

Регулярные выражения

Регулярное выражение — это шаблон задающий ограничение на вводимые пользователем значения. Оно может состоять как из обычных, так и из специальных символов и их комбинаций.

Обычные символы интерпретируются буквально, например регулярное выражение **абв** интерпретируется как строгая последовательность соответствующих букв в этом порядке.

Помимо этого, для составления более сложных регулярных выражений используются специальные символы. Их написание и значение подробнее расписаны ниже.

Специальные символы

Символ	Описание	Пример
Общие символы и символы количества		
\	Для обычных символов , означает что следующий символ будет интерпретироваться как специальный. Для специальных символов , означает что следующий символ будет интерпретироваться буквально.	\a — символ a становится специальным символом, означающим границу слова. * — специальный символ * интерпретируется как обычный символ "звездочка".
^	Следующий символ должен присутствовать в начале строки.	^A — символ A должен быть первым в строке.
\$	Предыдущий символ должен присутствовать в конце строки.	a\$ — символ a должен быть последним в строке.
*	Предыдущий символ может быть повторен 0 или больше раз.	a* — символ a может не быть в строке, а может повториться множество раз.
+	Предыдущий символ должен быть повторен 1 или больше раз.	a+ — символ a может повториться один раз в строке, а может повториться множество раз.
?	Предыдущий символ должен быть повторен 0 или 1 раз.	a? — символ a может не быть в строке, а может быть один раз.

.	Любой символ кроме переноса строки.	a.c равно abc , a1c и тд.
x y	В строке должен присутствовать либо x либо y.	abc xyz — в строке может быть abc или xyz .
{n}	Предыдущий символ должен быть повторен n раз.	a{2} — a должна быть повторена в строке дважды.
{n,m}	Предыдущий символ должен быть повторен не меньше n раз, но и не больше m раз.	a{2,5} — a должна быть повторена в строке хотя бы дважды и не больше пяти раз.
Скобочные группы		
(x)	Объединение символов в единый элемент (с запоминанием).	abc* — * относится только к c .
(?:x)	Объединение символов в единый элемент (без запоминания).	(abc)* — * относится ко всему выражению в скобках.
x(?:=y)	Соответствует x только если за ним идет y.	ab(?:=c) — после ab должна быть c .
x(?:!y)	Соответствует x только если за ним не идет y.	ab(?:!c) — после ab не должно быть c .
Соответствия набору символов		
[xyz]	Соответствует любому из символов в скобках.	[abcd] или [a-d] — строка должна состоять из любых символов от a до d .
[^xyz]	Соответствует любому из символов не в скобках.	[^abcd] или [^a-d] — строка должна исключать символы от a до d .
[\b]	Бэкспейс.	[\b] — находит символ бэкспейса в строке.
\b	Граница слова.	\b — перед или после предыдущего символа должен быть пробел или другая граница слова.
\d	Цифровой символ.	\d — любой из символов от 0 до 9.
\D	Нецифровой символ.	\D — любой из символов кроме символов от 0 до 9.
\s	Пробел.	\s — находит пробел, табуляцию, перенос строки.

<code>\S</code>	Не пробел.	<code>\S</code> — любой из символов кроме пробела, табуляции, переноса строки.
<code>\w</code>	Цифробуквенные символы и нижнее подчеркивание.	<code>\w</code> — любой из символов от 0 до 9, от А до Я и <code>_</code> .
<code>\W</code>	Нецифробуквенные символы.	<code>\W</code> — любой из символов кроме символов от 0 до 9, от А до Я и <code>_</code> .
<code>\0</code>	NULL	<code>\0</code> — находит пустое множество.

Примеры регулярных выражений и их разбор

Дата формата ДД.ММ.ГГГГ

Выражение: `(0?[1-9]|[12][0-9]|3[01])\.(0?[1-9]|1[012])\.(19|20)\d\d`

<code>(0?[1-9] [12][0-9] 3[01])</code>	
<code>0?[1-9]</code>	0 может быть повторен 1 или 0 раз, а за ним идет любая цифра кроме 0.
<code>[12][0-9]</code>	1 или 2, а за ними любая цифра.
<code>3[01]</code>	3 и 0 или 1.
<code> </code>	или.
<code>(0?[1-9] 1[012])</code>	
<code>1[012]</code>	1 и 0 или 1 или 2.
<code>\.(19 20)\d</code>	
<code>\.</code>	обычный символ точки.
<code>19 20</code>	19 или 20.
<code>\d</code>	любая цифра.

Возможная строка: 11.03.2005

Email

Выражение: [^@\\s]+@[^@\\s]+\\.([^@\\s]+)

[^@\\s]+	
[^@\\s]	любой из символов кроме @ и пробела.
+	предыдущий символ может быть повторен 1 или более раз.
@[^@\\s]+	
@	символ @.
\\.[^@\\s]+	
\\.	обычный символ точки.

Возможная строка: email@pp.com

Другие регулярные выражения

Email без латинских букв

```
(?:[a-z0-9!#$%&'*+|=?@^_`{|}~]+(?:\.[a-z0-9!#$%&'*+|=?@^_`{|}~]+)*)*(?:\[x01-\x08\x0b\x0c\x0e-\x1f\x21\x23-\x2f\x3d\x5d-\x7f]\|\\\[x01-\x09\x0b\x0c\x0e-\x7f])*)@(?:\.(?:[a-z0-9](?:[a-z0-9-]*[a-z0-9])?\.)+|[a-z0-9](?:[a-z0-9-]*[a-z0-9])?)|(?:(?:2(5[0-5])?[0-4][0-9])|1[0-9][0-9]|1[9]?[0-9])\.)\{3(?:2(5[0-5])?[0-4][0-9])|1[0-9][0-9]|1[9]?[0-9])|[a-z0-9-]*[a-z0-9](?:\[x01-\x08\x0b\x0c\x0e-\x1f\x21-\x2f\x3d\x5a\x53-\x7f]\|\\\[x01-\x09\x0b\x0c\x0e-\x7f])+\)\}
```

Для валидации номера телефона без рамок кодов страны

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

Возможная строка: +7 900 700 30 10 / 79007003010 / 8900 700 30 10 / + 37 900 700 30 10

Валидация номера с кодом определенных стран и возможностью выйти

$$(\wedge(\backslash+?7|\backslash+?37)\backslash d\{10\}$|отмена|назад|выход|стоп)$$

Возможная строка: +79007003010 / +379007003010 / +77007003010

Имя или слово с заглавной

[А-Я]{1}[а-я]{2,20}

Возможная строка: Метабот

Версия #3
Юрий Гарашко создал 12 June 2023 10:59:04
Ирина Петрова обновил 22 October 2024 09:20:59