

Package ‘rebus’

April 25, 2017

Type Package

Title Build Regular Expressions in a Human Readable Way

Version 0.1-3

Date 2017-04-25

Author Richard Cotton [aut, cre]

Maintainer Richard Cotton <richierocks@gmail.com>

Description Build regular expressions piece by piece using human readable code.
This package is designed for interactive use. For package development, use the rebus.* dependencies.

Depends R (>= 3.1.0)

Imports rebus.base (>= 0.0-3), rebus.datetimes, rebus.numbers,
rebus.unicode (>= 0.0-2)

Suggests testthat

License Unlimited

LazyLoad yes

LazyData yes

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

RoxygenNote 6.0.1

Collate 'export-base.R' 'export-datetimes.R' 'export-numbers.R'
'export-unicode.R' 'imports.R' 'regex-package.R'

NeedsCompilation no

Repository CRAN

Date/Publication 2017-04-25 21:42:46 UTC

R topics documented:

Anchors	2
as.regex	3
Backreferences	3
capture	3
CharacterClasses	3
char_class	3
ClassGroups	4
Concatenation	4
DateTime	4
escape_special	4
exactly	4
format.regex	5
get_weekdays	5
IsoClasses	5
literal	5
lookahead	5
modify_mode	6
number_range	6
or	6
rebus	6
recursive	8
regex	8
repeated	8
ReplacementCase	8
roman	9
SpecialCharacters	9
Unicode	9
UnicodeGeneralCategory	9
UnicodeOperators	9
UnicodeProperty	10
whole_word	10
WordBoundaries	10
Index	11

Anchors

The start or end of a string

Description
See [Anchors](#).

as.regex	<i>Convert or test for regex objects</i>
----------	--

Description

See [as.regex](#).

Backreferences	<i>Backreferences</i>
----------------	-----------------------

Description

See [Backreferences](#).

capture	<i>Capture a token, or not</i>
---------	--------------------------------

Description

See [capture](#).

CharacterClasses	<i>Class Constants</i>
------------------	------------------------

Description

See [CharacterClasses](#).

char_class	<i>A range or char_class of characters</i>
------------	--

Description

See [char_class](#).

ClassGroups	<i>Character classes</i>
-------------	--------------------------

Description

See [ClassGroups](#).

Concatenation	<i>Combine strings together</i>
---------------	---------------------------------

Description

See [Concatenation](#).

DateTime	<i>Date-time regexes</i>
----------	--------------------------

Description

See [DateTime](#).

escape_special	<i>Escape special characters</i>
----------------	----------------------------------

Description

See [escape_special](#).

exactly	<i>Make a regex exact</i>
---------	---------------------------

Description

See [exactly](#).

format.regex	<i>Print or format regex objects</i>
--------------	--------------------------------------

Description

See [format.regex](#).

get_weekdays	<i>Get the days of the week or months of the year</i>
--------------	---

Description

See [get_weekdays](#).

IsoClasses	<i>ISO 8601 date-time classes</i>
------------	-----------------------------------

Description

See [IsoClasses](#).

literal	<i>Treat part of a regular expression literally</i>
---------	---

Description

See [literal](#).

lookahead	<i>Lookaround</i>
-----------	-------------------

Description

See [lookahead](#).

modify_mode	<i>Apply mode modifiers</i>
-------------	-----------------------------

Description

See [modify_mode](#).

number_range	<i>Generate a regular expression for a number range</i>
--------------	---

Description

See [number_range](#).

or	<i>Alternation</i>
----	--------------------

Description

See [or](#).

rebus	<i>rebus: Regular Expression Builder, Um, Something</i>
-------	---

Description

Build regular expressions in a human readable way.

Details

Regular expressions are a very powerful tool, but the syntax is terse enough to be difficult to read. This makes bugs easy to introduce, and hard to find. This package contains functions to make building regular expressions easier.

Author(s)

Richard Cotton <richierocks@gmail.com>

See Also

`regex` and `regexpr` The `'stringr'` and `'stringi'` packages provide tools for matching regular expressions and nicely complement this package. <http://www.regular-expressions.info> has good advice on using regular expression in R. In particular, see <http://www.regular-expressions.info/rlanguage.html> and <http://www.regular-expressions.info/examples.html> <https://www.debuggex.com> is a visual regex debugging and testing site.

Examples

```
### Match a hex colour, like `"#99af01"`
# This reads *Match a hash, followed by six hexadecimal values.*

"#" %R% hex_digit(6)

# To match only a hex colour and nothing else, you can add anchors to the
# start and end of the expression.

START %R% "#" %R% hex_digit(6) %R% END

### Simple email address matching.
# This reads *Match one or more letters, numbers, dots, underscores, percents,
# plusses or hyphens. Then match an 'at' symbol. Then match one or more letters,
# numbers, dots, or hyphens. Then match a dot. Then match two to four letters.*

one_or_more(char_class(ASCII_ALNUM %R% "._%+-")) %R%
"@ " %R%
one_or_more(char_class(ASCII_ALNUM %R% ".-")) %R%
DOT %R%
ascii_alpha(2, 4)

### IP address matching.
# First we need an expression to match numbers between 0 and 255. Both the
# following syntaxes read *Match two then five then a number between zero and
# five. Or match two then a number between zero and four then a digit. Or match
# an optional zero or one followed by an optional digit folowed by a compulsory
# digit. Make this a single token, but don't capture it.*

# Using the %|% operator
ip_element <- group(
  "25" %R% char_range(0, 5) %|%
  "2" %R% char_range(0, 4) %R% ascii_digit() %|%
  optional(char_class("01")) %R% optional(ascii_digit()) %R% ascii_digit()
)

# The same again, this time using the or function
ip_element <- or(
  "25" %R% char_range(0, 5),
  "2" %R% char_range(0, 4) %R% ascii_digit(),
  optional(char_class("01")) %R% optional(ascii_digit()) %R% ascii_digit()
)

# It's easier to write using number_range, though it isn't quite as optimal
```

```
# as handcrafted regexes.
number_range(0, 255, allow_leading_zeroes = TRUE)

# Now an IP address consists of 4 of these numbers separated by dots. This
# reads *Match a word boundary. Then create a token from an `ip_element`
# followed by a dot, and repeat it three times. Then match another `ip_element`
# followed by a word boundary.*

BOUNDARY %R%
  repeated(group(ip_element %R% DOT), 3) %R%
  ip_element %R%
BOUNDARY
```

recursive	<i>Make the regular expression recursive.</i>
-----------	---

Description

See [recursive](#).

regex	<i>Create a regex</i>
-------	-----------------------

Description

See [regex](#).

repeated	<i>Repeat values</i>
----------	----------------------

Description

See [repeated](#).

