（七）下载全网所有小说

笔记本:	Markdown笔记		
创建时间:	2018/11/18 15:17	**更新时间:**	2018/11/20 11:39
作者:	FWJ		
URL:	https://stackoverflow.com/questions/30770213/no-schema-supplied-and-other-errors-with-using-request...		

环境

- 系统： deepin 15.8
- ide： vs code
- python : anaconda 3
- 主要库：re，multiprocessing， BeautifulSoup

Beautifulsoup库

beautifulsoup官方地址
https://www.crummy.com/software/BeautifulSoup/bs4/doc/index.zh.html

- find_all(name , attrs , recursive , text , **kwargs)
  按CSS搜索

```
soup.find_all("a", class_="sister")
# [<a href="http://example.com/elsie" >Elsie</a>,
#  <a href="http://example.com/lacie" >Lacie</a>,
#  <a href="http://example.com/tillie" >Tillie</a>]
```

  使用name参数
  name 参数可以查找所有名字为 name 的tag,字符串对象会被自动忽略掉.简单的用法如下:

```
soup.find_all("title")
#[<title>The Dormouse's story</title>]
```

- find
  使用上和find_all一样，唯一的区别是 find_all() 方法的返回结果是值包含一个元素的列表,而 find() 方法直接返回结果.find_all() 方法没有找到目标是返回空列表, find() 方法找不到目标时,返回 None .

```
print(soup.find("nosuchtag"))
# None
```

参考资料:
https://zhuanlan.zhihu.com/p/36895697

正则表达式

re.search

```
>>> pattern = re.compile("a")
>>> pattern.search("abcde")          # Match at index 0
>>> pattern.search("abcde", 1)    # No match;
```

re.compile

```
compiled_pattern = re.compile(pattern)
result = compiled_pattern.match(string)
```

参考资料:
http://tool.oschina.net/uploads/apidocs/jquery/regexp.html

https://www.ibm.com/developerworks/cn/opensource/os-cn-pythonre/index.html

多进程

multiprocessing库

```
    pool = multiprocessing.Pool(multiprocessing.cpu_count())
    for url_list in url_lists:
        pool.apply_async(get_all_txt, (url_list, ))
    pool.close()
    pool.join()
```

apply_async(func[, args[, kwds[, callback]]])
它是非阻塞，apply(func[, args[, kwds]])是阻塞的。

close()
关闭pool，使其不在接受新的任务。

join()
主进程阻塞，等待子进程的退出， join方法要在close或terminate之后使用。当然每个进程可以在各自的方法返回一个结果。apply或apply_async方法可以拿到这个结果并进一步进行处理。

参考资料：
https://cuiqingcai.com/3335.html

保存文件

with open

```
with open('./小说/{}.txt'.format(txt_name), 'a+') as f:
    f.write('小说标题: {} \n'.format(txt_name))
```

参考资料：
https://www.cnblogs.com/ymjyqsx/p/6554817.html

https://www.jianshu.com/p/5aaa4a4f793d

https://blog.csdn.net/qq_37383691/article/details/76060972

小错误

1. No schema supplied and other errors with using requests.get()
   大意