

分类： 系统运维

<http://renren.com/profile.do?id=223774257&from=0>

字符	描述
\	标记下一个字符是特殊字符或文字。例如, "n" 和字符 "n" 匹配。"n" 则和换行字符匹配。序列 "\\\" 和 "\\\" 匹配, 而 \"(\" 则和 \"(\" 匹配。
^	匹配输入的开头。
\$	匹配输入的末尾。
*	匹配前一个字符零或多次。例如, "zo*" 与 "z" 或 "zoo" 匹配。
+	匹配前一个字符一次或多次。例如, "zo+" 与 "zoo" 匹配, 但和 "z" 不匹配。
?	匹配前一个字符零或一次。例如, "a?ve?" 和 "never" 中的 "ve" 匹配。
.	匹配除换行字符外的任何单个字符。
(pattern)	匹配 pattern 并记住该匹配。匹配上的子字符串, 可以使用 Item [0]...[n], 来从生成的 Matches 集合中取回。要匹配圆括号字符 ( ), 则需使用 \"(\" 或 \")\"。
x y	匹配 x 或 y。例如, "z food" 和 "z" 或 "food" 匹配。"(z f)ood" 匹配 "zoo" 或 "food"。
{n}	n 是非负整数。共匹配 n 次。例如, "o{2}" 和 "Bob" 中的 "o" 不匹配, 但和 "foooooo" 中的前两个 o 匹配。
{n,}	n 是一个非负整数。至少匹配 n 次。例如, "o{2,}" 和 "Bob" 中的 "o" 不匹配, 但和 "foooooo" 中的所有 o 匹配。"o{1,}" 与 "o+" 等效。"o{0,}" 和 "o*" 等效。
{n,m}	m 和 n 是非负整数。至少匹配 n 次而至多匹配 m 次。例如, "o{1,3}" 和 "foooooo" 中的前三个 o 匹配。"o{0,1}" 和 "o?" 等效。
[ xyz]	字符集合。匹配括号内的任一字符。例如, "[abc]" 和 "plain" 中的 "a" 匹配。
[^xyz ]	否定字符集合。匹配非括号内的任何字符。例如, "[^abc]" 和 "plain" 中的 "p" 匹配。
[a-z]	字符范围。和指定范围内的任一字符匹配。例如, "[a-z]" 匹配 "a" 到 "z"范围内的任一小写的字母表字符。
[^m-z]	否定字符范围。匹配不在指定范围内的任何字符。例如, "[m-z]" 匹配不在 "m" 到 "z"范围内的任何字符。
\b	匹配字的边界, 也就是说, 在字和空格之间的位置。例如, "er\b" 和 "never" 中的 "er" 匹配, 但和 "verb"中的 "er" 不匹配。
\B	匹配非字边界。"ea*r\B" 和 "never early" 中的 "ear" 匹配。
\d	匹配数字字符。等价于 [0-9]。
\D	匹配非数字字符。等价于 [^0-9]。
\f	匹配换页字符。
\n	匹配换行字符。
\r	匹配回车字符。
\s	匹配任何空白, 包括空格、制表、换页等。与 "[ \f\n\r\t\v]" 等效。
\S	匹配任何非空白字符。与 "[^ \f\n\r\t\v]" 等效。
\t	匹配制表字符。
\v	匹配垂直制表符。
\w	匹配包括下划线在内的任何字字符。与 "[A-Za-z0-9_]" 等效。
\W	匹配任何非字字符。与 "[^A-Za-z0-9_]" 等效。
\num	匹配 num, 其中 num 是一个正整数。返回记住的匹配的引用。例如, "(.)\1" 匹配两个连续的同字符。
\n	匹配 n, 其中 n 是八进制换码值。八进制换码值必须是 1、2、或 3 位长。例如, "\\11" 和 "\\011" 都匹配制表字符。"\\0011" 和 "\\001" & "1" 是等效的。八进制换码值必须不超过 256。如果超过了, 则只有前两位组成表达式。允许在正则表达式中使用 ASCII 码。
\xn	匹配 n, 其中 n 是十六进制换码值。十六进制换码值必须正好是两位长。例如, "\\x41" 与 "A" 匹配。"\\x041"和 "\\x04" & "1" 是等效的。允许在正则表达式中使用 ASCII 码。

