

正则表达式

导入

正则表达式本身提供的功能是十分强大的，但是同时要做到熟练、精通使用正则表达式编程，处理各种复杂数据也并不是一件非常容易的事情。所以说正则表达式在各种复杂的数据处理中是非常重要的内容。

目录

- 1、概述
- 2、re 模块
- 3、贪婪模式与非贪婪模式

目标

- 1、区分正则表达式和 Python 正则表达式的相同点和不同点
- 2、使用 re 模块的相关方法处理数据
- 3、区分贪婪模式与非贪婪模式，使用非贪婪模式解决特殊的数据处理问题

一、概述

1、正则表达式

正则表达式，英文名为 **Regular Expression**，又称规则表达式。正则表达式通常被用来检索、替换那些符合某个模式(规则)的文本。

2、Python 正则表达式

Python 自 1.5 版本起增加了 `re` 模块，它提供 Perl 风格的正则表达式模式。`re` 模块使 Python 语言拥有全部的正则表达式功能。同时，`re` 模块是用 C 语言写的，其匹配速度非常快。

其中 `compile` 函数根据一个模式字符串和可选的标志参数生成一个正则表达式对象。该对象拥有一系列方法用于正则表达式匹配和替换。`re` 模块也提供了与这些方法功能完全一致的函数，这些函数使用一个模式字符串做为它们的第一个参数。

二、re 模块

1、re.match 方法

`re.match` 尝试从字符串的起始位置匹配一个规则，匹配成功就返回 `match` 对象，否则返回 `None`。可以使用 `group()` 获取匹配成功的字符串。

语法：`re.match(pattern, string, flags=0)`

参数说明：

参数	描述
<code>pattern</code>	匹配的正则表达式
<code>string</code>	要匹配的字符串。
<code>flags</code>	标志位，用于控制正则表达式的匹配方式，如：是否区分大小写，多行匹配等等。

示例 1（无标志位）：

```
>>> import re
>>> str = 'Python is the best language in the world'
>>> res = re.match('P', str)    # match 只能匹配以 xxx 开头的子字符串，第一个参数是正则，第二个参数是需要匹配的字符串
>>> res.group()    # 匹配成功使用 group 方法取出字符串
'P'
>>>
```

示例 2（有标志位）：

```
>>> import re
>>> str = 'Python is the best language in the world'
>>> res = re.match('P',str,re.I)    # re.I 表示忽略大小写
>>> res.group()    # 匹配成功使用 group 方法取出字符串
'P'
>>>
```

如果同时使用多个标志位使用|分割，比如 re.I | re.M

flags 可选标志位

修饰符	描述
re.I	使匹配对大小写不敏感
re.L	做本地化识别（locale-aware）匹配
re.M	多行匹配，影响 ^ 和 \$
re.S	使 . 匹配包括换行在内的所有字符
re.U	根据 Unicode 字符集解析字符。这个标志影响 \w, \W, \b, \B.
re.X	该标志通过给予你更灵活的格式以便你将正则表达式写得更易于理解。

我们可以使用 group(num) 或 groups() 匹配对象函数来获取匹配表达式。

匹配对象方法	描述
group(num=0)	匹配的整个表达式的字符串，group() 可以一次输入多个组号，在这种情况下它将返回一个包含那些组所对应值的元组。
groups()	返回一个包含所有小组字符串的元组，从 1 到所含的小组号。

```
>>> str = 'Python is the best language in the world'
>>> res = re.match('(.*) is (.*) .*', str)
>>> res
<_sre.SRE_Match object; span=(0, 40), match='Python is the best language
in the world'>
>>> res.group()
'Python is the best language in the world'
>>> res.group(1)
'Python'
>>> res.group(2)
'the'
>>>
```

2、常用的匹配规则

(1) 匹配字符

符号	匹配规则
.	匹配任意 1 个字符除了换行符\n
[abc]	匹配 abc 中的任意一个字符
\d	匹配一个数字，即 0-9
\D	匹配非数字，即不是数字
\s	匹配空白，即空格，tab 键
\S	匹配非空白，除空格，tab 键之类的
\w	匹配单词字符，即 a-z、A-Z、0-9、_
\W	匹配非单词字符

.(点): 匹配任意除换行符之外的字符

```
>>> import re
>>> str = 'aaaa'
>>> re.match('.', str).group()
'a'
>>> re.match('..', str).group()
'aa'
>>>
```

`[]`：匹配中括号内的任意一个字符

```
>>> str = 'hello'
>>> re.match('[he]', str).group()
'h'
>>> str = 'Hello'
>>> re.match('[He]', str).group()
'H'
>>> str = 'eee'
>>> re.match('[He]', str).group()
'e'
>>>
```

`\d`：匹配一个数字

```
>>> str = '124asb'
>>> re.match('\d', str).group()
'1'
>>> re.match('\d\d', str).group()
'12'
>>>
```

`\D`：匹配非数字，即不是数字

```
>>> str = 'a124asb'
>>> re.match('\D', str).group()
'a'
>>> str = '124asb' # 数字开头匹配错误
>>> re.match('\D', str).group()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'group'
```

`\s`：匹配空白字符串

```
>>> # \s 匹配空白字符
...
>>> str = '  hello'
>>>
>>> re.match('\s', str).group()
' '
>>>
```

`\S`：匹配非空白字符

```

>>> str = '    hello'
>>> re.match('\S',str).group()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'group'
>>>

>>> str = 'hello'
>>> re.match('\S',str).group()
'h'
>>>

```

\W: 匹配非单词字符

\w: 匹配单词字符，即 a-z、A-Z、0-9、_

```

>>> str = '    ssdf'
>>> re.match('\W',str).group()
' '

>>> str = 'ssdf'
>>> re.match('\w',str).group()
's'
>>>

