链滴

正则指引第一章学习笔记

作者：zhuangyan

原文链接：https://ld246.com/article/1371180155810

来源网站：链滴

第一章讲的是字符组（Character Class）。
在正则表达式中，它表示"在同一个位置可能出的各种字符"，其写法是在一对方括号[和]之间列出所有可能出现的字符，简单的字符组比如[ab]、[31]、[#.?]
需要注意的地方有
1）-范围表示法的要按ASCII顺序写， [0-9]是合法的，[9-0]会错
<pre>In [1]: import re
In [2]: re.search("^[0-9]$", "2") != None
Out[2]: True
<h2 id="In--3---re-search---9-0----2-----None">In [3]: re.search("^[9-0]$", "2") != None</h>
error   Traceback (most recent call last)
Traceback (most recent call last):
error: bad character range

 </pre>
2）匹配数字和字母不要用[0-z]，因为这个范围包括很多标点符号，最好写成[0-9a-zA-Z]
<pre>In [4]: re.search("^[0-z]$", "A") != None
Out[4]: True
In [5]: re.search("^[0-z]$", ":") != None
Out[5]: True </pre>
3）元字符做为普通字符使用时需要转义，在python代码中，普通字符串的转义需要加两个"\"，生字符串需要一个"\" 推荐使用原生字符串

<pre>#原生字符串和字符串的等价
In [6]: r"^[0\-9]$" == "^[0\\-9]$"
Out[6]: True
#原生字符串的转义要简单许多
In [7]: re.search(r"^[0\-9]$", "3") != None
Out[7]: False
In [8]: re.search(r"^[0\-9]$", "-") != None
Out[8]: True</pre>
4)]出现在不同位置，含义不同,正则表达式将]与前面最近的[匹配
<pre>#未转义的]
In [9]: re.search(r"^[012]345]$", "2345") != None
Out[9]: False
In [10]: re.search(r"^[012]345]$", "2345]") != None
Out[10]: True
In [11]: re.search(r"^[012]345]$", "5") != None
Out[11]: False
In [12]: re.search(r"^[012]345]$", "]") != None
Out[12]: False
#转义的]
In [13]: re.search(r"^[012]345]$", "2345") != None
Out[13]: False
In [14]: re.search(r"^[012]345]$", "5") != None
Out[14]: True
In [15]: re.search(r"^[012]345]$", "]") != None
Out[15]: True </pre>

5)字符组简记法
常见的字符组简记法有\d、\w、\s。从表面上看，它们与[...]完全没联系，其实是一致的。其中\等价于[0-9]，其中的d代表"数字（digit）"； \w等价于[0-9a-zA-Z_]，其中的w代表"单词字符（wor）"； \s等价于[\t\r\n\v\f]（第一个字符是空格），s表示"空白字符（space）"。
<pre>re.search(r"^\d$", "8") != None   # => True
re.search(r"^\d$", "a") != None   # => False
re.search(r"^\w$", "8") != None   # => True
re.search(r"^\w$", "a") != None   # => True

re.search(r"^\w$", "_") != None # => True
re.search(r"^\s$", " ") != None # => True
re.search(r"^\s$", "\t") != None # => True
re.search(r"^\s$", "\n") != None # => True

</pre>
6)相对于\d、\w和\s这三个普通字符组简记法，正则表达式也提供了对应排除型字符组的简记法\D、\W和\S--字母完全一样，只是改为大写。这些简记法匹配的字符互补：\s能匹配的字符，\S一定能匹配；\w能匹配的字符，\W一定不能匹配；\d能匹配的字符，\D一定不能匹配。例1-19示范了这个字符组简记法的应用。
<pre>#\d和\D
re.search(r"^\d$", "8") != None # => True
re.search(r"^\d$", "a") != None # => False
re.search(r"^\D$", "8") != None # => False
re.search(r"^\D$", "a") != None # => True
#\w和\W
re.search(r"^\w$", "c") != None # => True
re.search(r"^\w$", "!") != None # => False
re.search(r"^\W$", "c") != None # => False
re.search(r"^\W$", "!") != None # => True
#\s和\S
re.search(r"^\s$", "\t") != None # => True
re.search(r"^\s$", "0") != None # => False
re.search(r"^\S$", "\t") != None # => False
re.search(r"^\S$", "0") != None # => True </pre>