ReplacementCase	<i>Force the case of replacement values</i>
-----------------	---

Description

See [ReplacementCase](#).

roman	<i>Roman numerals</i>
-------	-----------------------

Description

See [roman](#).

SpecialCharacters	<i>Special characters</i>
-------------------	---------------------------

Description

See [SpecialCharacters](#).

Unicode	<i>Unicode classes</i>
---------	------------------------

Description

See [Unicode](#).

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

Description

See [UnicodeGeneralCategory](#).

UnicodeOperators	<i>Unicode Operators</i>
------------------	--------------------------

Description

See [UnicodeOperators](#).

UnicodeProperty	<i>Unicode Properties</i>
-----------------	---------------------------

Description

See [UnicodeProperty](#).

whole_word	<i>Match a whole word</i>
------------	---------------------------

Description

See [whole_word](#).

WordBoundaries	<i>Word boundaries</i>
----------------	------------------------

Description

See [WordBoundaries](#).

Index

`%R%` (Concatenation), 4
`%c%` (Concatenation), 4

`ADDITIONAL_ARROWS` (Unicode), 9
`additional_arrows` (Unicode), 9
`AEGEAN_NUMBERS` (Unicode), 9
`aegean_numbers` (Unicode), 9
`ALCHEMICAL_SYMBOLS` (Unicode), 9
`alchemical_symbols` (Unicode), 9
`ALNUM` (CharacterClasses), 3
`alnum` (ClassGroups), 4
`ALPHA` (CharacterClasses), 3
`alpha` (ClassGroups), 4
`ALPHABETIC_PRESENTATION_FORMS` (Unicode), 9
`alphabetic_presentation_forms` (Unicode), 9
`AM_PM` (DateTime), 4
 `Anchors`, 2, 2
`ANCIENT_GREEK_MUSICAL_NOTATION` (Unicode), 9
`ancient_greek_musical_notation` (Unicode), 9
`ANCIENT_GREEK_NUMBERS` (Unicode), 9
`ancient_greek_numbers` (Unicode), 9
`ANCIENT_SYMBOLS` (Unicode), 9
`ancient_symbols` (Unicode), 9
`ANY_CHAR` (CharacterClasses), 3
`any_char` (ClassGroups), 4
`ARABIC` (Unicode), 9
`arabic` (Unicode), 9
`ARABIC_EXTENDED_A` (Unicode), 9
`arabic_extended_a` (Unicode), 9
`ARABIC_MATHEMATICAL_ALPHANUMERIC_SYMBOLS` (Unicode), 9
`arabic_mathematical_alphanumeric_symbols` (Unicode), 9
`ARABIC_PRESENTATION_FORMS_A` (Unicode), 9
`arabic_presentation_forms_a` (Unicode), 9
`ARABIC_PRESENTATION_FORMS_B` (Unicode), 9
`arabic_presentation_forms_b` (Unicode), 9
`ARABIC_SUPPLEMENT` (Unicode), 9
`arabic_supplement` (Unicode), 9
`ARMENIAN` (Unicode), 9
`armenian` (Unicode), 9
`ARMENIAN_LIGATURES` (Unicode), 9
`armenian_ligatures` (Unicode), 9
`as.regex`, 3, 3
`as_lower` (ReplacementCase), 8
`as_upper` (ReplacementCase), 8
`ASCII_ALNUM` (CharacterClasses), 3
`ascii_alnum` (ClassGroups), 4
`ASCII_ALPHA` (CharacterClasses), 3
`ascii_alpha` (ClassGroups), 4
`ASCII_DIGIT` (CharacterClasses), 3
`ascii_digit` (ClassGroups), 4
`ASCII_LOWER` (CharacterClasses), 3
`ascii_lower` (ClassGroups), 4
`ASCII_UPPER` (CharacterClasses), 3
`ascii_upper` (ClassGroups), 4
`AVESTAN` (Unicode), 9
`avestan` (Unicode), 9

Backreferences, 3, 3
`BACKSLASH` (SpecialCharacters), 9
`BALINESE` (Unicode), 9
`balinese` (Unicode), 9
`BAMUN` (Unicode), 9
`bamun` (Unicode), 9
`BAMUN_SUPPLEMENT` (Unicode), 9
`bamun_supplement` (Unicode), 9
`BASSA_VAH` (Unicode), 9
`bassa_vah` (Unicode), 9
`BATAK` (Unicode), 9
`batak` (Unicode), 9
`BENGALI_AND_ASSAMESE` (Unicode), 9
`bengali_and_assamese` (Unicode), 9
`BLANK` (CharacterClasses), 3
`blank` (ClassGroups), 4
`BLOCK_ELEMENTS` (Unicode), 9

- block_elements (Unicode), 9
- BOPOMOFO (Unicode), 9
- bopomofo (Unicode), 9
- BOPOMOFO_EXTENDED (Unicode), 9
- bopomofo_extended (Unicode), 9
- BOUNDARY (WordBoundaries), 10
- BOX_DRAWING (Unicode), 9
- box_drawing (Unicode), 9
- BRAHMI (Unicode), 9
- brahmi (Unicode), 9
- BRAILLE_PATTERNS (Unicode), 9
- braille_patterns (Unicode), 9
- BUGINESE (Unicode), 9
- buginese (Unicode), 9
- BUHID (Unicode), 9
- buhid (Unicode), 9
- BYZANTINE_MUSICAL_SYMBOLS (Unicode), 9
- byzantine_musical_symbols (Unicode), 9

- capture, 3, 3
- CARD_SUITS (Unicode), 9
- card_suits (Unicode), 9
- CARET (SpecialCharacters), 9
- CARIAN (Unicode), 9
- carian (Unicode), 9
- case_insensitive (modify_mode), 6
- CAUCASIAN_ALBANIAN (Unicode), 9
- caucasian_albanian (Unicode), 9
- CENTURY (DateTime), 4
- CENTURY_IN (DateTime), 4
- CHAKMA (Unicode), 9
- chakma (Unicode), 9
- CHAM (Unicode), 9
- cham (Unicode), 9
- char_class, 3, 3
- char_range (ClassGroups), 4
- CharacterClasses, 3, 3
- CHEROKEE (Unicode), 9
- cherokee (Unicode), 9
- CHESS_CHECKERS_DRAUGHTS (Unicode), 9
- chess_checkers draughts (Unicode), 9
- CJK_COMPATIBILITY (Unicode), 9
- cjk_compatibility (Unicode), 9
- CJK_COMPATIBILITY_FORMS (Unicode), 9
- cjk_compatibility_forms (Unicode), 9
- CJK_COMPATIBILITY_IDEOGRAPHS (Unicode), 9
- cjk_compatibility_ideographs (Unicode), 9

