

RegEx квантификаторы и символные классы

Движок регулярных выражений:

- 1. Пробует найти совпадение с шаблоном на позиции строки.
- 2. Если нет совпадения, переходит к следующей позиции.

18 типов RegEx квантификаторов:

Тип квантификатора	Жадный	Ленивый	Ревнивый
Ноль или один раз	?	??	?+
Ноль или больше раз	*	*?	*+
Один или больше раз	+	++	++
m раз	{m}	{m}?	{m}+
Не менее m раз	{m, }	{m, }?	{m, }+
Не менее m, не более n раз	{m, n}	{m, n}?	{m, n}+

Жадный, Нетерпеливый и Щедрый (Greedy, Eager and Generous) - в жадном режиме (по умолчанию) квантификатор повторяется столько раз, сколько это возможно, то есть, выражению соответствует максимально длинная строка из возможных.

Ленивый, Неохотный (Lazy, Reluctant) - в ленивом режиме квантификатор повторяется наименьшее количество раз, то есть, выражению соответствует самая короткая строка из возможных.

Ревнивый, Притяжательный, Сверх-жадный (Possessive) - в ревнивом режиме квантификация не только старается найти максимально длинный вариант, но ещё и не позволяет алгоритму возвращаться к предыдущим шагам поиска для того, чтобы найти возможные соответствия для оставшейся части регулярного выражения.

Символьные классы:

Символ	Эквивалент	Значение
.	[^\n\r\u2028\u2029]	Любой символ, кроме новой строки.
\d	[0-9]	Символ цифры.
\D	[^0-9]	Любой символ, кроме цифры.
\w	[A-Za-z0-9_]	Любая буква латинского алфавита, любая цифра, подчеркивание.
\W	[^A-Za-z0-9_]	Символ не буквы латинского алфавита, не цифры и не подчеркивания.
\s	[\f\n\r\t\v\u00a0 \u1680\u180e \u2001\u2003\u2004\u2006 \u2007\u2009\u2009\u2028 \u2029\u202f\u205f\u2060]	Одиночный пробельный символ: пробел, табуляция и другие пробельные символы юникода.

<code>\s</code>	<code>[^\f\n\r\t\v\u00a0 \u1680\u180e\u2001\u2003\u2004\u2006\u2007\u2009\u2009\u2028\u2029\u202f\u205f\u3000]</code>	Любой не-пробельный одиночный символ юникода.
<code>\n</code>		Символ перевода строки.
<code>\</code>		Обратный слэш указывает, что следующий символ является специальным и не должен интерпретироваться буквально.

Границы слова:

Символ	Пример	Значение
<code>^</code>	/^Б/ не сопоставляется с «Б» в «буква Б», но сопоставляется с первой «Б» в «Буква Б»	Сопоставляется с началом ввода.
<code>\$</code>	/д\$/ не сопоставляется с «д» в «проходи», но сопоставляется с «д» в «проход»	Сопоставляется с концом ввода.
<code>\b</code>	/\bпол/ сопоставляется с «пол» в «в полдень»; /но\b/ сопоставляется с «но» в «возможно завтра»	Сопоставляется с границей слова нулевой ширины, например с позицией между буквой и пробелом (не путайте с набором символов <code>[\b]</code>).
<code>\B</code>	/\Bдень/ сопоставляется с «день» в «в полдень»; /за\b/ сопоставляется с «за» в «возможно завтра»	Сопоставляется с границей не-слов нулевой ширины, например с позицией между двумя буквами или двумя пробелами.

Во время изучения
чего-то нового,
я самозабвенно
выдумываю
невероятные
ситуации, в которых
это умение поможет
мне спасти мир

О нет! Убийца должно
быть последовал
за ней в отпуск!



Но чтобы узнать где он, нам нужно
прочесать 200 Мб писем в поисках
чего-то схожего по формату с адресом!



Это безнадежно!

Всем расступиться



Я знаю регулярные
выражения



кляц
кляц



[Инструкция, как спасти день =\).](#)