



АЮСА

Software Testing Company

Автоматизация тестирования

Регулярные выражения

План лекции

- Понятие и применение
- Элементы языка регулярных выражений
 - Простейшее регулярное выражение
 - Метасимволы и их значение
 - . (Точка)
 - Символьные классы
 - Квантификаторы
 - | (Выбор)
 - Группы
- Литература

Понятие и применение

Проверка корректности пользовательского ввода

facebook [Log in to existing account](#)

Sign Up

It's free and always will be.

galina makhotkina

galinamakhotkina@gmailcom

galinamakhotkina@gmailcom

.....

Birthday

Month Day Year Why do I need to provide my birthday?

Female Male

By clicking Sign Up, you agree to our [Terms](#) and that you have read our [Data Use Policy](#), including our [Cookie Use](#).

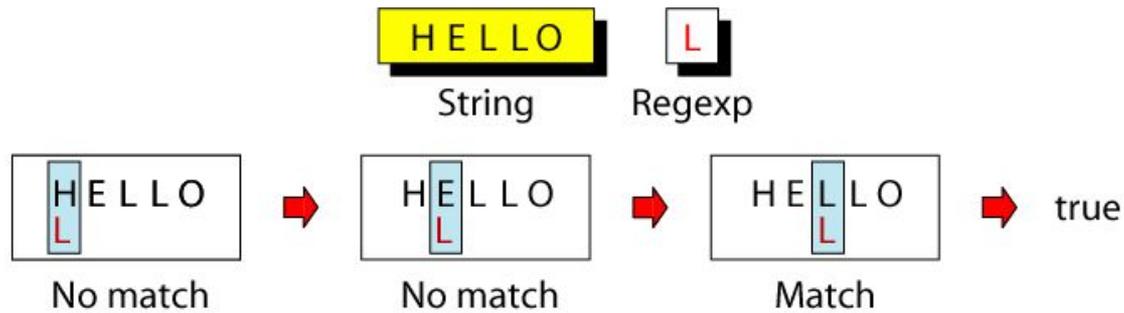
Sign Up

Please enter a valid email address.

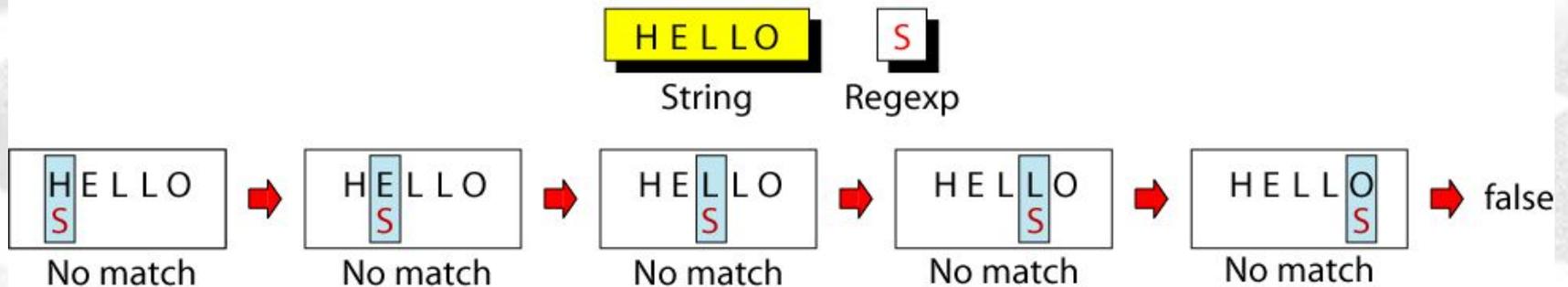
Mobile Find Friends Badges People Pages Places Apps Games Music
About Create Ad Create Page Developers Careers Privacy Cookies Terms Help

Facebook © 2013 · English (US)

Простейшее регулярное выражение



(a) Successful Pattern Match



(b) Unsuccessful Pattern Match

Метасимволы и их значение

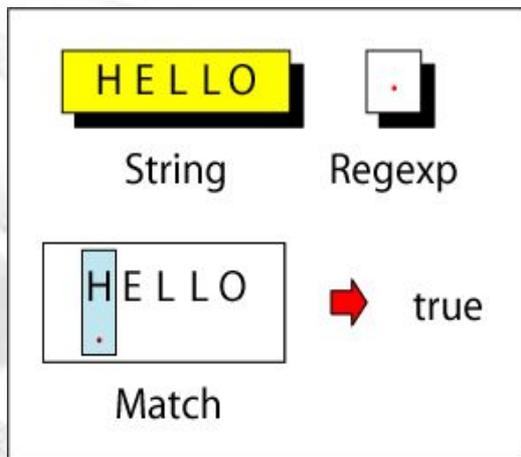
\wedge \$ \ . [] ? * + { } | ()