- CJK_COMPATIBILITY_IDEOGRAPHS_SUPPLEMENT (Unicode), 9
- cjk_compatibility_ideographs_supplement (Unicode), 9
- CJK_IDEOGRAPHIC_DESCRIPTION_CHARACTERS (Unicode), 9
- cjk_ideographic_description_characters (Unicode), 9
- CJK_STROKES (Unicode), 9
- cjk_strokes (Unicode), 9
- CJK_SYMBOLS_AND_PUNCTUATION (Unicode), 9
- cjk_symbols_and_punctuation (Unicode), 9
- CJK_UNIFIED_IDEOGRAPHS (Unicode), 9
- cjk_unified_ideographs (Unicode), 9
- CJK_UNIFIED_IDEOGRAPHS_EXTENSION_A (Unicode), 9
- cjk_unified_ideographs_extension_a (Unicode), 9
- CJK_UNIFIED_IDEOGRAPHS_EXTENSION_B (Unicode), 9
- cjk_unified_ideographs_extension_b (Unicode), 9
- CJK_UNIFIED_IDEOGRAPHS_EXTENSION_C (Unicode), 9
- cjk_unified_ideographs_extension_c (Unicode), 9
- CJK_UNIFIED_IDEOGRAPHS_EXTENSION_D (Unicode), 9
- cjk_unified_ideographs_extension_d (Unicode), 9
- ClassGroups, 4, 4
- CLOSE_BRACKET (SpecialCharacters), 9
- CLOSE_PAREN (SpecialCharacters), 9
- CNTRL (CharacterClasses), 3
- cntrl (ClassGroups), 4
- COMBINING_DIACRITIC_EXTENDED (Unicode), 9
- combining_diacritic_extended (Unicode), 9
- COMBINING_DIACRITIC_MARKS (Unicode), 9
- combining_diacritic_marks (Unicode), 9
- COMBINING_DIACRITIC_MARKS_FOR_SYMBOLS (Unicode), 9
- combining_diacritic_marks_for_symbols (Unicode), 9
- COMBINING_DIACRITIC_SUPPLEMENT (Unicode), 9
- combining_diacritic_supplement

- (Unicode), 9
- COMBINING_HALF_MARKS (Unicode), 9
- combining_half_marks (Unicode), 9
- COMMON_INDIC_NUMBER_FORMS (Unicode), 9
- common_indic_number_forms (Unicode), 9
- Concatenation, 4, 4
- CONTROL_PICTURES (Unicode), 9
- control_pictures (Unicode), 9
- COPTIC (Unicode), 9
- coptic (Unicode), 9
- COPTIC_EPACT_NUMBERS (Unicode), 9
- coptic_epact_numbers (Unicode), 9
- COUNTING_ROD_NUMERALS (Unicode), 9
- counting_rod_numerals (Unicode), 9
- CUNEIFORM (Unicode), 9
- cuneiform (Unicode), 9
- CUNEIFORM_NUMBERS_AND_PUNCTUATION (Unicode), 9
- cuneiform_numbers_and_punctuation (Unicode), 9
- CURRENCY_SYMBOLS (Unicode), 9
- currency_symbols (Unicode), 9
- CYPRIOT_SYLLABARY (Unicode), 9
- cyprriot_syllabary (Unicode), 9
- CYRILLIC (Unicode), 9
- cyrillic (Unicode), 9
- CYRILLIC_EXTENDED_A (Unicode), 9
- cyrillic_extended_a (Unicode), 9
- CYRILLIC_EXTENDED_B (Unicode), 9
- cyrillic_extended_b (Unicode), 9
- CYRILLIC_SUPPLEMENT (Unicode), 9
- cyrillic_supplement (Unicode), 9
- DateTime, 4, 4
- datetime (DateTime), 4
- DAY (DateTime), 4
- DAY_IN (DateTime), 4
- DAY_OF_YEAR (DateTime), 4
- DAY_OF_YEAR_IN (DateTime), 4
- DAY_SINGLE (DateTime), 4
- DESERET (Unicode), 9
- deseret (Unicode), 9
- DEVANAGARI (Unicode), 9
- devanagari (Unicode), 9
- DEVANAGARI_EXTENDED (Unicode), 9
- devanagari_extended (Unicode), 9
- DGT (CharacterClasses), 3
- dgt (ClassGroups), 4
- DIGIT (CharacterClasses), 3
- digit (ClassGroups), 4
- DINGBATS (Unicode), 9
- dingbats (Unicode), 9
- DMY (DateTime), 4
- DMY_IN (DateTime), 4
- DOLLAR (SpecialCharacters), 9
- DOMINO_TILES (Unicode), 9
- domino_tiles (Unicode), 9
- DOT (SpecialCharacters), 9
- DTSEP (DateTime), 4
- duplicate_group_names (modify_mode), 6
- DUPLOYAN (Unicode), 9
- duployan (Unicode), 9
- DYM (DateTime), 4
- DYM_IN (DateTime), 4
- EGYPTIAN_HIEROGLYPHS (Unicode), 9
- egyptian_hieroglyphs (Unicode), 9
- ELBASAN (Unicode), 9
- elbasan (Unicode), 9
- EMOTICONS (Unicode), 9
- emoticons (Unicode), 9
- ENCLOSED_ALPHANUMERIC_SUPPLEMENT (Unicode), 9
- enclosed_alphanumeric_supplement (Unicode), 9
- ENCLOSED_ALPHANUMERICS (Unicode), 9
- enclosed_alphanumerics (Unicode), 9
- ENCLOSED_CJK_LETTERS_AND_MONTHS (Unicode), 9
- enclosed_cjk_letters_and_months (Unicode), 9
- ENCLOSED_IDEOGRAPHIC_SUPPLEMENT (Unicode), 9
- enclosed_ideographic_supplement (Unicode), 9
- END (Anchors), 2
- engroup (capture), 3
- escape_special, 4, 4
- ETHIOPIC (Unicode), 9
- ethiopic (Unicode), 9
- ETHIOPIC_EXTENDED (Unicode), 9
- ethiopic_extended (Unicode), 9
- ETHIOPIC_EXTENDED_A (Unicode), 9
- ethiopic_extended_a (Unicode), 9
- ETHIOPIC_SUPPLEMENT (Unicode), 9
- ethiopic_supplement (Unicode), 9
- exactly, 4, 4

