MARKS (Unicode), 7
- combining\_half\_marks (Unicode), 7
- COMMON\_INDIC\_NUMBER\_FORMS (Unicode), 7
- common\_indic\_number\_forms (Unicode), 7
- CONTROL\_PICTURES (Unicode), 7
- control\_pictures (Unicode), 7
- COPTIC (Unicode), 7
- coptic (Unicode), 7
- COPTIC\_EPACT\_NUMBERS (Unicode), 7
- coptic\_epact\_numbers (Unicode), 7
- COUNTING\_ROD\_NUMERALS (Unicode), 7
- counting\_rod\_numerals (Unicode), 7
- CUNEIFORM (Unicode), 7
- cuneiform (Unicode), 7
- CUNEIFORM\_NUMBERS\_AND\_PUNCTUATION (Unicode), 7
- cuneiform\_numbers\_and\_punctuation (Unicode), 7
- CURRENCY\_SYMBOLS (Unicode), 7
- currency\_symbols (Unicode), 7
- CYPRIOT\_SYLLABARY (Unicode), 7
- cyriot\_syllabary (Unicode), 7
- CYRILLIC (Unicode), 7
- cyrillic (Unicode), 7
- CYRILLIC\_EXTENDED\_A (Unicode), 7
- cyrillic\_extended\_a (Unicode), 7
- CYRILLIC\_EXTENDED\_B (Unicode), 7
- cyrillic\_extended\_b (Unicode), 7
- CYRILLIC\_SUPPLEMENT (Unicode), 7
- cyrillic\_supplement (Unicode), 7
- DESERET (Unicode), 7
- deseret (Unicode), 7
- DEVANAGARI (Unicode), 7
- devanagari (Unicode), 7

- DEVANAGARI\_EXTENDED (Unicode), 7
- devanagari\_extended (Unicode), 7
- DINGBATS (Unicode), 7
- dingbats (Unicode), 7
- DOMINO\_TILES (Unicode), 7
- domino\_tiles (Unicode), 7
- DUPLOYAN (Unicode), 7
- duployan (Unicode), 7
  
- EGYPTIAN\_HIEROGLYPHS (Unicode), 7
- egyptian\_hieroglyphs (Unicode), 7
- ELBASAN (Unicode), 7
- elbasan (Unicode), 7
- EMOTICONS (Unicode), 7
- emoticons (Unicode), 7
- ENCLOSED\_ALPHANUMERIC\_SUPPLEMENT (Unicode), 7
- enclosed\_alphanumeric\_supplement (Unicode), 7
- ENCLOSED\_ALPHANUMERICS (Unicode), 7
- enclosed\_alphanumerics (Unicode), 7
- ENCLOSED\_CJK\_LETTERS\_AND\_MONTHS (Unicode), 7
- enclosed\_cjk\_letters\_and\_months (Unicode), 7
- ENCLOSED\_IDEOGRAPHIC\_SUPPLEMENT (Unicode), 7
- enclosed\_ideographic\_supplement (Unicode), 7
- ETHIOPIA (Unicode), 7
- ethiopic (Unicode), 7
- ETHIOPIA\_EXTENDED (Unicode), 7
- ethiopic\_extended (Unicode), 7
- ETHIOPIA\_EXTENDED\_A (Unicode), 7
- ethiopic\_extended\_a (Unicode), 7
- ETHIOPIA\_SUPPLEMENT (Unicode), 7
- ethiopic\_supplement (Unicode), 7
  
- FLOORS\_AND\_CEILINGS (Unicode), 7
- floors\_and\_ceilings (Unicode), 7
- FULLWIDTH\_ASCII\_DIGITS (Unicode), 7
- fullwidth\_ascii\_digits (Unicode), 7
- FULLWIDTH\_ASCII\_PUNCTUATION (Unicode), 7
- fullwidth\_ascii\_punctuation (Unicode), 7
  
- GENERAL\_PUNCTUATION (Unicode), 7
- general\_punctuation (Unicode), 7
- GEOMETRIC\_SHAPES (Unicode), 7
- geometric\_shapes (Unicode), 7
  
