# Regex

## []

تمام حروفی که داخل  کروشه هستند یک حرف حساب می شوند [jhb]ack در مثال بالا  jack,hack,back انتخاب می شوند البته می توان یک بازه را نیز داد [A-Za-z]ack تمام کلماتی که آخر آنها ack باشد

## کاراکتر .

نقطه یعنی هر کاراکتری باشد 1.+1   
یعنی هر کلمه ای که با ۱ شروع شود و به ۱ ختم شود مثلا  
1asdf3tt34ttg5767456454@@#$$%%%$#@@##$1  
مچ می شود

## [^ ]

^ در داخل براکت یعنی این کاراکترها نباشد [^aeiou][a-z]+ تمام کلماتی که با حروف صدادار شروع نمی شوند

## خود [ + \*

\[ , \+ , \\* خود کاراکتر را می جوید [ ] + \*

## کاراکتر \* :

کاراکتر قبل از آن صفر یا تعدادی تکرار شود  
10\*1  
معادل است  
11,101,1001,10001,100001

## ()

گروه ساخته می شود ضمنا داخل این پرانتزها می توان از شرطها استفاده کرد (\band\b|\bor\b) تمام and, or  های داخل فایل را انتخاب می کند

## +? , \*?

یعنی غیر حریصانه پترن را چک می کند  
^1.+1  
1000000000010000000001000001000  
مچ می شود حالا اگر میخواستیم فقط ۱ تا ۱ اول مچ شود  
از شمارنده زیر استفاده میشود  
+? ---> ^1.+?1  
1000000000010000000001000001000

## ^ , $

^  
 یعنی ابتدای فایل یا خط  
$  
یعنی انتهای فایل یا خط

## *x*{*n*,*m*}

Where "n" is 0 or a positive integer, "m" is a positive integer, and m > n, matches at least "n" and at most "m" occurrences of the preceding item "x". For example, /a{1,3}/ matches nothing in "cndy", the "a" in "candy", the two "a"'s in "caandy", and the first three "a"'s in "caaaaaaandy".

## Character classes

|  |  |
| --- | --- |
| [abc] | A single character of: a, b or c |
| [^abc] | A character except: a, b or c |
| [a-z] | A character in the range: a-z |
| [^a-z] | A character not in the range: a-z |
| [0-9] | A digit in the range: 0-9 |
| [a-zA-Z] | A character in the range: a-z or A-Z |
| [a-zA-Z0-9] | A character in the range: a-z, A-Z or 0-9 |

## Meta characters

|  |  |
| --- | --- |
| . | Any single character |
| \s | Any whitespace character |
| \S | Any non-whitespace character |
| \d | Any digit, Same as [0-9] |
| \D | Any non-digit, Same as [^0-9] |
| \w | Any word character |
| \W | Any non-word character |
| \X | Any Unicode sequences, linebreaks included |
| \C | Match one data unit |
| \R | Unicode newlines |
| \v | Vertical whitespace character |
| \V | Negation of \v - anything except newlines and vertical tabs |
| \h | Horizontal whitespace character |
| \H | Negation of \h |

## Quantifiers

|  |  |
| --- | --- |
| a? | Zero or one of a |
| a\* | Zero or more of a |
| a+ | One or more of a |
| [0-9]+ | One or more of 0-9 |
| a{3} | Exactly 3 of a |
| a{3,} | 3 or more of a |
| a{3,6} | Between 3 and 6 of a |
| a\* | Greedy quantifier |
| a\*? | Lazy quantifier |
| a\*+ | Possessive quantifier |

## Anchors

|  |  |
| --- | --- |
| ^ | Start of string |
| $ | End of string |
| \A | Start of string |
| \Z | End of string |
| \z | Absolute end of string |
| \b | A word boundary |
| \B | Non-word boundary |

## POSIX Character Classes

|  |  |  |
| --- | --- | --- |
| Character Class | Same as | Meaning |
| [[:alnum:]] | [0-9A-Za-z] | Letters and digits |
| [[:alpha:]] | [A-Za-z] | Letters |
| [[:ascii:]] | [\x00-\x7F] | ASCII codes 0-127 |
| [[:blank:]] | [\t ] | Space or tab only |
| [[:cntrl:]] | [\x00-\x1F\x7F] | Control characters |
| [[:digit:]] | [0-9] | Decimal digits |
| [[:graph:]] | [[:alnum:][:punct:]] | Visible characters (not space) |
| [[:lower:]] | [a-z] | Lowercase letters |
| [[:print:]] | [ -~] == [ [:graph:]] | Visible characters |
| [[:punct:]] | [!"#$%&’()\*+,-./:;<=>?@[]^\_`{|}~] | Visible punctuation characters |
| [[:space:]] | [\t\n\v\f\r ] | Whitespace |
| [[:upper:]] | [A-Z] | Uppercase letters |
| [[:word:]] | [0-9A-Za-z\_] | Word characters |
| [[:xdigit:]] | [0-9A-Fa-f] | Hexadecimal digits |
| [[:<:]] | [\b(?=\w)] | Start of word |
| [[:>:]] | [\b(?<=\w)] | End of word |