- FLOORS_AND_CEILINGS (Unicode), 9
- floors_and_ceilings (Unicode), 9
- format.regex, 5, 5
- FRACTIONAL_SECOND (DateTime), 4
- FRACTIONAL_SECOND_IN (DateTime), 4
- free_spacing (modify_mode), 6
- FULLWIDTH_ASCII_DIGITS (Unicode), 9
- fullwidth_ascii_digits (Unicode), 9
- FULLWIDTH_ASCII_PUNCTUATION (Unicode), 9
- fullwidth_ascii_punctuation (Unicode), 9

- GENERAL_PUNCTUATION (Unicode), 9
- general_punctuation (Unicode), 9
- GEOMETRIC_SHAPES (Unicode), 9
- geometric_shapes (Unicode), 9
- GEOMETRIC_SHAPES_EXTENDED (Unicode), 9
- geometric_shapes_extended (Unicode), 9
- GEORGIAN (Unicode), 9
- georgian (Unicode), 9
- GEORGIAN_SUPPLEMENT (Unicode), 9
- georgian_supplement (Unicode), 9
- get_months (get_weekdays), 5
- get_weekdays, 5, 5
- GLAGOLITIC (Unicode), 9
- glagolitic (Unicode), 9
- GOTHIC (Unicode), 9
- gothic (Unicode), 9
- GRANTHA (Unicode), 9
- grantha (Unicode), 9
- GRAPH (CharacterClasses), 3
- graph (ClassGroups), 4
- GRAPHEME (CharacterClasses), 3
- grapheme (ClassGroups), 4
- GREEK_AND_COPTIC (Unicode), 9
- greek_and_coptic (Unicode), 9
- GREEK_EXTENDED (Unicode), 9
- greek_extended (Unicode), 9
- group (capture), 3
- GUJARATI (Unicode), 9
- gujarati (Unicode), 9
- GURMUKHI (Unicode), 9
- gurmukhi (Unicode), 9

- HALFWIDTH_AND_FULLWIDTH_FORMS (Unicode), 9
- halfwidth_and_fullwidth_forms (Unicode), 9
- HANGUL_COMPATIBILITY_JAMO (Unicode), 9
- hangul_compatibility_jamo (Unicode), 9

- HANGUL_JAMO (Unicode), 9
- hangul_jamo (Unicode), 9
- HANGUL_JAMO_EXTENDED_A (Unicode), 9
- hangul_jamo_extended_a (Unicode), 9
- HANGUL_JAMO_EXTENDED_B (Unicode), 9
- hangul_jamo_extended_b (Unicode), 9
- HANGUL_SYLLABLES (Unicode), 9
- hangul_syllables (Unicode), 9
- HANUNOO (Unicode), 9
- hanunoo (Unicode), 9
- HEBREW (Unicode), 9
- hebrew (Unicode), 9
- HEX_DIGIT (CharacterClasses), 3
- hex_digit (ClassGroups), 4
- HIRAGANA (Unicode), 9
- hiragana (Unicode), 9
- HM (DateTime), 4
- HM_IN (DateTime), 4
- HMS (DateTime), 4
- HMS_IN (DateTime), 4
- HOUR12 (DateTime), 4
- HOUR12_IN (DateTime), 4
- HOUR12_SINGLE (DateTime), 4
- HOUR24 (DateTime), 4
- HOUR24_IN (DateTime), 4
- HOUR24_SINGLE (DateTime), 4

- ICU_REF1 (Backreferences), 3
- ICU_REF2 (Backreferences), 3
- ICU_REF3 (Backreferences), 3
- ICU_REF4 (Backreferences), 3
- ICU_REF5 (Backreferences), 3
- ICU_REF6 (Backreferences), 3
- ICU_REF7 (Backreferences), 3
- ICU_REF8 (Backreferences), 3
- ICU_REF9 (Backreferences), 3
- IMPERIAL_ARAMAIC (Unicode), 9
- imperial_aramaic (Unicode), 9
- INVISIBLE_OPERATORS (Unicode), 9
- invisible_operators (Unicode), 9
- IPA_EXTENSIONS (Unicode), 9
- ipa_extensions (Unicode), 9
- is.regex (as.regex), 3
- ISO_DATE (DateTime), 4
- iso_date (IsoClasses), 5
- ISO_DATE_IN (DateTime), 4
- ISO_DATETIME (DateTime), 4
- iso_datetime (IsoClasses), 5
- ISO_DATETIME_IN (DateTime), 4

- ISO_TIME (DateTime), 4
- iso_time (IsoClasses), 5
- ISO_TIME_IN (DateTime), 4
- IsoClasses, 5, 5
- IsoDateTime (IsoClasses), 5

- JAPANESE_CHESS (Unicode), 9
- japanese_chess (Unicode), 9
- JAVANESE (Unicode), 9
- javanese (Unicode), 9

- KAITHI (Unicode), 9
- kaithi (Unicode), 9
- KANA_SUPPLEMENT (Unicode), 9
- kana_supplement (Unicode), 9
- KANBUN (Unicode), 9
- kanbun (Unicode), 9
- KANGXI_RADICALS (Unicode), 9
- kangxi_radicals (Unicode), 9
- KANGXI_RADICALS_SUPPLEMENT (Unicode), 9
- kangxi_radicals_supplement (Unicode), 9
- KANNADA (Unicode), 9
- kannada (Unicode), 9
- KATAKANA (Unicode), 9
- katakana (Unicode), 9
- KATAKANA_PHONETIC_EXTENSIONS (Unicode), 9
- katakana_phonetic_extensions (Unicode), 9

- KAYAH_LI (Unicode), 9
- kayah_li (Unicode), 9
- KHAROSHTHI (Unicode), 9
- kharoshthi (Unicode), 9
- KHMER (Unicode), 9
- khmer (Unicode), 9
- KHMER_SYMBOLS (Unicode), 9
- khmer_symbols (Unicode), 9
- KHOJKI (Unicode), 9
- khojki (Unicode), 9
- KHUDAWADI (Unicode), 9
- khudawadi (Unicode), 9