- GEOMETRIC\_SHAPES\_EXTENDED (Unicode), 7
- geometric\_shapes\_extended (Unicode), 7
- GEORGIAN (Unicode), 7
- georgian (Unicode), 7
- GEORGIAN\_SUPPLEMENT (Unicode), 7
- georgian\_supplement (Unicode), 7
- GLAGOLITIC (Unicode), 7
- glagolitic (Unicode), 7
- GOTHIC (Unicode), 7
- gothic (Unicode), 7
- GRANTHA (Unicode), 7
- grantha (Unicode), 7
- GREEK\_AND\_COPTIC (Unicode), 7
- greek\_and\_coptic (Unicode), 7
- GREEK\_EXTENDED (Unicode), 7
- greek\_extended (Unicode), 7
- GUJARATI (Unicode), 7
- gujarati (Unicode), 7
- GURMUKHI (Unicode), 7
- gurmukhi (Unicode), 7
  
- HALFWIDTH\_AND\_FULLWIDTH\_FORMS (Unicode), 7
- halfwidth\_and\_fullwidth\_forms (Unicode), 7
- HANGUL\_COMPATIBILITY\_JAMO (Unicode), 7
- hangul\_compatibility\_jamo (Unicode), 7
- HANGUL\_JAMO (Unicode), 7
- hangul\_jamo (Unicode), 7
- HANGUL\_JAMO\_EXTENDED\_A (Unicode), 7
- hangul\_jamo\_extended\_a (Unicode), 7
- HANGUL\_JAMO\_EXTENDED\_B (Unicode), 7
- hangul\_jamo\_extended\_b (Unicode), 7
- HANGUL\_SYLLABLES (Unicode), 7
- hangul\_syllables (Unicode), 7
- HANUNOO (Unicode), 7
- hanunoo (Unicode), 7
- HEBREW (Unicode), 7
- hebrew (Unicode), 7
- HIRAGANA (Unicode), 7
- hiragana (Unicode), 7
  
- IMPERIAL\_ARAMAIC (Unicode), 7
- imperial\_aramaic (Unicode), 7
- INVISIBLE\_OPERATORS (Unicode), 7
- invisible\_operators (Unicode), 7
- IPA\_EXTENSIONS (Unicode), 7
- ipa\_extensions (Unicode), 7

- JAPANESE\_CHESS (Unicode), 7  
 japanese\_chess (Unicode), 7  
 JAVANESE (Unicode), 7  
 javanese (Unicode), 7  
  
 KAITHI (Unicode), 7  
 kaithi (Unicode), 7  
 KANA\_SUPPLEMENT (Unicode), 7  
 kana\_supplement (Unicode), 7  
 KANBUN (Unicode), 7  
 kanbun (Unicode), 7  
 KANGXI\_RADICALS (Unicode), 7  
 kangxi\_radicals (Unicode), 7  
 KANGXI\_RADICALS\_SUPPLEMENT (Unicode), 7  
 kangxi\_radicals\_supplement (Unicode), 7  
 KANNADA (Unicode), 7  
 kannada (Unicode), 7  
 KATAKANA (Unicode), 7  
 katakana (Unicode), 7  
 KATAKANA\_PHONETIC\_EXTENSIONS (Unicode),  
 7  
 katakana\_phonetic\_extensions (Unicode),  
 7  
 KAYAH\_LI (Unicode), 7  
 kayah\_li (Unicode), 7  
 KHAROSHTHI (Unicode), 7  
 kharoshthi (Unicode), 7  
 KHMER (Unicode), 7  
 khmer (Unicode), 7  
 KHMER\_SYMBOLS (Unicode), 7  
 khmer\_symbols (Unicode), 7  
 KHOJKI (Unicode), 7  
 khojki (Unicode), 7  
 KHUDAWADI (Unicode), 7  
 khudawadi (Unicode), 7  
  