```

(2) 匹配字符数量

符号	匹配规则
*	匹配前一个字符出现 0 次或者无限次，即可有可无
+	匹配前一个字符出现 1 次或者无限次，即至少有 1 次
\?	匹配前一个字符出现 1 次或者 0 次，即要么有 1 次，要么没有
{m}	匹配前一个字符出现 m 次
{m,}	匹配前一个字符至少出现 m 次
{n,m}	匹配前一个字符出现从 n 到 m 次

*****: 匹配前一个字符出现 0 次或者无限次，即可有可无

```

>>> import re
>>> ret = re.match('[A-Z][A-Z]*','Mn')

```

```

>>> ret.group()
'M'
>>> ret = re.match('[A-Z][a-z]*', 'Mn')
>>> ret.group()
'Mn'
>>> ret = re.match('[A-Z][a-z]*', 'Mnsdfsdfs')
>>> ret.group()
'Mnsdfsdfs'
>>>

```

+：使用实例，写一个正则匹配 **python** 变量名（变量名的语法规则：不能以数字开头，变量名只能包含字母，数字，下划线）

```

>>> ret = re.match('[a-zA-Z_]+[\w]*', 'name')
>>> ret
<_sre.SRE_Match object; span=(0, 4), match='name'>
>>> ret = re.match('[a-zA-Z_]+[\w]*', '_name')
>>> ret
<_sre.SRE_Match object; span=(0, 5), match='_name'>
>>> ret = re.match('[a-zA-Z_]+[\w]*', '33name')    # 数字开头不是变量名，
无法匹配
>>> ret
>>> ret.group()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'group'
>>>

```

?：使用实例，匹配 0-99 的数字

```

>>> re.match('[0-9]?[0-9]', '25').group()
'25'
>>> re.match('[0-9]?[0-9]', '09').group()
'09'
>>> re.match('[0-9]?[0-9]', '9').group()
'9'
>>>

```

{n,m}：匹配前一个字符出现从 **n** 到 **m** 次,使用实例，匹配 qq 邮箱

```

>>> re.match('[0-9]{6,11}@qq.com', '123654@qq.com').group()
>>> '123654@qq.com'

>>> re.match('[0-9]{6,11}@qq.com', '1236545999999@qq.com').group() # 大于 11 位匹配失败
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'group'

>>> re.match('[0-9]{6,11}@qq.com', '1236@qq.com').group() # 少于 6 位匹配失败
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'group'
>>>

```

(3) 原生字符串

python 字符串中\作为转义字符开头，比如\n 表示换行，\t 表示 tab 键，为了表示\本身，再加一个\，成为\\形式在 python 中表示路径 'G:\py 资料\1-上课资料\4-正则表达式课件\html'

```

>>> mypath = 'G:\py 资料\1-上课资料\4-正则表达式课件\html'
>>> print(mypath) # 路劲输出异常
G:\py 资料⊙-上课资料◆-正则表达式课件\html
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>> mypath = 'G:\\py 资料\\1-上课资料\\4-正则表达式课件\\html'
>>> print(mypath) # 将\转义之后正常
G:\py 资料\1-上课资料\4-正则表达式课件\html
>>>
>>>
>>>

```

下面使用正则表达式来匹配路径


```
>>> re.match('c:\\a.txt', 'c:\\a.txt').group()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'group'
>>>
>>>
>>> re.match('c:\\\\a.txt', 'c:\\a.txt').group() # 正则中\也表示转义,
所以匹配字符串中的\需要再次转义
'c:\\a.txt'
>>>
>>>
>>>
```

正则表达式里使用"\"作为转义字符，这就可能造成反斜杠困扰。假如你需要匹配文本中的字符"\"，那么使用编程语言表示的正则表达式里将需要 4 个反斜杠"\\\"。

在 python 中有原生字符串，在字符串前面加上 r 表示字符串中的\不转义。

```
>>> re.match('c:\\a.txt', 'c:\\a.txt').group()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'group'
>>>
>>>
>>> re.match(r'c:\\a.txt', 'c:\\a.txt').group() # 在正则前加 r, 表示原生字符串, python 字符串不转义。
'c:\\a.txt'
>>>
```

(4) 匹配开头结尾

符号	匹配规则
^	匹配字符串开头
\$	匹配字符串结尾

示例：匹配 163.com 的邮箱地址

```
>>> import re
>>>
>>> # 正确地址
...
>>> ret = re.match('[\w]{4,20}@163\.com', 'python@163.com')
>>> ret.group()
'python@163.com'
>>>
>>> # 不正确地址
...
>>> ret = re.match('[\w]{4,20}@163\.com', 'python@163.comsdsdds')
>>> ret.group()
'python@163.com'
>>> # 通过$来限定结尾
...
>>> ret = re.match('[\w]{4,20}@163\.com$', 'python@163.comsdsdds')
>>> ret.group()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'NoneType' object has no attribute 'group'
>>>
```

(5) 分组匹配

符号	匹配规则
	匹配左右任意一个表达式
(ab)	将括号中字符作为一个分组
\num	引用分组 num 匹配到的字符串
(?P)	分组起别名
(?P=name)	引用别名为 name 分组匹配到的字符串

|: 匹配出 0-100 之间的数字

```
# 匹配出两位数字
ret= re.match(' [1-9]? \d', '8')
ret.group()
>>> ret= re.match(' [1-9]? \d$ | 100', '78')
>>>
>>> ret.group()
'78'
>>> ret= re.match(' [1-9]? \d$ | 100', '100')
>>> ret.group()
'100'
```

(ab) 分组：分别匹配电号码区号和电话号码

```
>>> re.match(' ([^-]*)-(\d*)', '0756-12345678').group()
'0756-12345678'
>>> re.match(' ([^-]*)-(\d*)', '0756-12345678').group(0)
'0756-12345678'
>>> re.match(' ([^-]*)-(\d*)', '0756-12345678').group(1)
'0756'
>>> re.match(' ([^-]*)-(\d*)', '0756-12345678').group(2)
'12345678'
>>>
```

\: 匹配 html 中的文本

```
>>> re.match(' <[a-zA-Z]*>(.*)</[a-zA-Z]*>', '<div>hello world</div>').group()
'<div>hello world</div>'
>>>
>>> re.match(' <[a-zA-Z]*>(.*)</[a-zA-Z]*>', '<div>hello world</div>').group(1)
'hello world'
>>>
```

其实我们发现 html 的标签都是成对出现的，这样匹配规则 html 就可以引用组来实现，避免重复写。

```
>>> re.match(r'<([a-zA-Z]*)>.*</\1>', '<div>hello world</div>').group()
'<div>hello world</div>' # \1 表示引用前面第一组
```

```
>>>  
>>>
```

在使用组引用的时候如果有多个组就容易造成混乱，比如下面多个组引用，这样很混乱也不容易阅读

```
>>> ret =  
re.match(r'<(\w*)><(\w*)>.*</\2></\1>', '<div><h1>www.baidu.com</h1></div>')  
>>> ret.group()  
'<div><h1>www.baidu.com</h1></div>'  
>>>
```

解决上面这个问题得给分组起别名

(?P)分组起别名

(?P=name)引用别名为 name 分组匹配到的字符串

```
>>> ret =  
re.match(r'<(P<name1>\w*)><(P<name2>\w*)>.*</(P=name2)></(P=name1)>',  
'<div><h1>www.baidu.com</h1></div>')  
>>>  
>>> ret.group()  
'<div><h1>www.baidu.com</h1></div>'  
>>>
```

(?P)和(?P=name)中的字母 p 大写

3、re.compile 方法

compile 将正则表达式模式编译成一个正则表达式对象

reg = re.compile(pattern) result = reg.match(string)

等效于 result = re.match(pattern, string)

但使用 re.compile()和保存所产生的正则表达式对象重用效率更高时该表达式会在单个程序中多次使用。

4、re.search 方法

search 在全文中匹配一次，匹配到就返回

语法：re.search(pattern, string, flags=0)

函数参数说明：

参数	描述
pattern	匹配的正则表达式
string	要匹配的字符串。
flags	标志位，用于控制正则表达式的匹配方式，如：是否区分大小写，多行匹配等等。

示例：扫描字符串，寻找的第一个由该正则表达式模式产生匹配的位置，并返回相应的 MatchObject 实例。

```
>>> import re
>>>
>>> re.search('bc','abcd')    # 在整个字符串中匹配，匹配到第一个就返回
<_sre.SRE_Match object; span=(1, 3), match='bc'>
>>> re.search('bc','abcd').group()
'bc'
>>>
```

5、re.findall 方法

findall 匹配所有返回一个列表，这个方法使用频率较高。

语法：findall(string[, pos[, endpos]])

参数说明：

参数	描述
string	待匹配的字符串。
pos	可选参数，指定字符串的起始位置，默认为 0。

endpos	可选参数，指定字符串的结束位置，默认为字符串的长度。
--------	----------------------------

示例：使用 `findall` 来匹配出段子网中的所有段子



通过观察网页源代码，可以看出段子都包含在`(.*?)`中

```
# coding=utf-8
import re

html = """<ul><li class="list_li"><div class="content">如果你得罪了老板，失去的只是一份工作；如果你得罪了客户，失去的不过是一份订单；是的，世上只有一个人可以得罪：你给她脸色看，你冲她发牢骚，你大声顶撞她，甚至当着她的面摔碗，她都不会记恨你，原因很简单，因为她是你的母亲。</div><div class="info"><span class="right tags"><label class="tag"></label><a href="/member/8">短语</a>,<a href="http://ishuo.cn/special/laoban">老板</a>,<a href="http://ishuo.cn/special/muqin">母亲</a></span><a href="/subject/5899">因为她是你的母亲</a>04 月 29 日 0 个评论 0 人喜欢</div></li><li class="list_li"><div class="content">有位非常漂亮的女同事，有天起晚了没有时间化妆便急忙冲到公司。结果那天她被记旷工了……吃惊</div><div class="info"><span class="right tags"><label class="tag"></label><a href="/member/4">笑话</a>,<a href="http://ishuo.cn/special/nahan">内涵</a>,<a href="http://ishuo.cn/special/tongshi">同事</a></span><a href="/subject/5898">这个笑话好内涵，你懂了没？</a>04 月 29 日 0 个评论 0 人喜欢</div></li><li class="list_li"><div class="content">悟空和唐僧一起上某卫视非诚勿扰，悟空上台,24 盏灯全灭。理由:1. 没房没车只有一根破棍。2. 保镖职业危险。3. 动不动打妖精,对女生不温柔。4. 坐过牢,曾被压五指山下 500 年。</div><div class="info"><span class="right tags"><label class="tag"></label><a href="/member/8">短语</a>,<a href="http://ishuo.cn/special/laoban">老板</a>,<a href="http://ishuo.cn/special/muqin">母亲</a></span><a href="/subject/5899">因为她是你的母亲</a>04 月 29 日 0 个评论 0 人喜欢</div></li><li class="list_li"><div class="content">有位非常漂亮的女同事，有天起晚了没有时间化妆便急忙冲到公司。结果那天她被记旷工了……吃惊</div><div class="info"><span class="right tags"><label class="tag"></label><a href="/member/4">笑话</a>,<a href="http://ishuo.cn/special/nahan">内涵</a>,<a href="http://ishuo.cn/special/tongshi">同事</a></span><a href="/subject/5898">这个笑话好内涵，你懂了没？</a>04 月 29 日 0 个评论 0 人喜欢</div></li><li class="list_li"><div class="content">悟空和唐僧一起上某卫视非诚勿扰，悟空上台,24 盏灯全灭。理由:1. 没房没车只有一根破棍。2. 保镖职业危险。3. 动不动打妖精,对女生不温柔。4. 坐过牢,曾被压五指山下 500 年。</div><div class="info"><span class="right tags"><label class="tag"></label><a href="/member/8">短语</a>,<a href="http://ishuo.cn/special/laoban">老板</a>,<a href="http://ishuo.cn/special/muqin">母亲</a></span><a href="/subject/5899">因为她是你的母亲</a>04 月 29 日 0 个评论 0 人喜欢</div></li></ul>"""
```

唐僧上台，哗！灯全亮。理由：1. 公务员； 2. 皇上兄弟，后台最硬 3. 精通梵文等外语 4. 长得帅 5. 最关键一点：有宝马！

笑话

[悟空](/member/4), [唐僧](http://ishuo.cn/special/wukong), [非诚勿扰](http://ishuo.cn/special/feichengwurao)

[悟空和唐僧一起上某卫视非诚勿扰](/subject/5897)

04月29日0个评论 0人喜欢

Save your heart for someone who cares. 为了某个在乎你的人，请节约你的真心！

短语

[真心](/member/8)

[为了某个在乎你的人，请节约你的真心](/subject/5896)

04月29日0个评论 0人喜欢

最新版 i pod ！潮爆啦！！

笑话

[限量版](/member/4), [ipod](http://ishuo.cn/special/ipod)

[最新版 i pod](/subject/5895)

04月29日0个评论 0人喜欢

有一个人叫真咯嗦，娶了个老婆叫要你管，生了个儿子叫麻烦。有一天麻烦不见了！夫妻俩就去报案。警察问爸爸：请问这位男士你叫啥名字？爸爸说：真咯嗦。警察很生气，然后他又问妈妈叫啥名字。妈妈说：要你管。警察非常生气的说：你们要干什么？夫妻俩说：找麻烦。

警察

[笑话](/member/4), [警察](http://ishuo.cn/special/jingcha), [麻烦](http://ishuo.cn/special/mafan)

[你们要干什么？](/subject/5894)

04月29日0个评论 0人喜欢

一年奔波，尘缘遇了谁；一句珍重，天涯别了谁；一点灵犀，凭栏忆了谁；一种相思，闲愁予了谁；一江明月，豪情酬了谁；一场冬雪，烟波忘了谁；一壶浊酒，相逢醉了谁；一世浮生，轻狂撩了谁；一封短信，才情念了谁；一番思量，谁是谁的谁；一枚围脖，转发回复谁.....

段子

[尘缘](/member/2), [相逢](http://ishuo.cn/special/chenyuan)

[一年奔波，尘缘遇了谁](/subject/5893)

04月29日0个评论 0人喜欢

人与人间的信任，就像是纸片，一旦破损，就不会再回到原来的样子。

短语

[信任](/member/8)

[人与人间的信任，就像是纸片](/subject/5892)04月29日0条评论0人喜欢

爱是一种感觉，跟时间长短没关系。——《不敢说爱你》

短语

感觉

[爱是一种感觉，跟时间长短没关系](/subject/5891)04月29日0条评论0人喜欢

师太和贫僧牵手了！！（请用四个字表达此刻感受！）

笑话

师太

贫僧

[师太和贫僧牵手了](/subject/5890)04月29日0条评论0人喜欢

这孩子长大了就是一祸害，你同意吗？

笑话

孩子

[这孩子长大了就是一祸害，你同意吗？](/subject/5889)04月29日0条评论0人喜欢

有位语文老师为学生朗诵了一首陆游的《卧春》，要求学生听写出来……

笑话

听写

[糗事百科之听写](/subject/5888)04月29日0条评论0人喜欢

1. 暧昧王:天蝎; 2. 犹豫不决:处女; 3. 朝三暮四:双子; 4. 逆向思维:水瓶; 5. 真爱无罪:双鱼; 6. 以史为鉴:摩羯; 7. 权衡利弊:金牛; 8. 宁缺毋滥:天秤; 9. 社会道德:巨蟹; 10. 惜时如金:白羊; 11. 无暇顾及:射手; 12. 目标明确:狮子。你是这样的吗？

短语

星座

恋爱

[十二星座的恋爱特点](/subject/5887)04月29日0条评论0人喜欢

白羊：记不清人；金牛：悔未上床；双子：他（她）是青蛙（恐龙）；巨蟹：善哉善哉；狮子：一脚蹬开；处女：倒大霉了；天秤：感觉很好；天蝎：你死我活；射手：有多少次；摩羯：气死我也；水瓶：奶奶个熊；双鱼：还是想他（她）{点开大图更精彩} @乌托邦办公室

