# Reference - What does this regex mean?

Asked 5 years, 6 months ago Active 2 months ago Viewed 105k times

52

856

## What is this?

This is a collection of common Q&A. This is also a Community Wiki, so everyone is invited to participate in maintaining it.

## Why is this?

regex is suffering from *give me ze code* type of questions and poor answers with no explanation. This reference is meant to provide links to quality Q&A.

## What's the scope?

This reference is meant for the following languages: php, perl, javascript, python, ruby, java, .net.

This might be too broad, but these languages share the same syntax. For specific features there's the tag of the language behind it, example:

- What are regular expression Balancing Groups? .net

regex

edited Oct 31 '18 at 16:52

community wiki
18 revs, 13 users 32%
HamZa

**locked** by Robert Harvey Apr 8 '14 at 18:46

This question’s answers are a collaborative effort. If you see something that can be improved, just edit the answer to improve it! *No additional answers can be added here.*

Read more about locked posts here.

1 **I created a meta discussion, everyone is invited >>>** – HamZa Apr 8 '14 at 18:26

comments disabled on deleted / locked posts / reviews

## 1 Answer

827

## The Stack Overflow Regular Expressions FAQ

### Online tutorials

- RegexOne
- Regular Expressions Info

### Quantifiers

- Zero-or-more: `*` :greedy, `*?` :reluctant, `*+` :possessive
- One-or-more: `+` :greedy, `+?` :reluctant, `++` :possessive
- `?` :optional (zero-or-one)
- Min/max ranges (all inclusive): `{n,m}` :between n & m, `{n,}` :n-or-more, `{n}` :exactly n
- Differences between greedy, reluctant (a.k.a. "lazy", "ungreedy") and possessive quantifier:
  - Greedy vs. Reluctant vs. Possessive Quantifiers
  - In-depth discussion on the differences between greedy versus non-greedy
  - What's the difference between `{n}` and `{n}?`
  - Can someone explain Possessive Quantifiers to me? php, perl, java, ruby
  - Emulating possessive quantifiers .net
  - Non-Stack Overflow references: From Oracle, regular-expressions.info

### Character Classes

- What is the difference between square brackets and parentheses?

- `[...]` : any one character, `[^...]` : negated/any character but
- `[^]` matches any one character *including* newlines javascript
- `[\w-[\d]]` / `[a-z-[qz]]` : set subtraction .net , xml-schema , xpath , JGSoft
- `[\w&&[^\d]]` : set intersection java , ruby 1.9+
- `[[:alpha:]]` :POSIX character classes
- Why do `[^\\D2]` , `[^[^0-9]2]` , `[^2[^0-9]]` get different results in Java? java
- Shorthand:
  - Digit: `\d` :digit, `\D` :non-digit
  - Word character (Letter, digit, underscore): `\w` :word character, `\W` :non-word character
  - Whitespace: `\s` :whitespace, `\S` :non-whitespace
- Unicode categories ( `\p{L}, \P{L}` , etc.)

### Escape Sequences

- Horizontal whitespace: `\h` :space-or-tab, `\t` :tab
- Newlines:
  - `\r` , `\n` :carriage return and line feed
  - `\R` :generic newline php
- Negated whitespace sequences: `\H` :Non horizontal whitespace character, `\V` :Non vertical whitespace character, `\N` :Non line feed character pcre php5 java-8
- Other: `\v` :vertical tab, `\e` :the escape character

### Anchors

- `^` :start of line/input, `\b` :word boundary, and `\B` :non-word boundary, `$` :end of line/input
- `\A` :start of input, `\Z` :end of input php , perl , ruby
- `\z` :the very end of input ( `\Z` in Python) .net , php , pcre , java , ruby , icu , swift , objective-c
- `\G` :start of match php , perl , ruby

*(Also see "Flavor-Specific Information → Java → The functions in `Matcher` ")*

## Groups

- `(...)` :capture group, `(?:)` :non-capture group
  - Why is my repeating capturing group only capturing the last match?
- `\1` :backreference and capture-group reference, `$1` :capture group reference
  - What's the meaning of a number after a backslash in a regular expression?
  - `\g<1>123` :How to follow a numbered capture group, such as `\1` , with a number?: python
- What does a subpattern `(?i:regex)` mean?
- What does the 'P' in `(?P<group_name>regexp)` mean?
- `(?>)` :atomic group or independent group, `(?|)` :branch reset
  - Equivalent of branch reset in .NET/C# .net
- Named capture groups:
  - General named capturing group reference at `regular-expressions.info`
  - java : `(?<groupname>regex)` : Overview and naming rules *(Non-Stack Overflow links)*
  - Other languages: `(?P<groupname>regex)` python , `(?<groupname>regex)` .net , `(?<groupname>regex)` perl , `(?P<groupname>regex)` and `(?<groupname>regex)` php

## Lookarounds

- Lookaheads: `(?=...)` :positive, `(?!...)` :negative
- Lookbehinds: `(?<=...)` :positive, `(?<!...)` :negative (not supported by javascript )
- Lookbehind limits in:
  - Lookbehinds need to be constant-length php , perl , python , ruby
  - Lookarounds of limited length `{0,n}` java
  - Variable length lookbehinds are allowed .net
- Lookbehind alternatives:

- Using `\K` php , perl (Flavors that support `\K` )
- Alternative regex module for Python python
    - The hacky way
    - JavaScript negative lookbehind equivalents External link