 LAO (Unicode), 7  
 lao (Unicode), 7  
 LATIN (Unicode), 7  
 latin (Unicode), 7  
 LATIN\_1\_PUNCTUATION (Unicode), 7  
 latin\_1\_punctuation (Unicode), 7  
 LATIN\_1\_SUPPLEMENT (Unicode), 7  
 latin\_1\_supplement (Unicode), 7  
 LATIN\_EXTENDED\_A (Unicode), 7  
 latin\_extended\_a (Unicode), 7  
 LATIN\_EXTENDED\_ADDITIONAL (Unicode), 7  
 latin\_extended\_additional (Unicode), 7  
 LATIN\_EXTENDED\_B (Unicode), 7  
 latin\_extended\_b (Unicode), 7  
 LATIN\_EXTENDED\_C (Unicode), 7  
 latin\_extended\_c (Unicode), 7  
 LATIN\_EXTENDED\_D (Unicode), 7  
 latin\_extended\_d (Unicode), 7  
 LATIN\_EXTENDED\_E (Unicode), 7  
 latin\_extended\_e (Unicode), 7  
 LATIN\_LIGATURES (Unicode), 7  
 latin\_ligatures (Unicode), 7  
 LEPCHA (Unicode), 7  
 lepcha (Unicode), 7  
 LETTERLIKE\_SYMBOLS (Unicode), 7  
 letterlike\_symbols (Unicode), 7  
 LIMBU (Unicode), 7  
 limbu (Unicode), 7  
 LINEAR\_A (Unicode), 7  
 linear\_a (Unicode), 7  
 LINEAR\_B\_IDEOGRAMS (Unicode), 7  
 linear\_b\_ideograms (Unicode), 7  
 LINEAR\_B\_SYLLABARY (Unicode), 7  
 linear\_b\_syllabary (Unicode), 7  
 LISU (Unicode), 7  
 lisu (Unicode), 7  
 LYCIAN (Unicode), 7  
 lycian (Unicode), 7  
 LYDIAN (Unicode), 7  
 lydian (Unicode), 7  
  
 MAHAJANI (Unicode), 7  
 mahajani (Unicode), 7  
 MAHJONG\_TILES (Unicode), 7  
 mahjong\_tiles (Unicode), 7  
 MALAYALAM (Unicode), 7  
 malayalam (Unicode), 7  
 MANDAIC (Unicode), 7  
 mandaic (Unicode), 7  
 MANICHAEAN (Unicode), 7  
 manichaeian (Unicode), 7  
 MATH\_ARROWS (Unicode), 7  
 math\_arrows (Unicode), 7  
 MATHEMATICAL\_ALPHANUMERIC\_SYMBOLS  
 (Unicode), 7  
 mathematical\_alphanumeric\_symbols  
 (Unicode), 7  
 MEETEI\_MAYEK (Unicode), 7  
 meetei\_mayek (Unicode), 7  
 MEETEI\_MAYEK\_EXTENSIONS (Unicode), 7  
 meetei\_mayek\_extensions (Unicode), 7  
 MENDE\_KIKAKUI (Unicode), 7



- mende\_kikakui (Unicode), 7
- MEROITIC\_CURSIVE (Unicode), 7
- meroitic\_cursive (Unicode), 7
- MEROITIC\_HIEROGLYPHS (Unicode), 7
- meroitic\_hieroglyphs (Unicode), 7
- MIAO (Unicode), 7
- miao (Unicode), 7
- MISCELLANEOUS\_MATHEMATICAL\_SYMBOLS\_A (Unicode), 7
- miscellaneous\_mathematical\_symbols\_a (Unicode), 7
- MISCELLANEOUS\_MATHEMATICAL\_SYMBOLS\_B (Unicode), 7
- miscellaneous\_mathematical\_symbols\_b (Unicode), 7
- MISCELLANEOUS\_SYMBOLS\_AND\_PICTOGRAPHS (Unicode), 7
- miscellaneous\_symbols\_and\_pictographs (Unicode), 7
- MISCELLANEOUS\_TECHNICAL (Unicode), 7
- miscellaneous\_technical (Unicode), 7
- MODI (Unicode), 7
- modi (Unicode), 7
- MODIFIER\_TONE\_LETTERS (Unicode), 7
- modifier\_tone\_letters (Unicode), 7
- MONGOLIAN (Unicode), 7
- mongolian (Unicode), 7
- MRO (Unicode), 7
- mro (Unicode), 7
- MUSICAL\_SYMBOLS (Unicode), 7
- musical\_symbols (Unicode), 7
- MYANMAR (Unicode), 7
- myanmar (Unicode), 7
- MYANMAR\_EXTENDED\_A (Unicode), 7
- myanmar\_extended\_a (Unicode), 7
- MYANMAR\_EXTENDED\_B (Unicode), 7
- myanmar\_extended\_b (Unicode), 7
  
- NABATAEAN (Unicode), 7
- nabataean (Unicode), 7
- NEW\_TAI\_LUE (Unicode), 7
- new\_tai\_lue (Unicode), 7
- NKO (Unicode), 7
- nko (Unicode), 7
- NUMBER\_FORMS (Unicode), 7
- number\_forms (Unicode), 7
  
