# 正则表达式

<https://deerchao.net/tutorials/regex/regex.htm#getstarted>

入门

\b 是元字符 单词的开头或者结尾，单词的分界处

.是元字符，匹配除了换行符以外的任意字符

\*是数量，前面的内容可以连续重复使用任意次数

\d 一个数字

\s 任意的空白符

\w 数字字母下划线汉字等等

[a-z]括号内的一个字符

^匹配字符串的开始

$匹配字符串的结束

eg: ^\d{5,12}$

重复

|  |  |
| --- | --- |
| \* | 重复零次或更多次 |
| + | 重复一次或更多次 |
| ? | 重复零次或一次 |
| {n} | 重复n次 |
| {n,} | 重复n次或更多次 |
| {n,m} | 重复n到m次 |

分支条件

|

eg: \(0\d{2}\)[- ]?\d{8}|0\d{2}[- ]?\d{8}

转义字符

\. \\* \\

分组

(\d){3}

(\d{1,3}\.){3}\d{1,3}

反义

|  |  |
| --- | --- |
| \W | 匹配任意不是字母，数字，下划线，汉字的字符 |
| \S | 匹配任意不是空白符的字符 |
| \D | 匹配任意非数字的字符 |
| \B | 匹配不是单词开头或结束的位置 |
| [^x] | 匹配除了x以外的任意字符 |
| [^aeiou] | 匹配除了aeiou这几个字母以外的任意字符 |

## 后向引用

\1 分组1匹配的文本

分组0对应整个正则表达式

分组的命名

(?<name>\w+) 或者 (?'name'\w+)

反向引用就\k<name>

|  |  |  |
| --- | --- | --- |
| **常用分组语法** | | |
| **分类** | **代码/语法** | **说明** |
| **捕获** | (exp) | 匹配exp,并捕获文本到自动命名的组里 |
| (?<name>exp) | 匹配exp,并捕获文本到名称为name的组里，也可以写成(?'name'exp) |
| (?:exp) | 匹配exp,不捕获匹配的文本，也不给此分组分配组号 |
| **零宽断言** | (?=exp) | 匹配exp前面的位置 |
| (?<=exp) | 匹配exp后面的位置 |
| (?!exp) | 匹配后面跟的不是exp的位置 |
| (?<!exp) | 匹配前面不是exp的位置 |
| **注释** | (?#comment) | 这种类型的分组不对正则表达式的处理产生任何影响，用于提供注释让人阅读 |

## 零宽断言

| **符号** | **解释** | **示例** | **说明** |
| --- | --- | --- | --- |
| . | 匹配任意字符 | b.t | 可以匹配bat / but / b#t / b1t等 |
| \w | 匹配字母/数字/下划线 | b\wt | 可以匹配bat / b1t / b\_t等 但不能匹配b#t |
| \s | 匹配空白字符（包括\r、\n、\t等） | love\syou | 可以匹配love you |
| \d | 匹配数字 | \d\d | 可以匹配01 / 23 / 99等 |
| \b | 匹配单词的边界 | \bThe\b |  |
| ^ | 匹配字符串的开始 | ^The | 可以匹配The开头的字符串 |
| $ | 匹配字符串的结束 | .exe$ | 可以匹配.exe结尾的字符串 |
| \W | 匹配非字母/数字/下划线 | b\Wt | 可以匹配b#t / b@t等 但不能匹配but / b1t / b\_t等 |
| \S | 匹配非空白字符 | love\Syou | 可以匹配love#you等 但不能匹配love you |
| \D | 匹配非数字 | \d\D | 可以匹配9a / 3# / 0F等 |
| \B | 匹配非单词边界 | \Bio\B |  |
| [] | 匹配来自字符集的任意单一字符 | [aeiou] | 可以匹配任一元音字母字符 |
| [^] | 匹配不在字符集中的任意单一字符 | [^aeiou] | 可以匹配任一非元音字母字符 |
| \* | 匹配0次或多次 | \w\* |  |
| + | 匹配1次或多次 | \w+ |  |
| ? | 匹配0次或1次 | \w? |  |
| {N} | 匹配N次 | \w{3} |  |
| {M,} | 匹配至少M次 | \w{3,} |  |
| {M,N} | 匹配至少M次至多N次 | \w{3,6} |  |
| | | 分支 | foo|bar | 可以匹配foo或者bar |
| (?#) | 注释 |  |  |
| (exp) | 匹配exp并捕获到自动命名的组中 |  |  |
| (? <name>exp) | 匹配exp并捕获到名为name的组中 |  |  |
| (?:exp) | 匹配exp但是不捕获匹配的文本 |  |  |
| (?=exp) | 匹配exp前面的位置 | \b\w+(?=ing) | 可以匹配I'm dancing中的danc |
| (?<=exp) | 匹配exp后面的位置 | (?<=\bdanc)\w+\b | 可以匹配I love dancing and reading中的第一个ing |
| (?!exp) | 匹配后面不是exp的位置 |  |  |
| (?<!exp) | 匹配前面不是exp的位置 |  |  |
| \*? | 重复任意次，但尽可能少重复 | a.\*b a.\*?b | 将正则表达式应用于aabab，前者会匹配整个字符串aabab，后者会匹配aab和ab两个字符串 |
| +? | 重复1次或多次，但尽可能少重复 |  |  |
| ?? | 重复0次或1次，但尽可能少重复 |  |  |
| {M,N}? | 重复M到N次，但尽可能少重复 |  |  |
| {M,}? | 重复M次以上，但尽可能少重复 |  |  |