- LAO (Unicode), 9
- lao (Unicode), 9
- LATIN (Unicode), 9
- latin (Unicode), 9
- LATIN_1_PUNCTUATION (Unicode), 9
- latin_1_punctuation (Unicode), 9
- LATIN_1_SUPPLEMENT (Unicode), 9
- latin_1_supplement (Unicode), 9
- LATIN_EXTENDED_A (Unicode), 9
- latin_extended_a (Unicode), 9
- LATIN_EXTENDED_ADDITIONAL (Unicode), 9
- latin_extended_additional (Unicode), 9
- LATIN_EXTENDED_B (Unicode), 9
- latin_extended_b (Unicode), 9
- LATIN_EXTENDED_C (Unicode), 9
- latin_extended_c (Unicode), 9
- LATIN_EXTENDED_D (Unicode), 9
- latin_extended_d (Unicode), 9
- LATIN_EXTENDED_E (Unicode), 9
- latin_extended_e (Unicode), 9
- LATIN_LIGATURES (Unicode), 9
- latin_ligatures (Unicode), 9
- lazy (repeated), 8
- LEPCHA (Unicode), 9
- lepcha (Unicode), 9
- LETTERLIKE_SYMBOLS (Unicode), 9
- letterlike_symbols (Unicode), 9
- LIMBU (Unicode), 9
- limbu (Unicode), 9
- LINEAR_A (Unicode), 9
- linear_a (Unicode), 9
- LINEAR_B_IDEOGRAMS (Unicode), 9
- linear_b_ideograms (Unicode), 9
- LINEAR_B_SYLLABARY (Unicode), 9
- linear_b_syllabary (Unicode), 9
- LISU (Unicode), 9
- lisu (Unicode), 9
- literal, 5, 5
- lookahead, 5, 5
- lookbehind (lookahead), 5
- LOWER (CharacterClasses), 3
- lower (ClassGroups), 4
- LYCIAN (Unicode), 9
- lycian (Unicode), 9
- LYDIAN (Unicode), 9
- lydian (Unicode), 9

- MAHAJANI (Unicode), 9
- mahajani (Unicode), 9
- MAHJONG_TILES (Unicode), 9
- mahjong_tiles (Unicode), 9
- MALAYALAM (Unicode), 9
- malayalam (Unicode), 9
- MANDAIC (Unicode), 9
- mandaic (Unicode), 9
- MANICHAEAN (Unicode), 9

- manichaeian (Unicode), 9
- MATH_ARROWS (Unicode), 9
- math_arrows (Unicode), 9
- MATHEMATICAL_ALPHANUMERIC_SYMBOLS (Unicode), 9
- mathematical_alphanumeric_symbols (Unicode), 9
- MDY (DateTime), 4
- MDY_IN (DateTime), 4
- MEETEI_MAYEK (Unicode), 9
- meetui_mayek (Unicode), 9
- MEETEI_MAYEK_EXTENSIONS (Unicode), 9
- meetui_mayek_extensions (Unicode), 9
- MENDE_KIKAKUI (Unicode), 9
- mende_kikakui (Unicode), 9
- MEROITIC_CURSIVE (Unicode), 9
- meroitic_cursive (Unicode), 9
- MEROITIC_HIEROGLYPHS (Unicode), 9
- meroitic_hieroglyphs (Unicode), 9
- MIAO (Unicode), 9
- miao (Unicode), 9
- MINUTE (DateTime), 4
- MINUTE_IN (DateTime), 4
- MISCELLANEOUS_MATHEMATICAL_SYMBOLS_A (Unicode), 9
- miscellaneous_mathematical_symbols_a (Unicode), 9
- MISCELLANEOUS_MATHEMATICAL_SYMBOLS_B (Unicode), 9
- miscellaneous_mathematical_symbols_b (Unicode), 9
- MISCELLANEOUS_SYMBOLS_AND_PICTOGRAPHS (Unicode), 9
- miscellaneous_symbols_and_pictographs (Unicode), 9
- MISCELLANEOUS_TECHNICAL (Unicode), 9
- miscellaneous_technical (Unicode), 9
- MODI (Unicode), 9
- modi (Unicode), 9
- MODIFIER_TONE_LETTERS (Unicode), 9
- modifier_tone_letters (Unicode), 9
- modify_mode, 6, 6
- MONGOLIAN (Unicode), 9
- mongolian (Unicode), 9
- MONTH (DateTime), 4
- MONTH_IN (DateTime), 4
- MRO (Unicode), 9
- mro (Unicode), 9
- MS (DateTime), 4
- MS_IN (DateTime), 4
- multi_line (modify_mode), 6
- MUSICAL_SYMBOLS (Unicode), 9
- musical_symbols (Unicode), 9
- MYANMAR (Unicode), 9
- myanmar (Unicode), 9
- MYANMAR_EXTENDED_A (Unicode), 9
- myanmar_extended_a (Unicode), 9
- MYANMAR_EXTENDED_B (Unicode), 9
- myanmar_extended_b (Unicode), 9
- MYD (DateTime), 4
- MYD_IN (DateTime), 4
- NABATAEAN (Unicode), 9
- nabataean (Unicode), 9
- negate_and_group (char_class), 3
- negated_char_class (char_class), 3
- negative_lookahead (lookahead), 5
- negative_lookbehind (lookahead), 5
- NEW_TAI_LUE (Unicode), 9
- new_tai_lue (Unicode), 9
- NEWLINE (CharacterClasses), 3
- newline (ClassGroups), 4
- NKO (Unicode), 9
- nko (Unicode), 9
- no_backslash_escaping (modify_mode), 6
- NOT_BOUNDARY (WordBoundaries), 10
- NOT_DGT (CharacterClasses), 3
- not_dgt (ClassGroups), 4
- NOT_SPC (CharacterClasses), 3
- not_spc (ClassGroups), 4
- NOT_WRD (CharacterClasses), 3
- not_wrd (ClassGroups), 4
- NUMBER_FORMS (Unicode), 9
- number_forms (Unicode), 9
- number_range, 6, 6
- OGHAM (Unicode), 9
- ogham (Unicode), 9
- OL_CHIKI (Unicode), 9
- ol_chiki (Unicode), 9
- OLD_ITALIC (Unicode), 9
- old_italic (Unicode), 9
- OLD_NORTH_ARABIAN (Unicode), 9
- old_north_arabian (Unicode), 9
- OLD_PERMIC (Unicode), 9
- old_permic (Unicode), 9
- OLD_PERSIAN (Unicode), 9

- old_persian (Unicode), 9
- OLD_SOUTH_ARABIAN (Unicode), 9
- old_south_arabian (Unicode), 9
- OLD_TURKIC (Unicode), 9
- old_turkic (Unicode), 9
- one_or_more (repeated), 8
- OPEN_BRACE (SpecialCharacters), 9
- OPEN_BRACKET (SpecialCharacters), 9
- OPEN_PAREN (SpecialCharacters), 9
- OPT_LEADING_0 (DateTime), 4
- OPTICAL_CHARACTER_RECOGNITION (Unicode), 9
- optical_character_recognition (Unicode), 9
- optional (repeated), 8
- or, 6, 6
- or1 (or), 6
- ORIYA (Unicode), 9
- oriya (Unicode), 9
- ORNAMENTAL_DINGBATS (Unicode), 9
- ornamental_dingbats (Unicode), 9
- OSMANYA (Unicode), 9
- osmanya (Unicode), 9