- OGHAM (Unicode), 7
- ogham (Unicode), 7
  
- OL\_CHIKI (Unicode), 7
- ol\_chiki (Unicode), 7
- OLD\_ITALIC (Unicode), 7
- old\_italic (Unicode), 7
- OLD\_NORTH\_ARABIAN (Unicode), 7
- old\_north\_arabian (Unicode), 7
- OLD\_PERMIC (Unicode), 7
- old\_permic (Unicode), 7
- OLD\_PERSIAN (Unicode), 7
- old\_persian (Unicode), 7
- OLD\_SOUTH\_ARABIAN (Unicode), 7
- old\_south\_arabian (Unicode), 7
- OLD\_TURKIC (Unicode), 7
- old\_turkic (Unicode), 7
- OPTICAL\_CHARACTER\_RECOGNITION (Unicode), 7
- optical\_character\_recognition (Unicode), 7
- ORIYA (Unicode), 7
- oriya (Unicode), 7
- ORNAMENTAL\_DINGBATS (Unicode), 7
- ornamental\_dingbats (Unicode), 7
- OSMANYA (Unicode), 7
- osmanya (Unicode), 7
  
- PAHAHW\_HMONG (Unicode), 7
- pahawh\_hmong (Unicode), 7
- PAHLAVI\_INSCRIPTIONAL (Unicode), 7
- pahlavi\_inscriptional (Unicode), 7
- PAHLAVI\_PSALTER (Unicode), 7
- pahlavi\_psalter (Unicode), 7
- PALMYRENE (Unicode), 7
- palmyrene (Unicode), 7
- PAU\_CIN\_HAU (Unicode), 7
- pau\_cin\_hau (Unicode), 7
- PHAGS\_PA (Unicode), 7
- phags\_pa (Unicode), 7
- PHAISTOS\_DISC (Unicode), 7
- phaistos\_disc (Unicode), 7
- PHOENICIAN (Unicode), 7
- phoenician (Unicode), 7
- PHONETIC\_EXTENSIONS (Unicode), 7
- phonetic\_extensions (Unicode), 7
- PHONETIC\_EXTENSIONS\_SUPPLEMENT (Unicode), 7
- phonetic\_extensions\_supplement (Unicode), 7
- PLAYING\_CARDS (Unicode), 7
- playing\_cards (Unicode), 7

