

正则表达式语法

本课程仅用于教学用途

中华人民共和国网络安全法

中华人民共和国刑法

社会主义核心价值观

详情百度搜索以上“关键词”+site:gov.cn

参考资料

1~历史渊源

1951 年，一位名叫Stephen Kleene的数学科学家，他在Warren McCulloch和Walter Pitts早期工作的基础之上，发表了一篇题目是《神经网络事件的表示法》的论文，利用称之为正则集合的数学符号来描述此模型，引入了正则表达式的概念。正则表达式被作为用来描述其称之为“正则集的代数”的一种表达式，因而采用了“正则表达式”这个术语。

2~英国通信总部

<https://gchq.github.io/CyberChef/>

3~notepad++

<https://notepad-plus-plus.org/>

4~正则表达式测试工具与样例

<https://c.runoob.com/front-end/854>

https://mp.weixin.qq.com/s/IVCBtYa_cjYRUI7mBCO5ug

第一部分

从左往右，从上往下，运算级别高的优先

\

() []

* + ? {}

^ \$

|

第一部分（上）

1~^a a开头, 例如ab/a1

2~a\$ a结尾, 例如ba/1a

3~^ab\$ a开头, b结尾, 就是ab

4~ab. ab后面跟上1个字符, 点表示除\n（换行符）, \r（回车符）的任意字符, 例如abc/ab1

5~^a.b\$ a开头, b结尾, 中间一个任意字符, 例如a1b/acb

6~ab* a后面跟0个及以上的b, 例如a/ab/abb

第一部分（下）

7~ab+ a后面跟1个及以上的b, 例如ab/abb

8~ab? a后面跟0个或者1个b, 仅表示a/ab

9~ab.* ab后面跟0个及以上的字符, 例如ab/abc/ab123

10~.*ab ab之前有0个及以上的字符, 例如1ab/aab/1aaab

11~ab\.com 表示ab.com, 所有正则特殊符号如果按照原意, 必须转义处理

第二部分（上）

1~ $ab\{1\}$ a后跟1个b, 数字只能是自然数

2~ $ab\{1,3\}$ a后跟1到3个b, $ab/abb/abbb$ 中的1个

3~ $ab\{1,\}$ a后跟1个及以上的b, 例如 $ab/abb/abbbbb$

4~ $^.\{3\}\$$ 任意相连3个字符的字符串, 例如 $111/a1a/bbb$

5~ $a|b|2$ $a/b/2$ 中的1个

6~ $a|bc$ a/bc 中的1个

7~ $(a|b)cd$ acd/bcd 中的1个

第二部分（下）

8~[ab]cd acd/bcd中的1个

9~[^ab]cd 不能是acd/bcd中任何1个

10~[a-z]ab ab之前是a到z某个字符，例如aab/zab/rab

11~[A-Z]ab ab之前是A到Z的某个字符，例如Aab/Rab

12~[0-9]ab ab之前是0到9的某个字符，例如0ab/7ab

13~[0-8a-x]ab ab之前0-8或者a-x的某个字符

14~[^0-8a-x]ab ab之前不能是0-8或者a-x的某个字符

第三部分

1~a\b

\b是字符串边界（本身不是字符），a后面必须有空白字符，匹配a c

2~\ba a前面必须有空白字符

3~a\B \B是不能有边界，a后面不能有空白字符，a c不符合要求

4~\Ba a前面不能有空白字符

5~\d [0-9]中的1个

6~\D [^0-9]，不能有[0-9]这些字符

7~\w 大小写字母，数字，下划线的字符集[A-Za-z0-9_]中的1个，里面包含下划线

8~\W [^A-Za-z0-9_]，不能有[A-Za-z0-9_]中的字符

第四部分（上）

1~\t 水平制表符，对应 \cl或\x09，a\tb表示ab之间有1个水平制表符

2~\n 换行符，对应\cJ或\x0a，a\nb表示ab之间有1个换行符

3~\v 垂直制表符，对应\cK或\x0b，a\v b表示ab之间有1个垂直制表符

4~\f

换页符，对应\cL或\x0c，a\fb表示ab之间有1个换页符，文档上下页之间的空字符

5~\r 回车符，对应\cM或\x0d，a\r b表示ab之间有1个回车符

6~\s 匹配任何空白字符的集合，[\f\n\r\t\v]隶属于\s，[\f\n\r\t\v]不等于\s，例如空格

7~\S 匹配任何非空白字符的集合，与\s正好相反

第四部分（下）

`a\nb\nc`, `a\cJb\cJc`, `a\x0ab\x0ac` (16进制)

美国信息交换标准代码ASCII表 (其中16进制均可借鉴使用到正则表达式中)

| Bin | Oct | Dec | Hex | 英文名 | 中文名 |
|-----------|-----|-----|------|-----------------------------|-------|
| 0000 1001 | 011 | 9 | 0x09 | HT (horizontal tab) | 水平制表符 |
| 0000 1010 | 012 | 10 | 0x0A | LF (NL line feed, new line) | 换行键 |
| 0000 1011 | 013 | 11 | 0x0B | VT (vertical tab) | 垂直制表符 |
| 0000 1100 | 014 | 12 | 0x0C | FF (NP form feed, new page) | 换页键 |
| 0000 1101 | 015 | 13 | 0x0D | CR (carriage return) | 回车键 |

