

# Регулярные выражения

## Метасимволы для поиска символьных классов

Метасимвол	Назначение
\d	цифровой символ
\D	нецифровой символ
\s	символ пробела
\S	непробельный символ
\w	буквенно-цифровой символ или знак подчёркивания
\W	любой символ, кроме буквенного, цифрового или знака подчёркивания
.	любой символ

## Метасимволы для группировки символов

Метасимвол	Назначение
[абв]	любой из перечисленных (а, б, или в)
[^абв]	любой, кроме перечисленных (не а, б, в)
[a-zA-Z]	слияние диапазонов (латинские символы от а до z без учета регистра)
[a-d[m-p]]	объединение символов (от а до d и от m до p)

[a-z&&[def]]	пересечение символов (символы d, e, f)
[a-z&&[^bc]]	вычитание символов (символы a, d-z)

**Метасимволы для обозначения количества символов – квантификаторы. Квантификатор всегда следует после символа или группы символов.**

Метасимвол	Назначение
?	один или отсутствует
*	ноль или более раз
+	один или более раз
{n}	n раз
{n, }	n раз и более
{n,m}	не менее n раз и не более m раз

**Метасимволы для поиска символов редактирования текста**

Метасимвол	Назначение
\t	символ табуляции
\n	символ новой строки
\r	символ возврата каретки
\f	переход на новую страницу
\u 0085	символ следующей строки

<code>\u 2028</code>	символ разделения строк
<code>\u 2029</code>	символ разделения абзацев

## Метасимволы для поиска совпадений границ строк или текста

Метасимвол	Назначение
<code>^</code>	начало строки
<code>\$</code>	конец строки
<code>\b</code>	граница слова
<code>\B</code>	не граница слова
<code>\A</code>	начало ввода
<code>\G</code>	конец предыдущего совпадения
<code>\Z</code>	конец ввода
<code>\z</code>	конец ввода

Граница слова `\b` - это точка между двумя символами, у которой `\w` с одной стороны от нее и `\W` с другой стороны (в любом порядке).

# Задания

## 1. Квантификаторы – \* + ? и {}

1.	<b>abc*</b>	соответствует строке, в которой после ab следует 0 или более символов c
2.	<b>abc+</b>	соответствует строке, в которой после ab следует один или более символов c
3.	<b>abc?</b>	соответствует строке, в которой после ab следует 0 или один символ c
4.	<b>abc{2}</b>	соответствует строке, в которой после ab следует 2 символа c
5.	<b>abc{2,}</b>	соответствует строке, в которой после ab следует 2 или более символов c
6.	<b>abc{2,5}</b>	соответствует строке, в которой после ab следует от 2 до 5 символов c
7.	<b>a(bc)*</b>	соответствует строке, в которой после a следует 0 или более последовательностей символов bc
8.	<b>a(bc){2,5}</b>	соответствует строке, в которой после a следует от 2 до 5 последовательностей символов bc

Для наборов придуманных наборов строк, напишите проверку совпадения с шаблонами из представленной таблицы.

```
String testString="abcccccc";  
Pattern pattern = Pattern.compile("abc*");  
Matcher matcher = pattern.matcher(testString);  
System.out.println(matcher.matches());
```

## 2. Якоря – ^ и \$

<b>^Привет</b>	соответствует строке, начинающейся с Привет
<b>пока\$</b>	соответствует строке, заканчивающейся на пока
<b>^Привет пока\$</b>	точное совпадение (начинается и заканчивается как Привет пока)
<b>воробушки</b>	соответствует любой строке, в которой есть текст воробушки

- A. Напишите проверку для строк, которые начинаются с трех букв из диапазона [A-Z].
- B. Напишите проверку для строк, которые заканчиваются на 5 чисел из набора [1-9].
- C. Напишите проверку для строк, которые начинаются с пяти букв из диапазона [а-я] и заканчиваются на 4 числа из набора [1-9].
- D. Напишите проверку для строк, в которых есть слово пустыня.

### 3. Оператор ИЛИ – | или []

a(b c)	соответствует строке, в которой после a следует b или c
a[bc]	соответствует строке, в которой после a следует b или c

Выполните проверку для шаблонов, которые содержат оператор или.

### 4. Экранирование символов

Перед некоторыми символами необходимо ставить обратный слеш.

^.[\$( ) *+?{\ \	Эти символы необходимо экранировать
------------------	-------------------------------------

Выполните проверку для шаблонов, которые содержат экранированные символы.

### 5. Скобочные выражения []

[abc]	соответствует строке, которая содержит либо символ a или a b или a c -> такой же эффект от a b c
[a-c]	то же, что и выше
[a-fA-F0-9]	строка, представляющая одну шестнадцатеричную цифру без учёта регистра
[0-9]%	строка, содержащая символ от 0 до 9 перед знаком %
[^a-zA-Z]	строка, которая не имеет буквы от a до z или от A до Z. В этом случае ^ используется как отрицание в выражении

Внутри скобочных выражений все специальные символы (включая обратную косую черту \) теряют своё служебное значение, поэтому их ненужно экранировать.

Выполните проверку для шаблонов из таблицы.

## 6. Поиск тегов

Напишите несколько примеров строк, для которых проверка шаблона `<[^<>]+>` возвратит **a)** истину **b)** ложь

7. Напишите по три примера строк, для которых регулярные выражения из приведенной ниже таблицы, возвращают истину и ложь соответственно.

Выполните проверку.

```
String text = "AAABBBCCC";  
Pattern pattern = Pattern.compile("A|B");  
Matcher matcher = pattern.matcher(text);  
System.out.println(matcher.find());
```

#	Выражение
1.	<code>aa bb cc AA BB CC</code>
2.	<code>a(a b)b</code>
3.	<code>[a-cA-C]{2}</code>
4.	<code>\b\w+\b</code>
5.	<code>[a-d[m-p]]{3}</code>

8. Напишите регулярные выражения для форматов из приведенной ниже таблицы. Выполните проверку.

#	Формат
1.	<code>nn-nnnnn</code> (n - цифра от 1 до 9)
2.	<code>#XXXXXX</code> (X = 0 - 9, или A-F, или a-f )

3.	И.О.Фамилия
4.	nnn.nnn.nnn.nnn (n – цифра от 1 до 9)
5.	D:\\folder\\filename.ext