

# PYTHON

## 编程基础

# re模块使用-split、sub 和subn

# split



re模块中的split函数用于将字符串按与正则表达式匹配的子串分割。

re.split函数的语法格式如下：

```
re.split(pattern, string, maxsplit=0, flags=0)
```

其中，pattern是正则表达式；string是要分割的字符串；maxsplit是最大分割次数，默认为0表示不限制分割次数；flags与re.match等函数中的flags参数含义相同。

# split



例：re.split函数使用示例。

```
1 import re
2 str='sno:1810101,name:李晓明,age:19,major:计算机'
3 rlt=re.split(r'\W+', str)
4 print(rlt)
```

```
['sno', '1810101', 'name', '李晓明', 'age', '19', 'major', '计算机']
```

## sub



re模块的sub函数用于替换字符串中与正则表达式匹配的子串。re.sub函数的语法格式为：

```
re.sub(pattern, repl, string, count=0, flags=0)
```

其中，pattern是正则表达式；repl是要将匹配子串替换成的字符串；string是待做替换操作的字符串；count是最大替换次数，默认为0表示不限制替换次数（即将所有符合正则表达式的子串都替换成repl）；flags与re.match等函数中的flags参数含义相同。

## sub



### 例：re.sub函数使用示例。

```
1 import re
2 html="<h3 class='c-title'>
3 <a href='http://edu.people.com.cn/n1/2018/0905/c367001-30274290.html'>
    <em>南开大学</em>校长曹雪涛寄语新生</a>
4 </h3>"
5 content=re.sub(r'<[<]*>', "", html)
6 content=content.strip() #去除字符串content中两边的空白符
7 print('去除HTML标记后的内容为:', content)
```

## sub



例：re.sub函数使用示例。

去除HTML标记后的内容为：南开大学校长曹雪涛寄语新生

### 提示



第5行代码中，“<[<]\*>”表示匹配由一对尖括号括起来的字符串，通过“[<]”指定匹配的字符串中不能包含左尖括号“<”。

## subn



**re模块中的subn函数与re.sub函数功能完全相同，只是re.subn函数会以一个元组的形式同时返回替换匹配子串后得到的新字符串和替换的次数。re.subn函数的语法格式为：**

```
re.subn(pattern, repl, string, count=0, flags=0)
```

**各参数含义与re.sub函数相同。**



## subn



例：re.subn函数使用示例。

```
1 import re
2 html=""<h3 class="c-title">
3 <a href="http://edu.people.com.cn/n1/2018/0905/c367001-30274290.html">
    <em>南开大学</em>校长曹雪涛寄语新生</a>
4 </h3>""
5 content=re.subn(r'<[^<]*>', "", html)
6 print('去除HTML标记后的内容为: ', content)
```

## subn



例：re.subn函数使用示例。

去除HTML标记后的内容为: ('\\n 南开大学校长曹雪涛寄语新生\\n', 6)