笑话

星座

初恋

[十二星座坐上时光机回忆初恋](/subject/5886)04月29日0条评论0人喜欢

★做体操的猫咪

你伤不起! ~ (@精品搞笑) </div><div class="info"><label class="tag"></label>笑话, 体操做体操的猫咪你伤不起04月29日0条评论 0人喜欢</div><li class="list_li"><div class="content">哥,你是怎么用眼神杀死敌人的?? (@瞬间乐翻你 投稿)</div><div class="info"><label class="tag"></label>笑话, 眼神哥,你是怎么用眼神杀死敌人的?? 04月29日0条评论 0人喜欢</div><li class="list_li"><div class="content">这歌名起的一个比一个狠! </div><div class="info"><label class="tag"></label>笑话, 歌名这歌名起的一个比一个狠04月29日0条评论 0人喜欢</div><li class="list_li"><div class="content">今天同朋友来到一个地方,是新会、开平、台山三个地方的交界处(经国务院批准的)。因此,实现了2分钟内去新会上个厕所,过开平洗个手,再番来台山吃饭的历史创举。(@ZCHuang) </div><div class="info"><label class="tag"></label>笑话, 平生第一次可以同时出现在三个地方04月29日0条评论 0人喜欢</div><li class="list_li"><div class="content">如果给你一次去西天取经的机会,选3个做你的徒弟,你会选谁? 1.多啦A梦。2.柯南。3.蜡笔小新。4.樱木花道。5.悟空。5.大力水手。6.超人。7.功夫熊猫。8.美少女战士。9.路飞。10.漩涡鸣人。11.奥特曼。12.卡卡西。13.喜羊羊。14.蜘蛛侠。15.百变小子。16.忍者神龟。17.变形金刚。18.八神庵。19.杀生丸。20.阿童木</div><div class="info"><label class="tag"></label>段子, 西天取经, 徒弟如果给你一次去西天取经的机会04月29日0条评论 0人喜欢</div><li class="list_li"><div class="content">记者:说真的,你真的会给给孩子换尿布吗? 姚明:要不你躺下,我给你换一个! 实话告诉你,我用一只脚就能给孩子换尿布,喂奶啥的,都行。 记者:我不信! 姚明:真的,连灯都不用开。 记者:不可能! 你说怎么换? 姚明:一只脚把媳妇揣醒就得。</div><div class="info"><label class="tag"></label>笑话, 姚明姚明的幽默04月29日0条评论 0人喜欢</div><li class="list_li"><div class="content">失恋算个啥? 轻轻的,你走吧,千万别后悔,因为只要你一挥手,就会发现,已经有那等不及的意中人,正偷偷摸摸拉你的手! </div><div class="info"><span class="right

tags"><label class="tag"></label>短语, 失恋失恋算个啥? 04月29日0个评论 0人喜欢</div><li class="list_li"><div class="content">一生只谈三次恋爱最好, 一次懵懂, 一次刻骨, 一次一生。谈的太多会比较, 无法确定; 经历太多会麻木, 不再相信爱情, 行尸走肉, 最后与不爱的人结婚, 无法发自内心的爱对方, 日常表现的应付, 对方则抱怨你不够关心和不顾家, 最后这失败的爱情, 让你在遗憾和凑合中走完一生。</div><div class="info"><label class="tag"></label>段子, 恋爱, 爱情一生只谈三次恋爱最好04月29日0个评论 0人喜欢</div><li class="list_li"><div class="content">1 吃饭的时候咬到自己的舌头。2 在车上睡着了头敲到玻璃。3 玩夹子, 夹到手。4 坐在椅子上摇椅子, 把自己摔倒了。5 左脚拌右脚结果摔倒。6 碰到桌子或柜子的角, 巨疼。7 从身上硬撕下粘贴。8 拉衣服拉链夹到下巴上的肉。9 踩楼梯突然踩空摔倒。都经历过, 那无敌了! </div><div class="info"><label class="tag"></label>短语, 最痛肉体上最痛的, 你经历过几个04月29日0个评论 0人喜欢</div><li class="list_li"><div class="content">1 友情:《牛仔裤的夏天》 2 初恋:《情书》 3 孤独:《天使爱美丽》 4 坚强:《隐形的翅膀》 5 成长:《千与千寻》 6 自我:《穿普拉达的女王》 7 尊严:《成为简奥斯汀》 8 母爱:《黑暗中的舞者》 9 自由:《蓝》 10 智慧:《律政俏佳人》</div><div class="info"><label class="tag"></label>知识, 电影, 女人10 部值得一看的电影, 让女人受益一生】04月29日0个评论 0人喜欢</div><li class="list_li"><div class="content">有一种借口叫年轻, 可以不珍惜时光, 不珍惜爱, 不珍惜一切来之不易的东西。有一种感情叫错过, 错过爱, 错过可以相守的人, 错过一段刻骨铭心的情。有一种寂寞叫想念, 想念一个人, 一段往事, 一场相遇。寂静的夜里, 深深切切的想念, 于是深深切切地寂寞……</div><div class="info"><label class="tag"></label>段子, 借口, 珍惜, 寂寞有一种借口叫年轻04月29日0个评论 0人喜欢</div><li class="list_li"><div class="content">我们常常看到的风景是: 一个人总是仰望和羡慕着别人的幸福, 一回头, 却发现自己正被仰望和羡慕着。其实, 每个人都是幸福的。只是, 你的幸福, 常常在别人眼里 。[阳光] 早安