**Modifiers**

- Most flavors: `g` :global, `i` :case-insensitive, `u` :unicode, `x` :whitespace-extended
- `c` :current position perl `e` :expression php perl `o` :once ruby
- `m` :multiline php perl python javascript .net java , `m` :(non)multiline ruby
- `s` :single line (not supported by javascript or ruby ), `s` workaround javascript
- `S` :study php `U` :ungreedy php r
- How to convert preg_replace e to preg_replace_callback?
- What are inline modifiers?
- What is '?-mix' in a Ruby Regular Expression

**Other:**

- `|` :alternation (OR) operator, `.` :any character, `[.]` :literal dot character
- What special characters must be escaped?
- Control verbs ( php and perl ): `(*PRUNE)` , `(*SKIP)` , `(*FAIL)` and `(*F)`
    - php only: `(*BSR_ANYCRLF)`
- Recursion ( php and perl ): `(?R)` , `(?0)` and `(?1)` , `(?-1)` , `(?&groupname)`

**Common Tasks**

- Get a string between two curly braces: `{...}`
- Match (or replace) a pattern except in situations s1, s2, s3...
- How do I find all YouTube video ids in a string using a regex?

- Validation:
    - Internet: email addresses, URLs (host/port: regex and non-regex alternatives), passwords
    - Numeric: a number, min-max ranges (such as 1-31), phone numbers, date
    - *Parsing HTML with regex: See "General Information > When not to use Regex"*

**Advanced Regex-Fu**

- Strings and numbers:
    - Regular expression to match a line that doesn't contain a word
    - How does this PCRE pattern detect palindromes?
    - Match strings whose length is a fourth power
    - How does this regex find triangular numbers?
    - How to determine if a number is a prime with regex?
    - How to match the middle character in a string with regex?
- Other:
    - How can we match a^n b^n with Java regex?
    - Match nested brackets
        - Using a recursive pattern php , perl
        - Using balancing groups .net
    - “Vertical” regex matching in an ASCII “image”
    - List of highly up-voted regex questions on Code Golf
    - How to make two quantifiers repeat the same number of times?
    - An impossible-to-match regular expression: `(?!a)a`
    - Match/delete/replace `this` except in contexts A, B and C
    - Match nested brackets with regex without using recursion or balancing groups?

**Flavor-Specific Information**

*(Except for those marked with* `*` *, this section contains non-Stack Overflow links.)*

- Java

- Official documentation: Pattern Javadoc, Oracle's regular expressions tutorial
- The differences between functions in `java.util.regex.Matcher` :
  - `matches()` ): The match must be anchored to both input-start and -end
  - `find()` ): A match may be anywhere in the input string (substrings)
  - `lookingAt()` : The match must be anchored to input-start only
  - *(For anchors in general, see the section "Anchors")*
- The only `java.lang.String` functions that accept regular expressions: `matches(s)` , `replaceAll(s,s)` , `replaceFirst(s,s)` , `split(s)` , `split(s,i)`
- *An (opinionated and) detailed discussion of the disadvantages of and missing features in `java.util.regex`

- .NET
  - How to read a .NET regex with look-ahead, look-behind, capturing groups and back-references mixed together?
- Official documentation:
  - Boost regex engine: General syntax, Perl syntax *(used by TextPad, Sublime Text, UltraEdit, ...???)*
  - JavaScript 1.5 general info and RegExp object
  - .NET ┆ MySQL ┆ Oracle ┆ Perl5 version 18.2
  - PHP: pattern syntax, `preg_match`
  - Python: Regular expression operations, `search` vs `match` , how-to
  - Splunk: regex terminology and syntax and regex command
  - Tcl: regex syntax, manpage, `regexp` command
  - Visual Studio Find and Replace

### General information

*(Links marked with `*` are non-Stack Overflow links.)*

- Other general documentation resources: Learning Regular Expressions, *Regular-expressions.info, *Wikipedia entry, *RexEgg, Open-Directory Project
- DFA versus NFA
- Generating Strings matching regex
- Books: Jeffrey Friedl's *Mastering Regular Expressions*

- When to *not* use regular expressions:
  - *Some people, when confronted with a problem, think "I know, I'll use regular expressions." Now they have two problems.* (blog post written by Stack Overflow's founder)*
  - Do not use regex to parse HTML:
    - Don't. Please, just don't
    - Well, maybe...if you're *really* determined (other answers in this question are also good)

**Examples of regex that can cause regex engine to fail**

- Why does this regular expression kill the Java regex engine?

**Tools: Testers and Explainers**

*(This section contains non-Stack Overflow links.)*

- Online *(* includes replacement tester, + includes split tester)*:
  - Debuggex (Also has a repository of useful regexes) javascript , python , pcre
  - *Regular Expressions 101 php , pcre , python , javascript
  - Regex Pal, *regular-expressions.info* javascript
  - Rubular ruby RegExr Regex Hero dotnet
  - *+ regexstorm.net .net
  - *RegexPlanet: Java java , Go go , Haskell haskell , JavaScript javascript , .NET dotnet , Perl perl php PCRE php , Python python , Ruby ruby , XRegExp xregexp
  - `freeformatter.com` xregexp
  - *+ `regex.larsolavtorvik.com` php PCRE and POSIX, javascript
  - Refiddle javascript ruby .net
- Offline:
  - Microsoft Windows: RegexBuddy (analysis), RegexMagic (creation), Expresso (analysis, creation, free)

edited Aug 19 at 8:45 community wiki

142 revs, 39 users 25%
aliteralmind