

Regex.new('2.0.0')

Ruby 2.0.0 での正規表現の新機能

西山和広

日本Rubyの会

Onigmo

- Onigmo (Oniguruma-mod)
 - NEWS of Ruby 2.0.0 says following only:
 - Merge Onigmo
- <https://github.com/k-takata/Onigmo>

- Details are unknown

詳細不明

New feature (1) ¥K

examples without /¥K/

```
"foobar".sub(/(?<=foo)bar/, "") #=> "foo"  
"foobar".sub(/(?<=fo*)bar/, "")  
# SyntaxError: invalid pattern in look-behind: /(?<=fo*)bar/
```

examples with /¥K/

```
"foobar".sub(/foo¥Kbar/, "") #=> "foo"  
"foobar".sub(/fo*¥Kbar/, "") #=> "foo"
```

New feature (1) ¥K

Treat the first non-blank character of the line.

examples with /¥K/

```
gsub(/^ *¥K(¥d+)/) { $1.to_i+1 }
```

examples without /¥K/

```
gsub(/^ ( *) (¥d+)/) { "#{$1}#{$2.to_i+1}" }
```

New feature (2) ¥R

- Linebreak

改行文字

- Unicode:

```
(?>¥x0D¥x0A|[¥x0A-¥x0D¥x{85}¥x{2028}¥x{2029}])
```

- Not Unicode:

```
(?>¥x0D¥x0A|[¥x0A-¥x0D])
```

New feature (3) ¥X

- eXtended grapheme cluster

拡張書記素クラス

- Unicode:

(?>¥P{M}¥p{M}*)

- Not Unicode:

(?m..)

Extended grapheme cluster

● example:

```
"¥u{304B 3099}"[/¥X/].size #=> 2
```

● U+304B HIRAGANA LETTER KA

● U+3099 COMBINING KATAKANA-HIRAGANA
VOICED SOUND MARK

● see [UAX #29] for more detail

(Unicode標準附属書29)

New feature (4)

- conditional expression:

```
(?(cond)yes)
(?(cond)yes|no)
```

- example:

```
" :f o o "[/:([' "])?(?(1)[\w\s]+¥1 |¥w+)/]
#=> ":f"
"' :' f o o' "[/:([' "])?(?(1)[\w\s]+¥1 |¥w+)/]
#=> ":' f o o'"
```


(?adu)

- character set option (character range option)

文字集合オプション (文字範囲オプション)

- d: Default (compatible with Ruby 1.9.3)

- a: ASCII

- u: Unicode

see doc/RE in Onigmo for more detail

(?adu)

examples:

```
"¥u{3042}"[/¥w/]      #=> nil  
"¥u{3042}"[/(?a)¥w/]  #=> nil  
"¥u{3042}"[/(?d)¥w/]  #=> nil  
"¥u{3042}"[/(?u)¥w/]  #=> "あ"
```

```
/a¥b/      =~ "a¥u{3042}" #=> nil  
/(?a)a¥b/  =~ "a¥u{3042}" #=> 0  
/(?d)a¥b/  =~ "a¥u{3042}" #=> nil  
/(?u)a¥b/  =~ "a¥u{3042}" #=> nil
```

(?adu)

- (?-a), (?-d), (?-u) do not found
- unlike (?-i), (?-m), (?-x)

Character Property

- support for Unicode blocks
- example:

```
/¥p{InHiragana}/ =~ "¥u3042" #=> 0  
/¥p{InCJKUnifiedIdeographs}/ =~ "¥u3042" #=> nil
```

- see tool/enc-unicode.rb in Onigmo for more detail

/ ¥ Z /