正则表达式

2021年1月16日 10:23

?

? 前面的字符可出现0次或1次

used? 可以匹配到use、used

{}

{}用于指定前面字符出现的次数

ab{10}c匹配到 abbbbbbbbbbc

ab{2,3}c匹配到abbc、abbbc

ab{2,}c匹配到b出现两次以上，abb……c

*

*前面的字符可有可无（0~∞）

ab*c可以匹配到ac、abc、abbbbc……

+

+前面的字符可以出现多次（一次及以上）

ab+c可以匹配到abc、abbbbbbc

（）

如若要对多个字符进行匹配，可以将其用（）括起来

a（bc）*可以匹配到a、abc、abcbc……

| （或）

a （cat|dog）匹配到 a cat、a dog

[]

[]匹配仅含中间的元素部分

[a-z]匹配所有仅含小写字母的部分

[^]

[^]匹配不含其中间元素的部分

[^a]匹配不含a的部分

.

.表示任意字符（但不包含换行符）

\d匹配数字 等价于[0-9]

\D匹配非数字 等价于[^0-9]

\w匹配字母、数字、下划线等单词字符 等价于[0-9a-zA-Z_]

\W匹配非单词字符 等价于[^0-9a-zA-Z_]

\s匹配空格、换行、Tab等所有空白

\S匹配非空白

\b匹配单词边界，即字符末尾与空白之间的位置

^匹配行首 $匹配行尾

^a表示位于行首的a，$a表示位于行尾的a

贪婪模式&非贪婪模式

由于RE在使用的过程中，尽可能多的进行匹配。

- 如在html中aa test1 bb test2 cc
  - 用 .+ 进行一次匹配，当匹配到 test1 时，由于test2后面还有 ,贪婪模式下尽可能多的匹配，将自动匹配到test2右侧，即匹配结果为 test1 bb test2
  - 用 .+? 进行一次匹配，由于此时为非贪婪模式，在匹配到 test1 时结束了此次匹配。即匹配结果为 test1

由于*+./等字符在RE中有实际含义，而在一些地方匹配时会用到此些字符，（如IPv4地址中含有.）在匹配这些特殊字符时前面加上转义字符/