

文本读取和字符分割

在Python中常常要处理文本文件，还要进行各种处理。这里选择了一个内容是人名的文本，用多个字符进行分割。

1.生成文本

In [1]:

```
%%writefile test.txt
```

```
张三,杜子腾-王二麻子  
, -
```

```
麻花藤-李四,王五
```

Writing test.txt

2.读取文本

In [2]:

```
with open("test.txt") as f:  
    t = f.read()  
print(t)
```

```
张三,杜子腾-王二麻子  
, -
```

```
麻花藤-李四,王五
```

去除换行标志。

In [3]:

```
t1=t.replace('\n','')  
t1
```

Out[3]:

```
'张三,杜子腾-王二麻子,-麻花藤-李四,王五'
```

3.处理分割字符

因为分割字符不是唯一的，这里采用了用正则表达式，需要导入re库。

使用多个分割字符，用“|”分开。

In [4]:

```
import re
t2=re.split(r"|-", t1)
t2
```

Out[4]:

```
['张三', '杜子腾', '王二麻子', '', '麻花藤', '李四', '王五']
```

使用通配符，W+表示多个字符

In [5]:

```
import re
t2=re.split(r"\W+", t1)
t2
```

Out[5]:

```
['张三', '杜子腾', '王二麻子', '麻花藤', '李四', '王五']
```

4.参考资料

1.<https://www.runoob.com/python/python-reg-expressions.html> (<https://www.runoob.com/python/python-reg-expressions.html>)

2.<https://www.runoob.com/python/file-methods.html> (<https://www.runoob.com/python/file-methods.html>)

In []: