

2021/6/11--正则错误

笔记本: reptile_draft

创建时间: 2021/6/11 11:14

更新时间: 2021/7/13 18:14

作者: 134exetj717

URL: about:blank

```
Traceback (most recent call last):
  File "C:\my_practise\python_practise\reptile_basic.py", line 18, in <module>
    f.write(html)
UnicodeEncodeError: 'gbk' codec can't encode character '\u200b' in position 15035: illegal multibyte sequence

Traceback (most recent call last):
  File "C:\my_practise\python_practise\reptile_basic.py", line 17, in <module>
    f.write(html)
TypeError: write() argument must be str, not bytes
```

错误原因:

字符不匹配, 文件录入不能直接是html, 必须要先转为str, 然后一个一个录入

```
#找到图片资源
findPic = re.compile('sa_href="(.*?)" target="_blank">\n\n<span class="tit">(.*?)</span>\n</a>')
```

在使用正则表达式的过程中, 小括号代表我们需要提取出来的元素~!!!

在使用正则表达式搜索时, 当出现 [(1,2,3)] 的过程中, 我们不能用 [0]来表示里面的(1,2,3), 未知
两种打印都会出错???

```
Traceback (most recent call last):
  File "C:\my_practise\python_practise\reptile_basic.py", line 55, in <module>
    urls = analyseWeb(baseHtml)
  File "C:\my_practise\python_practise\reptile_basic.py", line 32, in analyseWeb
    print(pic[0][0])
IndexError: list index out of range

print(pic[0])
print(pic[0][0])
```

后面自己尝试的时候又可以了...

```
str = [(1,2,3)]
print(str)
print(str[0])
print(str[0][1])
```

```
[(1, 2, 3)]  
(1, 2, 3)  
2
```

字符串可以用split('/') 来进行有效切割

```
for url in urls:  
    file_name = './pic/' + url.split('/')[-1]  
    f = open(file_name, 'w')  
    f.write(str(url))  
    f.close()
```