第五部分

1~[:alpha:] 任何字母

2~[:digit:] 任何数字

3~[:alnum:] 任何字母和数字

4~[:space:] 任何空白字符

5~[:upper:] 任何大写字母

6~[:lower:] 任何小写字母

7~[:punct:] 任何标点符号

8~[:xdigit:] 任何16进制的数字, 含有[0-9a-fA-F]

使用范围有限, php编码, notepad++也ok

第六部分

1~a\w+b 匹配a1b2b, 不是a1b

2~a\w+?b 倾向匹配a1b2b中的a1b

3~? 尽可能少匹配字符, 适用于+*?{n,m}{n,}

4~(?i)ABCD (?i)表示忽略其后字符的大小写, 可以匹配ABCD, 也可以abcd

5~(?s)ab.*cd (?s)促使点可以表示所有字符, 包括\n, \r, 属于能量加成

6~(?m)^\d+\w+\$ 例如一个含有\n的完整字符被分割匹配

7~(?m)促使^可以匹配字符串开头以及行的开头, \n之后

8~(?m)促使\$可以匹配字符串结尾以及行的结尾, \n之前

第七部分（上）

1~ $a(?:b|c)$ 等同 $a(b|c)$ 或 $ab|ac$

2~ $a(?=b|c|d)$ a 的右侧必须有 $b/c/d$ 中的1个

3~ $a(?:!b|c|d)$ a 的右侧不能有 $b/c/d$ 中的1个

4~ $(?<!b|c|d)a$ a 的左侧不能有 $b/c/d$ 中的1个

5~ $(?<=b|c|d)a$ a 的左侧必须有 $b/c/d$ 中的1个

6~ $(\backslash d)\backslash 1$ 连续2个相同数字, $(\backslash d)\backslash 1\{1\}$

7~ $(\backslash d)(\backslash d)\backslash 2$ 连续3个相同的数字, $(\backslash d)\backslash 1\{2\}$

8~ $(\backslash d)(\backslash d)(\backslash d)\backslash 3$ 连续4个相同的数字, $(\backslash d)\backslash 1\{3\}$

9~ $(\backslash d)\backslash 1\{1\}(\backslash d)\backslash 2\{1\}$ 例如1122, 1133, 2244, 1111

第七部分（下）

1~(0(?=1)|1(?=2)|2(?=3)|3(?=4)|4(?=5)|5(?=6)|6(?=7)|7(?=8)|8(?=9)|9(?=0)){0}\d

任意1个数字匹配

2~(0(?=1)|1(?=2)|2(?=3)|3(?=4)|4(?=5)|5(?=6)|6(?=7)|7(?=8)|8(?=9)|9(?=0)){1}\d

2个连续升序数字为一组

3~(0(?=1)|1(?=2)|2(?=3)|3(?=4)|4(?=5)|5(?=6)|6(?=7)|7(?=8)|8(?=9)|9(?=0)){2}\d

3个连续升序数字为一组

4~((0(?=1)|1(?=2)|2(?=3)|3(?=4)|4(?=5)|5(?=6)|6(?=7)|7(?=8)|8(?=9)|9(?=0)){1}\d)\

1{1}

例如1212是一组

字母雷同，降序雷同

第八部分（上）

1~Domain

www.baidu.com

`[a-zA-Z0-9][-a-zA-Z0-9]{0,62}(\.[a-zA-Z0-9][-a-zA-Z0-9]{0,62})+`

`(\w{1,20}\.)+\w{1,10}` 简化版

65535（端口数量） 63（域名中子域名字段不超过） 253（域名总长度不超过）

2~IPV4

192.168.100.100

`((\d)|([1-9]\d)|(1\d{2})|(2[0-4]\d)|(25[0-5]))(\.){3}((25[0-5])|(2[0-4]\d)|(1\d{2})|([1-9]\d)|(\d))`

`(\d) ([1-9]\d) (1\d{2}) (2[0-4]\d) (25[0-5])`

`(\d{1,3}\.){3}\d{1,3}` 简化版

`\d{1,3}`

第八部分 (下)

3~邮箱

123456@qq.com

`\w+([-+.] \w+)*@ \w+([-.] \w+)*\ . \w+([-.] \w+)*`

`\w+@ \w+ \.[a-z]{2,}` 简化版

4~手机号

188 6450 0763

`((13[0-9])|(14[5,7])|(15[0-3,5-9])|(17[0,3,5-8])|(18[0-9])|166|198|199|147)\d{8}`

`1\d{10}` 简化版

5~身份证号

36 07 24 19981016 601 3

`[1-9]\d{5}(18|19|20)\d{2}((0[1-9])|(1[0-2]))(((0[2-9]|10|20|30|31)\d{3}[0-9Xx]`

`\d{17})[\dxX]` 简化版

第九部分

通配符

```
mkdir ceshi touch ceshi.txt 1.txt c.txt
```

```
ls *.txt
```

```
find /root/ceshi -name 'c*.txt'
```

找到root目录下，c开头的txt文档，找到c.txt，ceshi.txt

```
find /root/ceshi -name '?eshi.txt'
```

找到结尾是eshi的txt文档，找到ceshi.txt

```
find /root/ceshi -name '[ceshi].txt'
```

范围内的1个字符，找到c.txt

```
find /root/ceshi -name '[!ceshi].txt'
```

范围之外的1个字符，1.txt，!可以用^代替

制作视频相关软件

1-WPS Office

2~迅捷屏幕录像工具

3~迅捷视频转换器

共勉

早睡早起，学好英语。
少看手机，多多实践。

感谢筒子们观看咯！