
Урок 6

Регулярные выражения
Основные правила

Правило поиска для точки

. — любой символ

\. — точка



Диапазон допустимых значений

[нл] — только «н» и «л»

[а-я] — все русские буквы в нижнем регистре от «а» до «я» (кроме «ё»)

[А-Я] — все заглавные русские буквы

[А-Яа-яЁё] — все русские буквы

[a-z] — латиница мелким шрифтом

[a-zA-Z] — все английские буквы

[0-9] — любая цифра

[В-Ю] — буквы от «В» до «Ю» (да, диапазон — это не только от А до Я)

[А-ГО-Р] — буквы от «А» до «Г» и от «О» до «Р»

Единственный допустимый разделитель — это дефис

Например: [1-3] - верно, [1-31] - неверно



Исключение и дополнение допустимых значений

^ внутри [] означает исключение:

[^0-9] — любой символ, кроме цифр

[^ёЁ] — любой символ, кроме буквы «ё»

[^a-b8] — любой символ, кроме букв «a», «b», «v» и цифры 8

[a-я5,] — любой символ от a до я цифра 5 или запятая

Экранирование спец символов

Вот полный перечень этих символов:

`[]\^$.|?*+()`

Чтобы использовать специальный символ как обычный, добавьте к нему обратную косую черту: `\`.

Это называется «экранирование символа»

Перечисление вариантов

Чтобы указать допустимые значения для:

- Одного символа — используем `[]`
- Нескольких символов или целого слова — используем `|`



Круглые скобки обозначают группу символов.

Зачем нужны скобки?

Показать, где начинается и заканчивается группа.

Иначе вертикальная черта применится ко всем символам

A(н | л)я или **A(нн | л)я**

Метасимволы

Символ	Эквивалент	Пояснение
\d	[0-9]	Цифровой символ
\D	[^0-9]	Нецифровой символ
\s	[\f\n\r\t\v]	Пробельный символ
\S	[^ \f\n\r\t\v]	Непробельный символ
\w	[[:word:]]	Буквенный или цифровой символ или знак подчёркивания
\W	[^[:word:]]	Любой символ, кроме буквенного или цифрового символа или знака подчёркивания
.		Вообще любой символ

Квантификаторы

Квантификатор	Число повторений
?	Ноль или одно
*	Ноль или более
+	Один или более
{n}	Ровно n раз
{m,n}	От m до n включительно
{m,}	Не менее m

Regex: data{2}
Найдет: dataa
Не найдет: datadata

Regex: (data){2}
Найдет: datadata
Не найдет: dataa

Не забывайте — квантификатор применяется к последнему символу или группе символов!



Метасимволы, обозначающие позицию строки

Символ	Значение
\b	граница слова
\B	Не граница слова
^	начало текста (строки)
\$	конец текста (строки)

Какой у даты шаблон? Мы рассмотрим ДД.ММ.ГГГГ:

2 цифры дня

точка

2 цифры месяца

точка

4 цифры года

День месяца может быть максимум 31 — первая цифра [0-3]

Максимальный месяц 12 — первая цифра [01]

Год или 19.., или 20.. — первая цифра [12], а вторая [09]