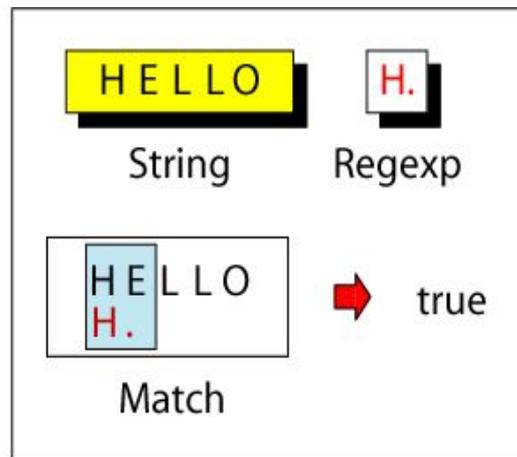
\. \^ \\$ \+ * и т. д.

. (Точка)

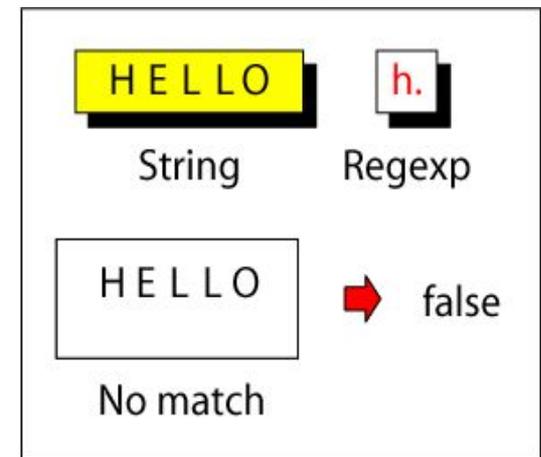
Определяет любой символ, кроме "\n" (новая строка).



(a) Single-Character



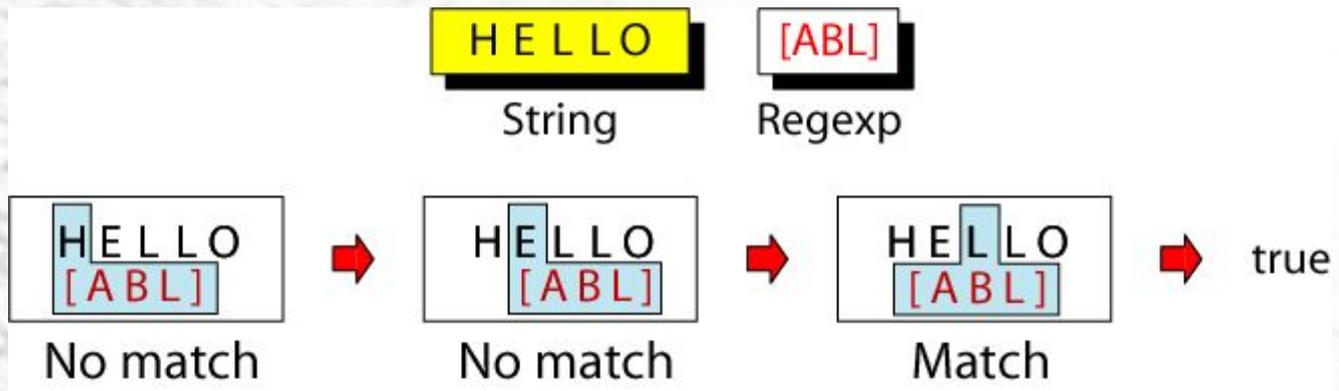
(b) Combination-True



(c) Combination-False

Символьные классы

- `[abcdef]` - Набор символов. Находит любой из указанных СИМВОЛОВ.



- `[^abcdef]` Исключающий набор символов. Находит любой из неуказанных символов.
- `[a-f]` Диапазон символов. Находит любой символ из диапазона.
- `[^a-f]` - Исключающий диапазон символов. Находит любой символ, не входящий в диапазон.