- PAHAHW_HMONG (Unicode), 9
- pahawh_hmong (Unicode), 9
- PAHLAVI_INSCRIPTIONAL (Unicode), 9
- pahlavi_inscriptional (Unicode), 9
- PAHLAVI_PSALTER (Unicode), 9
- pahlavi_psalter (Unicode), 9
- PALMYRENE (Unicode), 9
- palmyrene (Unicode), 9
- PAU_CIN_HAU (Unicode), 9
- pau_cin_hau (Unicode), 9
- PHAGS_PA (Unicode), 9
- phags_pa (Unicode), 9
- PHAISTOS_DISC (Unicode), 9
- phaistos_disc (Unicode), 9
- PHOENICIAN (Unicode), 9
- phoenician (Unicode), 9
- PHONETIC_EXTENSIONS (Unicode), 9
- phonetic_extensions (Unicode), 9
- PHONETIC_EXTENSIONS_SUPPLEMENT (Unicode), 9
- phonetic_extensions_supplement (Unicode), 9
- PIPE (SpecialCharacters), 9
- PLAYING_CARDS (Unicode), 9
- playing_cards (Unicode), 9

- PLUS (SpecialCharacters), 9
- PRINT (CharacterClasses), 3
- print.regex (format.regex), 5
- printable (ClassGroups), 4
- PRIVATE_USE_AREA (Unicode), 9
- private_use_area (Unicode), 9
- PUNCT (CharacterClasses), 3
- punct (ClassGroups), 4

- QUESTION (SpecialCharacters), 9

- rebus, 6
- rebus-package (rebus), 6
- recursive, 8, 8
- REF1 (Backreferences), 3
- REF2 (Backreferences), 3
- REF3 (Backreferences), 3
- REF4 (Backreferences), 3
- REF5 (Backreferences), 3
- REF6 (Backreferences), 3
- REF7 (Backreferences), 3
- REF8 (Backreferences), 3
- REF9 (Backreferences), 3
- regex, 7, 8, 8
- regexpr, 7
- REJANG (Unicode), 9
- rejang (Unicode), 9
- repeated, 8, 8
- ReplacementCase, 8, 8
- ROMAN (roman), 9
- roman, 9, 9
- RUMI_NUMERAL_SYMBOLS (Unicode), 9
- rumi_numeral_symbols (Unicode), 9
- RUNIC (Unicode), 9
- runic (Unicode), 9

- SAMARITAN (Unicode), 9
- samaritan (Unicode), 9
- SAURASHTRA (Unicode), 9
- saurashtra (Unicode), 9
- SECOND (DateTime), 4
- SECOND_IN (DateTime), 4
- SHARADA (Unicode), 9
- sharada (Unicode), 9
- SHAVIAN (Unicode), 9
- shavian (Unicode), 9
- SHORTHAND_FORMAT_CONTROLS (Unicode), 9
- shorthand_format_controls (Unicode), 9
- SIDDHAM (Unicode), 9

- siddham (Unicode), 9
- single_line (modify_mode), 6
- SINHALA (Unicode), 9
- sinhala (Unicode), 9
- SINHALA_ARCHAIC_NUMBERS (Unicode), 9
- sinhala_archaic_numbers (Unicode), 9
- SMALL_FORM_VARIANTS (Unicode), 9
- small_form_variants (Unicode), 9
- SORA_SOMPENG (Unicode), 9
- sora_sompeng (Unicode), 9
- SPACE (CharacterClasses), 3
- space (ClassGroups), 4
- SPACING_MODIFIER_LETTERS (Unicode), 9
- spacing_modifier_letters (Unicode), 9
- SPC (CharacterClasses), 3
- spc (ClassGroups), 4
- SpecialCharacters, 9, 9
- SPECIALS (Unicode), 9
- specials (Unicode), 9
- STAR (SpecialCharacters), 9
- START (Anchors), 2
- SUNDANESE (Unicode), 9
- sundanese (Unicode), 9
- SUNDANESE_SUPPLEMENT (Unicode), 9
- sundanese_supplement (Unicode), 9
- SUPERSCRIPTS_AND_SUBSCRIPTS (Unicode), 9
- superscripts_and_subscripts (Unicode), 9
- SUPPLEMENTAL_ARROWS_A (Unicode), 9
- supplemental_arrows_a (Unicode), 9
- SUPPLEMENTAL_MATHEMATICAL_OPERATORS (Unicode), 9
- supplemental_mathematical_operators (Unicode), 9
- SUPPLEMENTAL_PUNCTUATION (Unicode), 9
- supplemental_punctuation (Unicode), 9
- SUPPLEMENTARY_PRIVATE_USE_AREA_A (Unicode), 9
- supplementary_private_use_area_a (Unicode), 9
- SUPPLEMENTARY_PRIVATE_USE_AREA_B (Unicode), 9
- supplementary_private_use_area_b (Unicode), 9
- SYLOTI_NAGRI (Unicode), 9
- syloti_nagri (Unicode), 9
- SYRIAC (Unicode), 9
- syriac (Unicode), 9
- TAGALOG (Unicode), 9
- tagalog (Unicode), 9
- TAGBANWA (Unicode), 9
- tagbanwa (Unicode), 9
- TAGS (Unicode), 9
- tags (Unicode), 9
- TAI_LE (Unicode), 9
- tai_le (Unicode), 9
- TAI_THAM (Unicode), 9
- tai_tham (Unicode), 9
- TAI_VIET (Unicode), 9
- tai_viet (Unicode), 9
- TAI_XUAN_JING_SYMBOLS (Unicode), 9
- tai_xuan_jing_symbols (Unicode), 9
- TAKRI (Unicode), 9
- takri (Unicode), 9
- TAMIL (Unicode), 9
- tamil (Unicode), 9
- TELUGU (Unicode), 9
- telugu (Unicode), 9
- THAANA (Unicode), 9
- thaana (Unicode), 9
- THAI (Unicode), 9
- thai (Unicode), 9
- TIBETAN (Unicode), 9
- tibetan (Unicode), 9
- TIFINAGH (Unicode), 9
- tifinagh (Unicode), 9
- TIMEZONE (DateTime), 4
- TIMEZONE_OFFSET (DateTime), 4
- TIRHUTA (Unicode), 9
- tirhuta (Unicode), 9
- token (capture), 3
- TRANSPORT_AND_MAP_SYMBOLS (Unicode), 9
- transport_and_map_symbols (Unicode), 9
- UGARITIC (Unicode), 9
- ugaritic (Unicode), 9
- UGC_CASED_LETTER (UnicodeGeneralCategory), 9
- ugc_cased_letter (UnicodeGeneralCategory), 9
- UGC_CLOSE_PUNCTUATION (UnicodeGeneralCategory), 9
- ugc_close_punctuation (UnicodeGeneralCategory), 9
- UGC_CONNECTOR_PUNCTUATION (UnicodeGeneralCategory), 9
- ugc_connector_punctuation (UnicodeGeneralCategory), 9