- PRIVATE\_USE\_AREA (Unicode), 7  
private\_use\_area (Unicode), 7
- REJANG (Unicode), 7  
rejang (Unicode), 7
- RUMI\_NUMERAL\_SYMBOLS (Unicode), 7  
rumi\_numeral\_symbols (Unicode), 7
- RUNIC (Unicode), 7  
runic (Unicode), 7
- SAMARITAN (Unicode), 7  
samaritan (Unicode), 7
- SAURASHTRA (Unicode), 7  
saurashtra (Unicode), 7
- SHARADA (Unicode), 7  
sharada (Unicode), 7
- SHAVIAN (Unicode), 7  
shavian (Unicode), 7
- SHORTHAND\_FORMAT\_CONTROLS (Unicode), 7  
shorthand\_format\_controls (Unicode), 7
- SIDDHAM (Unicode), 7  
siddham (Unicode), 7
- SINHALA (Unicode), 7  
sinhala (Unicode), 7
- SINHALA\_ARCHAIC\_NUMBERS (Unicode), 7  
sinhala\_archaic\_numbers (Unicode), 7
- SMALL\_FORM\_VARIANTS (Unicode), 7  
small\_form\_variants (Unicode), 7
- SORA\_SOMPENG (Unicode), 7  
sora\_sompeng (Unicode), 7
- SPACING\_MODIFIER\_LETTERS (Unicode), 7  
spacing\_modifier\_letters (Unicode), 7
- SPECIALS (Unicode), 7  
specials (Unicode), 7
- SUNDANESE (Unicode), 7  
sundanese (Unicode), 7
- SUNDANESE\_SUPPLEMENT (Unicode), 7  
sundanese\_supplement (Unicode), 7
- SUPERSCRIPTS\_AND\_SUBSCRIPTS (Unicode), 7  
superscripts\_and\_subscripts (Unicode), 7
- SUPPLEMENTAL\_ARROWS\_A (Unicode), 7  
supplemental\_arrows\_a (Unicode), 7
- SUPPLEMENTAL\_MATHEMATICAL\_OPERATORS (Unicode), 7  
supplemental\_mathematical\_operators (Unicode), 7
- SUPPLEMENTAL\_PUNCTUATION (Unicode), 7  
supplemental\_punctuation (Unicode), 7
- SUPPLEMENTARY\_PRIVATE\_USE\_AREA\_A (Unicode), 7  
supplementary\_private\_use\_area\_a (Unicode), 7
- SUPPLEMENTARY\_PRIVATE\_USE\_AREA\_B (Unicode), 7  
supplementary\_private\_use\_area\_b (Unicode), 7
- SYLOTI\_NAGRI (Unicode), 7  
syloiti\_nagri (Unicode), 7
- SYRIAC (Unicode), 7  
syriac (Unicode), 7
- TAGALOG (Unicode), 7  
tagalog (Unicode), 7
- TAGBANWA (Unicode), 7  
tagbanwa (Unicode), 7
- TAGS (Unicode), 7  
tags (Unicode), 7
- TAI\_LE (Unicode), 7  
tai\_le (Unicode), 7
- TAI\_THAM (Unicode), 7  
tai\_tham (Unicode), 7
- TAI\_VIET (Unicode), 7  
tai\_viet (Unicode), 7
- TAI\_XUAN\_JING\_SYMBOLS (Unicode), 7  
tai\_xuan\_jing\_symbols (Unicode), 7
- TAKRI (Unicode), 7  
takri (Unicode), 7
- TAMIL (Unicode), 7  
tamil (Unicode), 7
- TELUGU (Unicode), 7  
telugu (Unicode), 7
- THAANA (Unicode), 7  
thaana (Unicode), 7
- THAI (Unicode), 7  
thai (Unicode), 7
- TIBETAN (Unicode), 7  
tibetan (Unicode), 7
- TIFINAGH (Unicode), 7  
tifinagh (Unicode), 7
- TIRHUTA (Unicode), 7  
tirhuta (Unicode), 7
- TRANSPORT\_AND\_MAP\_SYMBOLS (Unicode), 7  
transport\_and\_map\_symbols (Unicode), 7
- UGARITIC (Unicode), 7  
ugaritic (Unicode), 7
- 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, 7, 7, 43
- unicode\_general\_category, 43
- unicode\_general\_category
  - (ugc\_cased\_letter), 2
- unicode\_intersect (UnicodeOperators), 37
- unicode\_inverse (UnicodeOperators), 37
- unicode\_operators (UnicodeOperators), 37
- unicode\_property, 7
- unicode\_property (up\_alphabetic), 39
- unicode\_setdiff (UnicodeOperators), 37
- unicode\_union (UnicodeOperators), 37
- UnicodeGeneralCategory
  - (ugc\_cased\_letter), 2
- UnicodeOperators, 37
- UnicodeProperty (up\_alphabetic), 39
- UNIFIED\_CANADIAN\_ABORIGINAL\_SYLLABICS
  - (Unicode), 7
- unified\_canadian\_aboriginal\_syllabics
  - (Unicode), 7
- UNIFIED\_CANADIAN\_ABORIGINAL\_SYLLABICS\_EXTENDED
  - (Unicode), 7
- unified\_canadian\_aboriginal\_syllabics\_extended
  - (Unicode), 7
- UP\_ALPHABETIC (up\_alphabetic), 39
- up\_alphabetic, 39
- UP\_ASCII\_HEX\_DIGIT (up\_alphabetic), 39
- up\_ascii\_hex\_digit (up\_alphabetic), 39
- UP\_BIDI\_CONTROL (up\_alphabetic), 39
- up\_bidi\_control (up\_alphabetic), 39
- UP\_BIDI\_MIRRORED (up\_alphabetic), 39
- up\_bidi\_mirrored (up\_alphabetic), 39
- UP\_CASE\_IGNOREABLE (up\_alphabetic), 39
- up\_case\_ignorable (up\_alphabetic), 39
- UP\_CASE\_SENSITIVE (up\_alphabetic), 39
- up\_case\_sensitive (up\_alphabetic), 39
- UP\_CASED (up\_alphabetic), 39
- up\_cased (up\_alphabetic), 39
- UP\_CHANGES\_WHEN\_CASEFOLDED
  - (up\_alphabetic), 39
- up\_changes\_when\_casefolded
  - (up\_alphabetic), 39
- UP\_CHANGES\_WHEN\_CASEMAPPED
  - (up\_alphabetic), 39
- up\_changes\_when\_casemapped
  - (up\_alphabetic), 39
- UP\_CHANGES\_WHEN\_LOWERCASED
  - (up\_alphabetic), 39
- up\_changes\_when\_lowercased
  - (up\_alphabetic), 39
- UP\_CHANGES\_WHEN\_NFKC\_CASEFOLDED
  - (up\_alphabetic), 39
- up\_changes\_when\_nfkc\_casefolded
  - (up\_alphabetic), 39
- UP\_CHANGES\_WHEN\_TITLECASED
  - (up\_alphabetic), 39
- up\_changes\_when\_titlecased
  - (up\_alphabetic), 39
- UP\_CHANGES\_WHEN\_UPPERCASED
  - (up\_alphabetic), 39
- up\_changes\_when\_uppercased
  - (up\_alphabetic), 39
- UP\_DASH (up\_alphabetic), 39
- up\_dash (up\_alphabetic), 39
- UP\_DEFAULT\_IGNOREABLE\_CODE\_POINT
  - (up\_alphabetic), 39
- up\_default\_ignorable\_code\_point
  - (up\_alphabetic), 39
- UP\_DEPRECATED (up\_alphabetic), 39
- up\_deprecated (up\_alphabetic), 39
- UP\_DIACRITIC (up\_alphabetic), 39
- up\_diacritic (up\_alphabetic), 39
- UP\_EXTENDER (up\_alphabetic), 39
- up\_extender (up\_alphabetic), 39
- UP\_HEX\_DIGIT (up\_alphabetic), 39
- up\_hex\_digit (up\_alphabetic), 39
- UP\_HYPHEN (up\_alphabetic), 39
- up\_hyphen (up\_alphabetic), 39
- UP\_ID\_CONTINUE (up\_alphabetic), 39
- up\_id\_continue (up\_alphabetic), 39