先让我们看两个特殊的符号 '^'和 '\$'。他们的作用是分别指出一个字符串的开始和结束。例子如下：  
"^The": 表示所有以"The"开始的字符串 ("There", "The cat"等)；  
"of despair\$": 表示所以以"of despair"结尾的字符串；  
"^abc\$": 表示开始和结尾都是"abc"的字符串——呵呵，只有"abc"自己了；  
"notice": 表示任何包含"notice"的字符串。  
象最后那个例子，如果你不使用两个特殊字符，你就在表示要查找的串在被查找串的任意部分——你并不把它定位在某一个顶端。  
其它还有 '\*', '+'和 '?'这三个符号，表示一个或一序列字符重复出现的次数。它们分别表示“没有或更多”，“一次或更多”还有“没有或一次”。下面是几个例子：  
"ab\*": 表示一个字符串有一个a后面跟着零个或若干个b。（"a", "ab", "abbb",.....）；  
"ab+": 表示一个字符串有一个a后面跟着至少一个b或者更多；  
"ab?": 表示一个字符串有一个a后面跟着零个或者一个b；  
"a?b+\$": 表示在字符串的末尾有零个或一个a跟着一个或几个b。  
你也可以使用范围，用大括号括起，用以表示重复次数的范围。  
"ab{2}": 表示一个字符串有一个a跟着2个b ("abb")；

"ab{2,}": 表示一个字符串有一个a跟着至少2个b;  
"ab{3,5}": 表示一个字符串有一个a跟着3到5个b。  
请注意,你必须指定范围的下限(如: "{0,2}"而不是"{,2}")。还有,你可能注意到了, '\*'、'+'和 '?'相当于"{0,}"、"{1,}"和"{0,1}"。

表列出按标准 Unicode 字符属性进行匹配的语法。两个字母的缩写词与 Unicode 字符属性数据库中所列的一样。可将这些指定为字符集的一部分。例如,表达式 [:Nd:Nl:No] 匹配任何种类的数字。表达式语法说明  
大写字母 :Lu 匹配任何一个大写字母。例如,:Luhe 匹配"The"但不匹配"the"。  
小写字母 :Ll 匹配任何一个大写字母。例如,:Llhe 匹配"the"但不匹配"The"。  
词首大写字母 :Lt 匹配将大写字母和小写字母结合的字符,例如,Nj 和 Dz。  
修饰符字母 :Lm 匹配字母或标点符号,例如逗号、交叉重音符和双撇号,用于表示对前一字母的修饰。  
其他字母 :Lo 匹配其他字母,如哥特体字母 ahsa。  
十进制数 :Nd 匹配十进制数(如 0-9)和它们的双字节等效数。  
字母数字 :Nl 匹配字母数字,例如罗马数字和表意数字零。  
其他数字 :No 匹配其他数字,如旧斜体数字一。  
开始标点符号 :Ps 匹配开始标点符号,例如左方括号和左大括号。  
结束标点符号 :Pe 匹配结束标点符号,例如右方括号和右大括号。  
左引号 :Pi 匹配左双引号。  
右引号 :Pf 匹配单引号和右双引号。  
破折号 :Pd 匹配破折号标记。  
连接符号 :Pc 匹配下划线标记。  
其他标点符号 :Po 匹配逗号(,)、?、"、!、@、#、%、&、\*、\、冒号(:)、分号(;)、' 和 /。  
空白分隔符 :Zs 匹配空白。  
行分隔符 :Zl 匹配 Unicode 字符 U+2028。  
段落分隔符 :Zp 匹配 Unicode 字符 U+2029。  
无间隔标记 :Mn 匹配无间隔标记。  
组合标记 :Mc 匹配组合标记。  
封闭标记 :Me 匹配封闭标记。  
数学符号 :Sm 匹配 +、=、~、|、< 和 >。  
货币符号 :Sc 匹配 \$ 和其他货币符号。  
修饰符号 :Sk 匹配修饰符号,如抑扬音、抑音符号和长音符号。  
其他符号 :So 匹配其他符号,如版权符号、段落标记和度数符号。  
其他控制 :Cc 匹配行尾。  
其他格式 :Cf 格式化控制字符,例如双向控制字符。  
代理项 :Cs 匹配代理项对的一半。  
其他私用 :Co 匹配私用区域的任何字符。  
其他未分配的字符 :Cn 匹配未映射到 Unicode 字符的字符。

