Регулярные выражения

Королев Н.С.

16 ноября 2018 г.

Регулярные выражения

Возможности регулярных выражений в разных языках отличаются

1. **PCRE (perl, php)** - Perl Compatible Regular Expressions
2. **Python**
3. **javascript**
4. **Golang**

Протестировать регулярные выражения можно на regex101.com (там есть объяснения)

Примитивный поиск

aba	Точное совпадение	abababa abacus zabava
[xyzA-D]	Одно из множества	xhelloy camel Case
[^xyzA-D]	Всё кроме множества (один символ)	xhelloy camel Case
. (точка)	Любой символ (кроме \n)	anything I'm serious

Полезные сокращения

\d	[0-9]	xxxNagibator228xxx -1.2e5 128746123,33
\D	[^0-9]	xxxNagibator288xxx -1.2e5 128746123,33
\w	[A-Za-z\d_]	I'll be back $999 _class_attribute
\s	[\r\n\t\f\v]	Any whitespace character

Модификаторы

{n,m}	Предыдущий символ от n до m раз (жадно)
*	Предыдущий символ любое количество раз (эквивалент {0,}) (жадно)
+	Предыдущий символ ненулевое количество раз (эквивалент {1,}) (жадно)
?	Предыдущий символ ноль или один раз (эквивалент {0, 1}) (жадно)

Примеры с модификаторами

.*	Весь текст (без \n) (в том числе пустые строки)	I'll be back
\.jpe?g	.jpg и .jpeg	image.jpg 1986.png cat.jpeg fail.jpeg.zip
\[.+\]	Ссылки с википедии (с багом)	word [17] done [18][19] [20] failed [?]

Жадность и ленивость

{n,m}?	Предыдущий символ от n до m раз (лениво)
***?**	Предыдущий символ любое количество раз (эквивалент {0,}?) (лениво)
\[.*?\]	Ссылки с википедии

word [17]
done [18][19]
[20] failed [?]

Обратные ссылки (Backreferences)

(expr)	Группирование выражений
\1, \2, ..., \9	Ссылки на сгруппированные ранее выражения
(\w+) \1	Суффикс слова совпадает с префиксом следующего слова show off off and This this this This isn't the part
func\((.+), \1\)	Функция с двумя одинаковыми аргументами func(a, a) func(a, b) func(x.val, x.val) func(g(a, b), g(a, b)) func(g(a, g(a)), b)

Примеры для самостоятельного изучения

R(e)g\1xp?

[+-]?\d+[.,]?\d*

\w+@\w+\.[A-Za-z]{2,}

\d{1,3}(\.\d{1,3}){3}

\[[^\]]*\]

Якоря и ИЛИ

^	Начало строки
$	Конец строки
\b	На границе слова
expr1|expr2	expr1 или expr2
\bword\b	Слово "word"целиком
^word|word$	"word"в начале строки или в конце строки

word
bigword
wordplay
re-word

word by word
wording word words

Look-ahead, Look-behind

(?=regex)	Положительный просмотр вперёд (далее идёт regex)
(?!regex)	Отрицательный просмотр вперёд (далее идёт не regex)
(?<=regex)	Положительный просмотр назад
(?<!regex)	Отрицательный просмотр назад

.(?=0)	Символы после которых идёт 0	2018-2019 01000
(?<=10)20	Все 20, перед которыми есть 10	3151020 10x202020