- UP\_ID\_START (up\_alphabetic), 39
- up\_id\_start (up\_alphabetic), 39
- UP\_IDEOGRAPHIC (up\_alphabetic), 39
- up\_ideographic (up\_alphabetic), 39
- UP\_LOWERCASE (up\_alphabetic), 39
- up\_lowercase (up\_alphabetic), 39
- UP\_MATH (up\_alphabetic), 39
- up\_math (up\_alphabetic), 39
- UP\_NONCHARACTER\_CODE\_POINT  
(up\_alphabetic), 39
- up\_noncharacter\_code\_point  
(up\_alphabetic), 39
- UP\_POSIX\_ALNUM (up\_alphabetic), 39
- up\_posix\_alnum (up\_alphabetic), 39
- UP\_POSIX\_BLANK (up\_alphabetic), 39
- up\_posix\_blank (up\_alphabetic), 39
- UP\_POSIX\_GRAPH (up\_alphabetic), 39
- up\_posix\_graph (up\_alphabetic), 39
- UP\_POSIX\_PRINT (up\_alphabetic), 39
- up\_posix\_print (up\_alphabetic), 39
- UP\_POSIX\_XDIGIT (up\_alphabetic), 39
- up\_posix\_xdigit (up\_alphabetic), 39
- UP\_QUOTATION\_MARK (up\_alphabetic), 39
- up\_quotation\_mark (up\_alphabetic), 39
- UP\_SOFT\_DOTTED (up\_alphabetic), 39
- up\_soft\_dotted (up\_alphabetic), 39
- UP\_TERMINAL\_PUNCTUATION  
(up\_alphabetic), 39
- up\_terminal\_punctuation  
(up\_alphabetic), 39
- UP\_UPPERCASE (up\_alphabetic), 39
- up\_uppercase (up\_alphabetic), 39
- UP\_WHITE\_SPACE (up\_alphabetic), 39
- up\_white\_space (up\_alphabetic), 39
  
- VAI (Unicode), 7
- vai (Unicode), 7
- VARIATION\_SELECTORS (Unicode), 7
- variation\_selectors (Unicode), 7
- VARIATION\_SELECTORS\_SUPPLEMENT  
(Unicode), 7
- variation\_selectors\_supplement  
(Unicode), 7
- VEDIC\_EXTENSIONS (Unicode), 7
- vedic\_extensions (Unicode), 7
- VERTICAL\_FORMS (Unicode), 7
- vertical\_forms (Unicode), 7
  
- WARANG\_CITI (Unicode), 7
  
- warang\_citi (Unicode), 7
  
- YI\_RADICALS (Unicode), 7
- yi\_radicals (Unicode), 7
- YI\_SYLLABLES (Unicode), 7
- yi\_syllables (Unicode), 7
- YIJING\_HEXAGRAM\_SYMBOLS (Unicode), 7
- yijing\_hexagram\_symbols (Unicode), 7
- YIJING\_MONO\_DI\_AND\_TRIGRAMS (Unicode), 7
- yijing\_mono\_di\_and\_trigrams (Unicode), 7