</div><div class="info"><label class="tag"></label>短语, 幸福你的幸福，常常在别人眼里04月29日0条评论0人喜欢</div><li class="list_li"><div class="content">你病，或者不病倒，老板就在那里，不悲不喜；你休，或者不休假，工作就在那里，不来不去；你拼，或者不拼命，工资就在那里，不增不减；你辞，或者不辞职，地球还是会转，不歇不停；让我中500万，或者让我傍个大款，扯淡，蛋疼，淡定，悲催，关注，评论，转发!!!</div><div class="info"><label class="tag"></label>段子, 老板, 工作你拼，或者不拼命，工资就在那里，不增不减04月29日0条评论0人喜欢</div><li class="list_li"><div class="content">1、毕业后才知道校园恋爱是最纯洁的；2、毕业后才知道学习是最重要的；3、毕业后才知道校园生活是最幸福的；4、毕业后才知道宿舍生活是最好的；5、毕业后才知道食堂的饭菜是最便宜的；6、毕业后才知道上学是最美妙的事。7、毕业后才知道学生花钱最大手大脚.....</div><div class="info"><label class="tag"></label>短语, 毕业, 恋爱毕业后才知道这七宗“最”04月29日0条评论0人喜欢</div><li class="list_li"><div class="content">“我爱你”的含义是：无论贫穷，富贵，生老，病死，天灾，人祸我都不离弃的爱你。当你说出这三个字的时候，你是否有足够的勇气和顽强的毅力去承受他的人生。爱，不要轻易说出口。</div><div class="info"><label class="tag"></label>短语, 我爱你爱，不要轻易说出口04月29日0条评论0人喜欢</div><li class="list_li"><div class="content">好好去爱，去生活。青春如此短暂，不要叹老。偶尔可以停下来休息，但是别蹲下来张望。走了一条路的时候，记得别回头看。时不时问问自己，自己在干嘛?记住，每天的太阳都是新的，不要辜负了美好的晨光。</div><div class="info"><label class="tag"></label>段子, 生活每天的太阳都是新的，不要辜负了美好的晨光04月29日0条评论0人喜欢</div><li class="list_li"><div class="content">People cry, not because they\'re weak. It\'s because they\'ve been strong for too long. 哭泣，不代表脆弱，只因坚强了太久。</div><div class="info"><label class="tag"></label>短语, <a

[坚强](http://ishuo.cn/special/jianqiang), [哭泣](http://ishuo.cn/special/kuqi)

[哭泣](/subject/5869), 不代表脆弱, 只因坚强了太久

04月29日 0个评论 0人喜欢

他做到了, 他悬空了! 其实你也可以做到, 站在一滩水的右上角就 ok!

[笑话](/member/4), [悬空](http://ishuo.cn/special/xuankong)

[他做到了, 他悬空了!](/subject/5868)

04月29日 0个评论 0人喜欢

一年前, 学校广场上献血. 200CC 送一副修指甲的用具, 400CC 送个手表. 邻班一 MM 听了感觉很幸福, 跑过去问护士: "1000CC 送什么?" 护士淡定的说: "送个棺材……"

[笑话](/member/4), [糗事百科](http://ishuo.cn/special/qiushibaike), [献血](http://ishuo.cn/special/xianxue)

[糗事百科之献血](/subject/5867)

04月29日 0个评论 0人喜欢

1 白羊——孙悟空; 2 金牛——东海龙王; 3 双子——红孩儿; 4 巨蟹——观音娘娘; 5 狮子——玉皇大帝; 6 处女——唐僧; 7 天秤——太白金星; 8 天蝎——白骨精; 9 射手——六耳猕猴; 10 魔羯——铁扇公主; 11 水瓶——牛魔王; 12 双鱼——猪八戒。

[短语](/member/8), [星座](http://ishuo.cn/special/xingzuo), [西游记](http://ishuo.cn/special/xiyouji)

[12 星座所代表的西游记人物](/subject/5866)

04月29日 0个评论 0人喜欢

镜子前拉一下领带; 吹口哨下楼; 决然地摁灭香烟; 突然把车开过来, 摇下车窗笑; 弯腰轻抚可爱小孩的头; 从背后变出一朵玫瑰; 做俯卧撑, 满脸是汗珠; 台上讲话眼睛扫视全场; 思考时额头露出川字皱纹; 果决大步流星地走; 正在换灯泡或钉钉子; 买单时毫不迟疑拿出钱包。

[短语](/member/8), [女人](http://ishuo.cn/special/nvren), [男人](http://ishuo.cn/special/nanren)

[女人眼里男人最 man 十二个瞬间](/subject/5865)

04月29日 0个评论 0人喜欢

★这自行车酷吧! 喜欢吗? (@趣图语录精选)

[笑话](/member/4), [自行车](http://ishuo.cn/special/zixingche)

[这自行车酷吧! 喜欢吗?](/subject/5864)

04月29日 0个评论 88人喜欢

妈妈语重心长的对女儿说, “从小你就不聪明, 累死累活的才考上个大学, 毕业后还找不到工作, 现在司

机要男的、编辑要男的、会计要男的、连秘书也指定要男的，妈实在为你操碎了心啊。”女儿，“555555……。”妈妈一抹脸，坚定的说，“所以趁现在老婆还能是女人，赶紧上岗，要不然过两年……”

段子,[糗事百科](http://ishuo.cn/special/qiushibaike),[老婆](http://ishuo.cn/special/laopo)

[糗事百科之找工作真难](/subject/5863)04月29日0条评论 38人喜欢

真 de 太可爱了……愿世界大同……

笑话,[可爱](http://ishuo.cn/special/keai),[大同](http://ishuo.cn/special/datong)

[愿世界大同](/subject/5862)04月28日0条评论 235人喜欢

有时候，面对着身边的人，突然觉得说不出话。有时候，曾经一直坚持的东西一夜间面目全非。有时候，想放纵自己，希望自己痛痛快快歇斯底里地发一次疯。有时候，别人突然对你说，我觉得你变了，然后自己开始百感交集。有时候，觉得自己拥有着整个世界，一瞬间却又觉得自己其实一无所有。