UGC_CONTROL (UnicodeGeneralCategory), 9
 ugc_control (UnicodeGeneralCategory), 9
 UGC_CURRENCY_SYMBOL
 (UnicodeGeneralCategory), 9
 ugc_currency_symbol
 (UnicodeGeneralCategory), 9
 UGC_DASH_PUNCTUATION
 (UnicodeGeneralCategory), 9
 ugc_dash_punctuation
 (UnicodeGeneralCategory), 9
 UGC_DECIMAL_NUMBER
 (UnicodeGeneralCategory), 9
 ugc_decimal_number
 (UnicodeGeneralCategory), 9
 UGC_ENCLOSING_MARK
 (UnicodeGeneralCategory), 9
 ugc_enclosing_mark
 (UnicodeGeneralCategory), 9
 UGC_FINAL_PUNCTUATION
 (UnicodeGeneralCategory), 9
 ugc_final_punctuation
 (UnicodeGeneralCategory), 9
 UGC_FORMAT_CONTROL
 (UnicodeGeneralCategory), 9
 ugc_format_control
 (UnicodeGeneralCategory), 9
 UGC_INITIAL_PUNCTUATION
 (UnicodeGeneralCategory), 9
 ugc_initial_punctuation
 (UnicodeGeneralCategory), 9
 UGC_LETTER (UnicodeGeneralCategory), 9
 ugc_letter (UnicodeGeneralCategory), 9
 UGC_LETTER_NUMBER
 (UnicodeGeneralCategory), 9
 ugc_letter_number
 (UnicodeGeneralCategory), 9
 UGC_LINE_SEPARATOR
 (UnicodeGeneralCategory), 9
 ugc_line_separator
 (UnicodeGeneralCategory), 9
 UGC_LOWERCASE_LETTER
 (UnicodeGeneralCategory), 9
 ugc_lowercase_letter
 (UnicodeGeneralCategory), 9
 UGC_MARK (UnicodeGeneralCategory), 9
 ugc_mark (UnicodeGeneralCategory), 9
 UGC_MATH_SYMBOL
 (UnicodeGeneralCategory), 9
 ugc_math_symbol
 (UnicodeGeneralCategory), 9
 UGC_MODIFIER_LETTER
 (UnicodeGeneralCategory), 9
 ugc_modifier_letter
 (UnicodeGeneralCategory), 9
 UGC_MODIFIER_SYMBOL
 (UnicodeGeneralCategory), 9
 ugc_modifier_symbol
 (UnicodeGeneralCategory), 9
 UGC_NONSPACING_MARK
 (UnicodeGeneralCategory), 9
 ugc_nonspacing_mark
 (UnicodeGeneralCategory), 9
 UGC_NUMBER (UnicodeGeneralCategory), 9
 ugc_number (UnicodeGeneralCategory), 9
 UGC_OPEN_PUNCTUATION
 (UnicodeGeneralCategory), 9
 ugc_open_punctuation
 (UnicodeGeneralCategory), 9
 UGC_OTHER (UnicodeGeneralCategory), 9
 ugc_other (UnicodeGeneralCategory), 9
 UGC_OTHER_LETTER
 (UnicodeGeneralCategory), 9
 ugc_other_letter
 (UnicodeGeneralCategory), 9
 UGC_OTHER_NUMBER
 (UnicodeGeneralCategory), 9
 ugc_other_number
 (UnicodeGeneralCategory), 9
 UGC_OTHER_PUNCTUATION
 (UnicodeGeneralCategory), 9
 ugc_other_punctuation
 (UnicodeGeneralCategory), 9
 UGC_OTHER_SYMBOL
 (UnicodeGeneralCategory), 9
 ugc_other_symbol
 (UnicodeGeneralCategory), 9
 UGC_PARAGRAPH_SEPARATOR
 (UnicodeGeneralCategory), 9
 ugc_paragraph_separator
 (UnicodeGeneralCategory), 9
 UGC_PRIVATE_USE_CONTROL
 (UnicodeGeneralCategory), 9
 ugc_private_use_control
 (UnicodeGeneralCategory), 9
 UGC_PUNCTUATION
 (UnicodeGeneralCategory), 9

- ugc_punctuation (UnicodeGeneralCategory), 9
- UGC_SEPARATOR (UnicodeGeneralCategory), 9
- ugc_separator (UnicodeGeneralCategory), 9
- UGC_SPACE_SEPARATOR (UnicodeGeneralCategory), 9
- ugc_space_separator (UnicodeGeneralCategory), 9
- UGC_SPACING_MARK (UnicodeGeneralCategory), 9
- ugc_spacing_mark (UnicodeGeneralCategory), 9
- UGC_SURROGATE_CONTROL (UnicodeGeneralCategory), 9
- ugc_surrogate_control (UnicodeGeneralCategory), 9
- UGC_SYMBOL (UnicodeGeneralCategory), 9
- ugc_symbol (UnicodeGeneralCategory), 9
- UGC_TITLECASE_LETTER (UnicodeGeneralCategory), 9
- ugc_titlecase_letter (UnicodeGeneralCategory), 9
- UGC_UNASSIGNED_CONTROL (UnicodeGeneralCategory), 9
- ugc_unassigned_control (UnicodeGeneralCategory), 9
- UGC_UPPERCASE_LETTER (UnicodeGeneralCategory), 9
- ugc_uppercase_letter (UnicodeGeneralCategory), 9
- Unicode, 9, 9
- unicode_intersect (UnicodeOperators), 9
- unicode_inverse (UnicodeOperators), 9
- unicode_setdiff (UnicodeOperators), 9
- unicode_union (UnicodeOperators), 9
- UnicodeGeneralCategory, 9, 9
- UnicodeOperators, 9, 9
- UnicodeProperty, 10, 10
- UNIFIED_CANADIAN_ABORIGINAL_SYLLABICS (Unicode), 9
- unified_canadian_aboriginal_syllabics (Unicode), 9
- UNIFIED_CANADIAN_ABORIGINAL_SYLLABICS_EXTENDED (Unicode), 9
- unified_canadian_aboriginal_syllabics_extended (Unicode), 9
- UNMATCHABLE (CharacterClasses), 3
- UP_ALPHABETIC (UnicodeProperty), 10
- up_alphabetic (UnicodeProperty), 10
- UP_ASCII_HEX_DIGIT (UnicodeProperty), 10
- up_ascii_hex_digit (UnicodeProperty), 10
- UP_BIDI_CONTROL (UnicodeProperty), 10
- up_bidi_control (UnicodeProperty), 10
- UP_BIDI_MIRRORED (UnicodeProperty), 10
- up_bidi_mirrored (UnicodeProperty), 10
- UP_CASE_IGNOREABLE (UnicodeProperty), 10
- up_case_ignorable (UnicodeProperty), 10
- UP_CASE_SENSITIVE (UnicodeProperty), 10
- up_case_sensitive (UnicodeProperty), 10
- UP_CASED (UnicodeProperty), 10
- up_cased (UnicodeProperty), 10
- UP_CHANGES_WHEN_CASEFOLDED (UnicodeProperty), 10
- up_changes_when_casefolded (UnicodeProperty), 10
- UP_CHANGES_WHEN_CASEMAPPED (UnicodeProperty), 10
- up_changes_when_casemapped (UnicodeProperty), 10
- UP_CHANGES_WHEN_LOWERCASED (UnicodeProperty), 10
- up_changes_when_lowercased (UnicodeProperty), 10
- UP_CHANGES_WHEN_NFKC_CASEFOLDED (UnicodeProperty), 10
- up_changes_when_nfkc_casefolded (UnicodeProperty), 10
- UP_CHANGES_WHEN_TITLECASED (UnicodeProperty), 10
- up_changes_when_titlecased (UnicodeProperty), 10
- UP_CHANGES_WHEN_UPPERCASED (UnicodeProperty), 10
- up_changes_when_uppercased (UnicodeProperty), 10
- UP_DASH (UnicodeProperty), 10
- up_dash (UnicodeProperty), 10
- UP_DEFAULT_IGNOREABLE_CODE_POINT (UnicodeProperty), 10
- up_default_ignorable_code_point (UnicodeProperty), 10
- UP_DEPRECATED (UnicodeProperty), 10
- up_deprecated (UnicodeProperty), 10
- UP_DIACRITIC (UnicodeProperty), 10

- up_diacritic (UnicodeProperty), 10
- UP_EXTENDER (UnicodeProperty), 10
- up_extender (UnicodeProperty), 10
- UP_HEX_DIGIT (UnicodeProperty), 10
- up_hex_digit (UnicodeProperty), 10
- UP_HYPHEN (UnicodeProperty), 10
- up_hyphen (UnicodeProperty), 10
- UP_ID_CONTINUE (UnicodeProperty), 10
- up_id_continue (UnicodeProperty), 10
- UP_ID_START (UnicodeProperty), 10
- up_id_start (UnicodeProperty), 10
- UP_IDEOGRAPHIC (UnicodeProperty), 10
- up_ideographic (UnicodeProperty), 10
- UP_LOWERCASE (UnicodeProperty), 10
- up_lowercase (UnicodeProperty), 10
- UP_MATH (UnicodeProperty), 10
- up_math (UnicodeProperty), 10
- UP_NONCHARACTER_CODE_POINT
(UnicodeProperty), 10
- up_noncharacter_code_point
(UnicodeProperty), 10
- UP_POSIX_ALNUM (UnicodeProperty), 10
- up_posix_alnum (UnicodeProperty), 10
- UP_POSIX_BLANK (UnicodeProperty), 10
- up_posix_blank (UnicodeProperty), 10
- UP_POSIX_GRAPH (UnicodeProperty), 10
- up_posix_graph (UnicodeProperty), 10
- UP_POSIX_PRINT (UnicodeProperty), 10
- up_posix_print (UnicodeProperty), 10
- UP_POSIX_XDIGIT (UnicodeProperty), 10
- up_posix_xdigit (UnicodeProperty), 10
- UP_QUOTATION_MARK (UnicodeProperty), 10
- up_quotation_mark (UnicodeProperty), 10
- UP_SOFT_DOTTED (UnicodeProperty), 10
- up_soft_dotted (UnicodeProperty), 10
- UP_TERMINAL_PUNCTUATION
(UnicodeProperty), 10
- up_terminal_punctuation
(UnicodeProperty), 10
- UP_UPPERCASE (UnicodeProperty), 10
- up_uppercase (UnicodeProperty), 10
- UP_WHITE_SPACE (UnicodeProperty), 10
- up_white_space (UnicodeProperty), 10
- UPPER (CharacterClasses), 3
- upper (ClassGroups), 4

- VAI (Unicode), 9
- vai (Unicode), 9
- VARIATION_SELECTORS (Unicode), 9
- variation_selectors (Unicode), 9
- VARIATION_SELECTORS_SUPPLEMENT
(Unicode), 9
- variation_selectors_supplement
(Unicode), 9
- VEDIC_EXTENSIONS (Unicode), 9
- vedic_extensions (Unicode), 9
- VERTICAL_FORMS (Unicode), 9
- vertical_forms (Unicode), 9

- WARANG_CITI (Unicode), 9
- warang_citi (Unicode), 9
- WEEK_OF_YEAR (DateTime), 4
- WEEK_OF_YEAR_IN (DateTime), 4
- WEEKDAY0 (DateTime), 4
- WEEKDAY1 (DateTime), 4
- whole_word, 10, 10
- WordBoundaries, 10, 10
- WRD (CharacterClasses), 3
- wrd (ClassGroups), 4

- YDM (DateTime), 4
- YDM_IN (DateTime), 4
- YEAR (DateTime), 4
- YEAR2 (DateTime), 4
- YEAR4 (DateTime), 4
- YI_RADICALS (Unicode), 9
- yi_radicals (Unicode), 9
- YI_SYLLABLES (Unicode), 9
- yi_syllables (Unicode), 9
- YIJING_HEXAGRAM_SYMBOLS (Unicode), 9
- yijing_hexagram_symbols (Unicode), 9
- YIJING_MONO_DI_AND_TRIGRAMS (Unicode), 9
- yijing_mono_di_and_trigrams (Unicode), 9
- YMD (DateTime), 4
- YMD_IN (DateTime), 4

- zero_or_more (repeated), 8