表达式语法说明  
Alpha :Al 匹配任何一个字符。例如,:Alhe 匹配"The"、"then"、"reached"等单词。  
数字 :Nu 匹配任何一个数或数字。  
标点符号 :Pu 匹配任何一个标点符号,如 ?、@、' 等等。  
空白 :Wh 匹配所有类型的空格,包括印刷和表意文字的空格。  
Bidi :Bi 匹配诸如阿拉伯文和希伯来文这类从右向左书写的字符。  
朝鲜文 :Ha 匹配朝鲜文和组合朝鲜文字母。  
平假名 :Hi 匹配平假名字符。  
片假名 :Ka 匹配片假名字符。  
表意文字/汉字/日文汉字 :Id 匹配表意文字字符,如汉字和日文汉字  
"^n": 表示匹配前面表达式的 n 个搜索项。例如,[0-9]^4 匹配任意 4 位数字的序列。  
"@": 表示匹配前面表达式的 0 个或更多搜索项,匹配尽可能少的字符。  
"#": 匹配前面表达式的一个或更多搜索项,匹配尽可能少的字符。  
"[^...]": 匹配跟在 ^ 之后的不在字符集中的任何字符。  
"\$": 将匹配定位到行尾。  
"< ": 仅当词在文本中的此位置开始时才匹配。  
"> ": 仅当词在文本中的此位置结束时才匹配  
:a 匹配表达式 ([a-zA-Z0-9])。  
:c 匹配表达式 ([a-zA-Z])。  
:d 匹配表达式 ([0-9])。  
:h 匹配表达式 ([0-9a-fA-F])+。  
:i 匹配表达式 ([a-zA-Z\_\$\_][a-zA-Z0-9\_\$\_]\*)。  
:n 匹配表达式 ((([0-9]+.[0-9]\*)|([0-9]\*.[0-9]+)|([0-9]+)).  
:q 匹配表达式 ("^[^"]\*"|'[^']\*')  
:w 匹配表达式 ([a-zA-Z]+)  
:z 匹配表达式 ([0-9])+。  
\e Unicode U+001B。 转义符  
\g Unicode U+0007。 Bell  
\h Unicode U+0008。 退格符  
\n 匹配与平台无关的换行符。在"替换"表达式中,插入换行符。 换行符  
\t 匹配制表符,Unicode U+0009。 制表符

还有一个'|',表示"或"操作:  
"hi|hello":表示一个字符串里有"hi"或者"hello";

"(b|cd)ef": 表示"bef"或"cdef";

"(a|b)\*c": 表示一串"a""b"混合的字符串后面跟一个"c";

'.'可以替代任何字符;

"a.[0-9]": 表示一个字符串有一个"a"后面跟着一个任意字符和一个数字;

"^{3}\$": 表示有任意三个字符的字符串 (长度为3个字符);

方括号表示某些字符允许在一个字符串中的某一特定位置出现:

"[ab]": 表示一个字符串有一个"a"或"b" (相当于"a|b");

"[a-d]": 表示一个字符串包含小写的'a'到'd'中的一个 (相当于"a|b|c|d"或者"[abcd]");

"^[a-zA-Z]": 表示一个以字母开头的字符串;

"[0-9] %": 表示一个百分号前有一位的数字;

",[a-zA-Z0-9] \$": 表示一个字符串以一个逗号后面跟着一个字母或数字结束。

你也可以在方括号里用'^'表示不希望出现的字符, '^'应在方括号里的第一位。(如: "%^[a-zA-Z] %"表示两个百分号中不应该出现字母)。

为了逐字表达,你必须在"^.\$()|\*+?{\\"这些字符前加上转移字符'\'。

请注意在方括号中,不需要转义字符。