放纵,[一无所有](http://ishuo.cn/special/yiwusuoyou)

[有时候，觉得自己拥有着整个世界](/subject/5861)04月28日0条评论 179人喜欢

男人就应该象自己的小弟弟， 第一：从不外露炫耀； 第二： 关键是时刻硬的起撑的住； 第三： 能培育出接班人； 第四： 善于攻击而又使其感到愉悦； 第五： 既能制造摩擦又使大家同感快乐； 第六： 胜利后能谦恭的缩小自己。 总结：低调、有骨气、有能力。。。 (via@笑多了会怀孕)

男人,[小弟弟](http://ishuo.cn/special/xiaodidi)

[男人就应该象自己的小弟弟](/subject/5860)04月28日0条评论 194人喜欢

世界上最有钱的人是奥特曼，因为所有取款机上都印着他名字的缩写“ATM”。

奥特曼,[ATM](http://ishuo.cn/special/ATM)

[世界上最有钱的人是奥特曼](/subject/5859)04月28日1条评论 213人喜欢

语文好万能增长你的文学知识！英语能让你与鬼佬交流！历史能让你不背叛啊！地理能让你不至于迷路啊！政治能让你知道怎样维权啊！可是数学除了毁掉整个人生还能做什么啊！泥马！！你用函数买菜啊！你去黄鹤楼还去算长江里的船距离你多远啊！你看到一

排电话号码要想它们之间有没有通项公式啊！@侯蒙恩

[笑话](/member/4), [语文](http://ishuo.cn/special/yuwen), [英语](http://ishuo.cn/special/yingyu), [数学](http://ishuo.cn/special/shuxue)[数学除了毁掉整个人生还能做什么啊](/subject/5858)04月28日0条评论164人喜欢

[月经](/subject/5857)又被成为“好朋友”你知道原因吗？04月28日0条评论194人喜欢

[这个世界上，总有个人治得了你](/subject/5856)04月28日0条评论147人喜欢

[短语](/member/8), [爱情](http://ishuo.cn/special/aiqing)[所谓爱情也不过是](/subject/5855)04月28日0条评论238人喜欢

[朋友就像人民币，有真、也有假](/subject/5854)04月28日0条评论239人喜欢

[这个指挥太有激情了](/subject/5853)04月28日0条评论175人喜欢

[今天第一天上班。。有点小紧张呢~~我帅气吗？！~](/subject/5853)

[笑话](/member/4), [指挥](http://ishuo.cn/special/zhahui)

```

href="http://ishuo.cn/special/shuaiqi">帅气</a></span><a
href="/subject/5852">我帅气吗?! </a>04月28日0条评论 118人喜欢
</div></li><li class="list_li"><div class="content">一个小男孩拿着一张假
钱走进了玩具店，准备买一架玩具飞机。 服务员阿姨说：“小朋友，你的钱不是
真的。 ” 小男孩反问道：“阿姨，难道你的飞机是真的？” </div><div
class="info"><span class="right tags"><label class="tag"></label><a
href="/member/4">笑话</a>,<a href="http://ishuo.cn/special/wanju">玩具
</a></span><a href="/subject/5851">糗事百科之买玩具</a>04月28日0个评
论 35人喜欢</div></li><li class="list_li"><div class="content">据说失眠
的同学盯着看十分钟就能睡着了。</div><div class="info"><span class="right
tags"><label class="tag"></label><a href="/member/4">笑话</a>,<a
href="http://ishuo.cn/special/shimian">失眠</a></span><a
href="/subject/5850">据说失眠的同学盯着看十分钟就能睡着了</a>04月28日0
条评论 138人喜欢</div></li></ul>"""

# 将正则编译
reg = re.compile('<div class="content">(.*?)</div>')

# findall 匹配字符串中所有的符合正则的

# result_list = reg.findall(html)      # 两种用法都可以
result_list = re.findall(reg,html)
index = 0
print(result_list)
for i in result_list:
    # 循环将段子取出
    print('第 %s 条段子: %s' % (index, i))
    index += 1

```

6、re.sub 方法

sub 将匹配到的数据进行替换

语法：sub(pattern, repl, string, count=0, flags=0)

参数说明：

参数	描述
pattern	正则中的模式字符串。

repl	替换的字符串，也可为一个函数。
string	要被查找替换的原始字符串。
count	模式匹配后替换的最大次数，默认 0 表示替换所有的匹配。
flags	标志位，用于控制正则表达式的匹配方式

示例：

```
>>> # sub 将匹配中的字符串替换
...
>>> ret = re.sub('h','H','hello world')
>>>
>>> ret      # 原来 h 被替换成了 H
'Hello world'
>>>
```

7、re.split 方法

split 根据匹配进行切割字符串，并返回一个列表

语法：split(pattern, string, maxsplit=0, flags=0)

参数说明：

参数	描述
pattern	匹配的正则表达式
string	要匹配的字符串。
maxsplit	分隔次数，maxsplit=1 分隔一次，默认为 0，不限制次数。
flags	标志位，用于控制正则表达式的匹配方式

示例：

```
# 以，切割字符串，并且最多 2 次
>>> ret = re.split(',', 'hello,world,hello,world', 2)
>>> ret
['hello', 'world', 'hello,world']
>>>
```


三、贪婪模式与非贪婪模式

1、概述

Python 里数量词默认是贪婪的,总是尝试匹配尽可能多的字符,非贪婪则相反,总是尝试匹配尽可能少的字符。

在 "*", "?", "+", "{m,n}" 后面加上 ? , 使贪婪变成非贪婪.

