

正则表达式的常用法

1. `.*`

`.` 表示 匹配除换行符 `\n` 之外的任何单字符, `*` 表示零次或多次。所以 `.*` 在一起就表示任意字符出现零次或多次。没有 `?` 表示贪婪模式。比如 `a.*b`, 它将会匹配最长的以 `a` 开始, 以 `b` 结束的字符串。如果用它来搜索 `aabab` 的话, 它会匹配整个字符串 `aabab`。这被称为贪婪匹配。

又比如模式 `src=.*``, 它将会匹配最长的以 `src=`` 开始, 以 ``` 结束的最长的字符串。用它来搜索 `` 时, 将会返回 `src=`test.jpg` width=`60px` height=`80px``

2. `.*?`

`?` 跟在 `*` 或者 `+` 后边用时, 表示懒惰模式。也称非贪婪模式。就是匹配尽可能少的字符。就意味着匹配任意数量的重复, 但是在能使整个匹配成功的前提下使用最少的重复。

`a.*?b` 匹配最短的, 以 `a` 开始, 以 `b` 结束的字符串。如果把它应用于 `aabab` 的话, 它会匹配 `aab` (第一到第三个字符) 和 `ab` (第四到第五个字符)。

又比如模式 `src=.*?`, 它将会匹配 `src=`` 开始, 以 ``` 结束的尽可能短的字符串。且开始和结束中间可以没有字符, 因为 `*` 表示零到多个。用它来搜索 `` 时, 将会返回 `src=``。

3. `.+?`

同上, `?` 跟在 `*` 或者 `+` 后边用时, 表示懒惰模式。也称非贪婪模式。就意味着匹配

任意数量的重复，但是在能使整个匹配成功的前提下使用最少的重复。

`a.+?b` 匹配最短的，以 `a` 开始，以 `b` 结束的字符串，但 `a` 和 `b` 中间至少要有有一个字符。如果把它应用于 `ababccaab` 的话，它会匹配 `abab`（第一到第四个字符）和 `aab`（第七到第九个字符）。注意此时匹配结果不是 `ab,ab` 和 `aab`。因为 `a` 和 `b` 中间至少要有有一个字符。

又比如模式 `src=\.+?`，它将会匹配 `src=` 开始，以 `.` 结束的尽可能短的字符串。

且开始和结束中间必须有字符，因为 `+` 表示 1 到多个。用它来搜索 `` 时，将会返回 `src=`test.jpg``。注意与 `*?` 时的区别，此时不会匹配 `src=``，因为 `src=`` 和 ``` 之间至少有一个字符。

正则元字符

\	将下一个字符标记为一个特殊字符、或一个原义字符、或一个 向后引用、或一个八进制转义符。例如，'n' 匹配字符 "n"。'\n' 匹配一个换行符。序列 '\\ 匹配 "\" 而 \"(\" 则匹配 "("。
^	匹配输入字符串的开始位置。如果设置了 RegExp 对象的 Multiline 属性，^ 也匹配 '\n' 或 '\r' 之后的位置。
\$	匹配输入字符串的结束位置。如果设置了 RegExp 对象的 Multiline 属性，\$ 也匹配 '\n' 或 '\r' 之前的位置。
*	匹配前面的子表达式零次或多次。例如，zo* 能匹配 "z" 以及 "zoo"。* 等价于{0,}。
+	匹配前面的子表达式一次或多次。例如，'zo+' 能匹配 "zo" 以及 "zoo"，但不能匹配 "z"。 + 等价于 {1,}。
?	匹配前面的子表达式零次或一次。例如，"do(es)?" 可以匹配 "do" 或 "does" 。? 等价于 {0,1}。
{n}	n 是一个非负整数。匹配确定的 n 次。例如，'o{2}' 不能匹配 "Bob" 中的 'o'，但是能匹配 "food" 中的两个 o。
{n,}	n 是一个非负整数。至少匹配 n 次。例如，'o{2,}' 不能匹配 "Bob" 中的 'o'，但能匹配 "fooooo" 中的所有 o。'o{1,}' 等价于 'o+'。'o{0,}' 则等价于 'o*'。
{n,m}	m 和 n 均为非负整数，其中 n <= m。最少匹配 n 次且最多匹配 m 次。例如，"o{1,3}" 将匹配 "fooooood" 中的前三个 o。'o{0,1}' 等价于 'o?'。请注意在逗号和两个数之间不能有空格。

?	当该字符紧跟在任何一个其他限制符 (*, +, ?, {n}, {n,}, {n,m}) 后面时，匹配模式是非贪婪的。非贪婪模式尽可能少的匹配所搜索的字符串，而默认的贪婪模式则尽可能多的匹配所搜索的字符串。例如，对于字符串 "oooo", 'o+?' 将匹配单个 "o", 而 'o+' 将匹配所有 'o'。
.	匹配除换行符 (\n、\r) 之外的任何单个字符。要匹配包括 '\n' 在内的任何字符，请使用像 "(. \n)" 的模式。
(pattern)	匹配 pattern 并获取这一匹配。所获取的匹配可以从产生的 Matches 集合得到，在 VBScript 中使用 SubMatches 集合，在 JScript 中则使用 \$0...\$9 属性。要匹配圆括号字符，请使用 '\(' 或 '\)'。
(?:pattern)	匹配 pattern 但不获取匹配结果，也就是说这是一个非获取匹配，不进行存储供以后使用。这在使用 "或" 字符 () 来组合一个模式的各个部分是很有用。例如， 'industr(?:y ies) 就是一个比 'industry industries' 更简略的表达式。
(?=pattern)	正向肯定预查 (look ahead positive assert)，在任何匹配 pattern 的字符串开始处匹配查找字符串。这是一个非获取匹配，也就是说，该匹配不需要获取供以后使用。例如， "Windows(=95 98 NT 2000)" 能匹配 "Windows2000" 中的 "Windows", 但不能匹配 "Windows3.1" 中的 "Windows"。预查不消耗字符，也就是说，在一个匹配发生后，在最后一次匹配之后立即开始下一次匹配的搜索，而不是从包含预查的字符之后开始。
(?!pattern)	正向否定预查(negative assert)，在任何不匹配 pattern 的字符串开始处匹配查找字符串。这是一个非获取匹配，也就是说，该匹配不需要获取供以后使用。例如 "Windows(?!95 98 NT 2000)" 能匹配 "Windows3.1" 中的 "Windows", 但不能匹配 "Windows2000" 中的 "Windows"。预查不消耗字符，也就是说，在一个匹配发生后，在最后一次匹配之后立即开始下一次匹配的搜索，而不是从包含预查的字符之后开始。
(?<=pattern)	反向(look behind)肯定预查，与正向肯定预查类似，只是方向相反。例如， "(?<=95 98 NT 2000)Windows" 能匹配 "2000Windows" 中的 "Windows", 但不能匹配 "3.1Windows" 中的 "Windows"。

(?<!pattern)	反向否定预查，与正向否定预查类似，只是方向相反。例如 "(?<!95 98 NT 2000)Windows"能匹配"3.1Windows"中的"Windows"，但不能匹配"2000Windows"中的"Windows"。
x y	匹配 x 或 y。例如，'z food' 能匹配 "z" 或 "food"。'(z f)ood' 则匹配 "zood" 或 "food"。
[xyz]	字符集合。匹配所包含的任意一个字符。例如， '[abc]' 可以匹配 "plain" 中的 'a'。
[^xyz]	负值字符集合。匹配未包含的任意字符。例如， '[^abc]' 可以匹配 "plain" 中的 'p'、'l'、'i'、' '、'n'。
[a-z]	字符范围。匹配指定范围内的任意字符。例如， '[a-z]' 可以匹配 'a' 到 'z' 范围内的任意小写字母字符。
[^a-z]	负值字符范围。匹配任何不在指定范围内的任意字符。例如， '[^a-z]' 可以匹配任何不在 'a' 到 'z' 范围内的任意字符。
\b	匹配一个单词边界，也就是指单词和空格间的位置。例如， 'er\b' 可以匹配"never" 中的 'er'，但不能匹配 "verb" 中的 'er'。
\B	匹配非单词边界。'er\B' 能匹配 "verb" 中的 'er'，但不能匹配 "never" 中的 'er'。
\cx	匹配由 x 指明的控制字符。例如， \cM 匹配一个 Control-M 或回车符。x 的值必须为 A-Z 或 a-z 之一。否则，将 c 视为一个原义的 'c' 字符。
\d	匹配一个数字字符。等价于 [0-9]。
\D	匹配一个非数字字符。等价于 [^0-9]。
\f	匹配一个换页符。等价于 \x0c 和 \cL。
\n	匹配一个换行符。等价于 \x0a 和 \cJ。
\r	匹配一个回车符。等价于 \x0d 和 \cM。
\s	匹配任何空白字符，包括空格、制表符、换页符等等。等价于 [\f\n\r\t\v]。

\S	匹配任何非空白字符。等价于 <code>[^\f\n\r\t\v]</code> 。
\t	匹配一个制表符。等价于 <code>\x09</code> 和 <code>\cl</code> 。
\v	匹配一个垂直制表符。等价于 <code>\x0b</code> 和 <code>\cK</code> 。
\w	匹配字母、数字、下划线。等价于 <code>'[A-Za-z0-9_]'</code> 。
\W	匹配非字母、数字、下划线。等价于 <code>'[^A-Za-z0-9_]'</code> 。
\xn	匹配 <code>n</code> ，其中 <code>n</code> 为十六进制转义值。十六进制转义值必须为确定的两个数字长。例如， <code>'\x41'</code> 匹配 <code>"A"</code> 。 <code>'\x041'</code> 则等价于 <code>'\x04' & "1"</code> 。正则表达式中可以使用 <code>ASCII</code> 编码。
\num	匹配 <code>num</code> ，其中 <code>num</code> 是一个正整数。对所获取的匹配的引用。例如， <code>'(.)\1'</code> 匹配两个连续的相同字符。
\n	标识一个八进制转义值或一个向后引用。如果 <code>\n</code> 之前至少 <code>n</code> 个获取的子表达式，则 <code>n</code> 为向后引用。否则，如果 <code>n</code> 为八进制数字 <code>(0-7)</code> ，则 <code>n</code> 为一个八进制转义值。
\nm	标识一个八进制转义值或一个向后引用。如果 <code>\nm</code> 之前至少有 <code>nm</code> 个获得子表达式，则 <code>nm</code> 为向后引用。如果 <code>\nm</code> 之前至少有 <code>n</code> 个获取，则 <code>n</code> 为一个后跟文字 <code>m</code> 的向后引用。如果前面的条件都不满足，若 <code>n</code> 和 <code>m</code> 均为八进制数字 <code>(0-7)</code> ，则 <code>\nm</code> 将匹配八进制转义值 <code>nm</code> 。
\nml	如果 <code>n</code> 为八进制数字 <code>(0-3)</code> ，且 <code>m</code> 和 <code>l</code> 均为八进制数字 <code>(0-7)</code> ，则匹配八进制转义值 <code>nm</code> <code>l</code> 。
\un	匹配 <code>n</code> ，其中 <code>n</code> 是一个用四个十六进制数字表示的 <code>Unicode</code> 字符。例如， <code>\u00A9</code> 匹配版权符号 <code>(?)</code> 。