Символьные классы (сокращения)

.	любой символ
\d	любая цифра, т.е. [0-9]
\D	любой символ кроме цифры, т.е. [^\d]
\w	любой символ, образующий «слово», т.е. [A-Za-z0-9_]
\W	любой символ, НЕ образующий «слово», т.е. [^\w]
\s	любой whitespace символ, т.е. [.\t\v\r\n\f]
\S	любой НЕ whitespace символ, т.е. [^\s]

Можно использовать внутри символьного класса:
[\d\s] – цифры и whitespace символы

Квантификатор

Квантификатор	Описание	Пример
?	0 или 1 элемент {0,1}	colou?r
		color, colour colour
*	0 или более элементов {0,}	colou*r
		color, colour, colouur colour
+	1 или более элементов {1,}	colou+r
		colour, colouur color
{n}	ровно n элементов	co{2}l
		cool cl, col, coool
{m,n}	от m от n элементов	co{1,2}l
		col, cool cl, coool

| (Выбор)

gray | grey

gray, grey

other

gr (a | e) y

gray, grey

gry, graey

red | green | blue

red, green, blue

other

sw (i | a | u) m

swim, swam , swum

swiim, swiam...

Группы

Выделяет часть выражения в группу. Чтобы определить символы "(" и ")", следует использовать "\"(" и "\")".

A(BC){3}



ABCBCBC

(F(BC){2}G){2}



FBCBCGFBCBCG

Если в регулярном выражении необходимо сослаться на какую-то из предшествующих групп, следует использовать $\backslash n$, где вместо n подставить номер нужной группы.

regular expressions 101 — an online regex tester for javascript, php, pcre and python.

Your regular expression in: Python

```
" ([^\s]+(?:\.(jpg|gif|png))\.\2)" flags: gmixsu
```

Your test string

```
gffdgdgfg.jpg
```

Your regular expression explained

- 1st Capturing group `([^\s]+(?:\.(jpg|gif|png))\.\2)`
 - `[^\s]+` match a single character not present in the list below
 - Quantifier: Between one and unlimited times, as many times as possible, giving back as needed [greedy]
 - `\s` match any white space character `[\r\n\t\f]`
 - `(?:\.(jpg|gif|png))` Positive Lookahead - Assert that the regex below can be matched
 - `\.` matches the character `.` literally
 - 2nd Capturing group `(jpg|gif|png)`
 - 1st Alternative: `jpg`
`jpg` matches the characters `jpg` literally (case sensitive)
 - 2nd Alternative: `gif`
`gif` matches the characters `gif` literally (case sensitive)
 - 3rd Alternative: `png`
`png` matches the characters `png` literally (case sensitive)
 - `\.` matches the character `.` literally
 - `\2` matches the same text as most recently matched by the 2nd backreference

Your regular expression explained

- 1st Capturing group `([^\s]+(?:\.(jpg|gif|png))\.\2)`
 - `[^\s]+` match a single character not present in the list below
 - Quantifier: Between one and unlimited times, as many times as possible, giving back as needed [greedy]
 - `\s` match any white space character `[\r\n\t\f]`
 - `(?:\.(jpg|gif|png))` Positive Lookahead - Assert that the regex below can be matched
 - `\.` matches the character `.` literally
 - 2nd Capturing group `(jpg|gif|png)`
 - 1st Alternative: `jpg`
`jpg` matches the characters `jpg` literally (case sensitive)
 - 2nd Alternative: `gif`
`gif` matches the characters `gif` literally (case sensitive)
 - 3rd Alternative: `png`
`png` matches the characters `png` literally (case sensitive)
 - `\.` matches the character `.` literally
 - `\2` matches the same text as most recently matched by the 2nd backreference

Match groups

MATCH 1		
1.	[0-12]	<code>'gffdgdgfg.jpg'</code>
2.	[9-12]	<code>'jpg'</code>

http://gskinner.com/RegExr/

The screenshot shows a Mozilla Firefox browser window titled "RegExr - Mozilla Firefox". The address bar displays "gskinner.com/RegExr/". The main content area is divided into two columns. The left column contains a "Match" tab with the input field containing the regular expression `\be(\w*)s\b`. Below the input field are checkboxes for "global" (checked), "ignoreCase", "extended", "dotall", and "multiline", along with a "Share Link" button. A welcome message follows, describing the tool's features: real-time results, code hinting, detailed results, a built-in regex guide, online and desktop availability, saving expressions, and searching community expressions. At the bottom of this section, it lists the tools used: Flex 3 and Spelling Plus Library. The right column has tabs for "Samples", "My Saved", and "Community". Under "Samples", there is a "show all" dropdown and a list of sample expressions: `.`, `\w`, `\W`, `\d`, `\D`, `\s`, `\S`, `[ABC]`, and `[^ABC]`. Below the list, a description states: "Matches any character, except for line breaks if dotall is false." At the bottom of the browser window, the footer shows "0.3.1b built by gskinner.com" and links for "About RegExr" and "Desktop Version".

*Некоторые люди, сталкиваясь с проблемой, думают:
"Знаю, я воспользуюсь регулярными выражениями".
Теперь у них две проблемы.*

Джейми Завински

Благодарю за
внимание  Software Testing Company