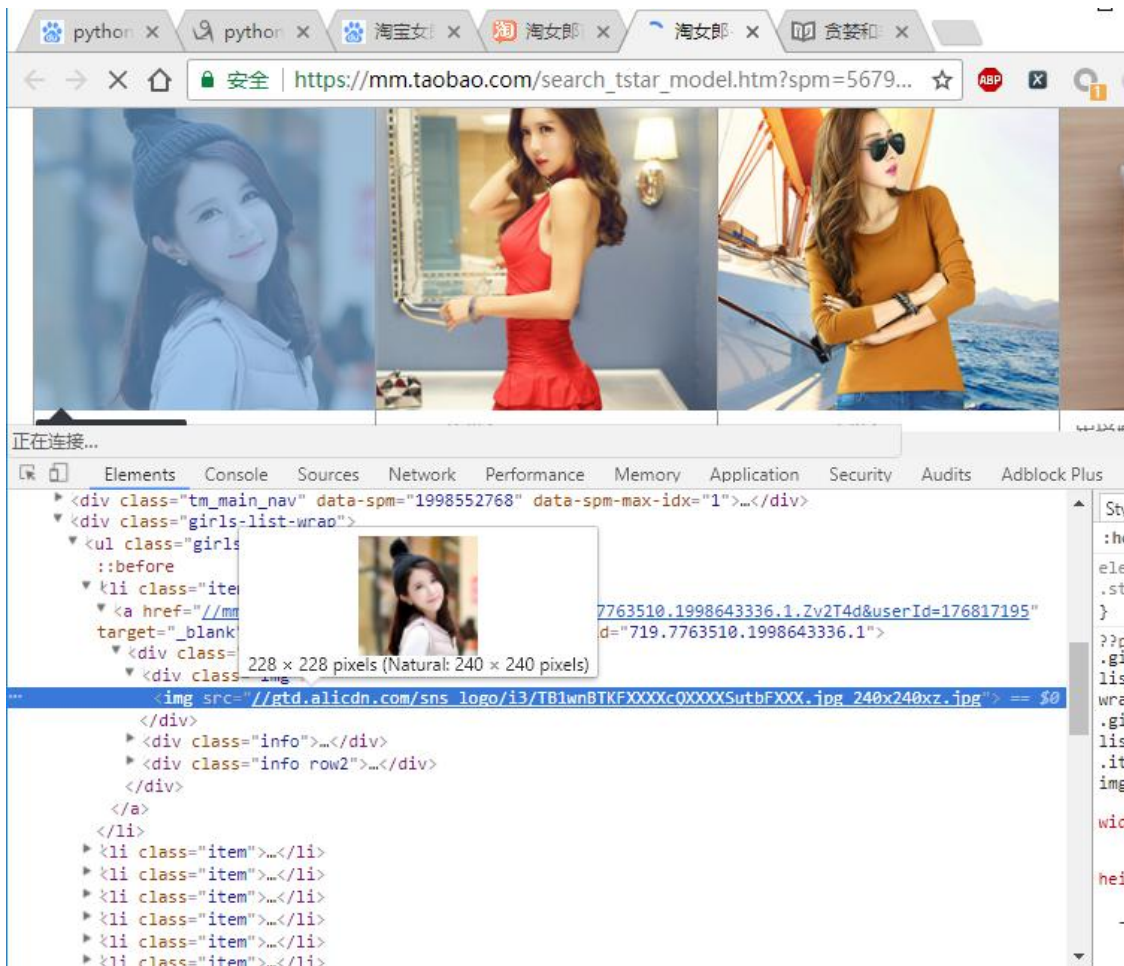
示例:

```
>>> # 贪婪模式匹配
...
>>> re.match('.*\d', 'a22222').group()
'a22222'
>>>
>>> # 加? 号非贪婪模式
...
>>> re.match('.*?\d', 'a11112').group()
'a1'
>>>
```

2、案例

匹配淘女郎模特首页所有图片 url 链接 网站

url:<https://mm.taobao.com/searchtstarmodel.htm?spm=5679.126488.640745.2.24bacb080L6tCt>



图片 url

正则表达式应该写成 '', html)
1400 for i in res:
1401 print(i)
1402
```

贪婪模式，url之外的内容也匹配到了

```
C:\Users\huangguifeng\AppData\Local\Programs\Python\Python35\python3.exe C:/Users/huangguifeng/Des
//img.alicdn.com/tps/i4/TB1IOxzIFXXXXXQaXXXWp6r0pXX-425-380.png_450x10000q90.jpg" alt="凌智贤"
//img.alicdn.com/tps/i1/TB1AJVMIFXXXXaKXFXWp6r0pXX-425-380.png_450x10000q90.jpg" alt="董俐DONG"
//img.alicdn.com/tps/i2/TB1TPhSIFXXXXbnXpXXWp6r0pXX-425-380.png_450x10000q90.jpg" alt="陈洁kiki"
//img.alicdn.com/tps/i3/TB12BRxIFXXXXadaXXXWp6r0pXX-425-380.png_450x10000q90.jpg" alt="赵方方F"
//img.alicdn.com/tps/i1/TB1RMh1IFXXXXa7XXXXWp6r0pXX-425-380.png_450x10000q90.jpg" alt="斯方圆"
//img.alicdn.com/tps/i2/TB10ajnIFXXXXaKXXXXWp6r0pXX-425-380.png_450x10000q90.jpg" alt="杨杨"
```

## 使用非贪婪模式

```
main.py x
33
34
35 html = '''...'''
非贪婪模式
1399 res = re.findall('三生三世，十里桃花九州海上牧云记莫斯科行动' 请使用正则将标签中的全部内容匹配出来,请写 Python 代码完成匹配。