

Package 'rebus.unicode'

May 9, 2026

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

Index	45
--------------	-----------

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

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_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

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())

```

UnicodeOperators

Unicode Pattern Operators

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
 manichaeen (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