

这一篇也是关于爬取微信文章的，从十二号开始，以下是十二号的记录，引到这里

12号这一天做的是爬取微信文章的爬虫项目，与之前的相比，主体框架差不多，不过因为微信对于反爬处理比较严格，所以这一项目会用到比较多的异常处理内容，这在之前的项目中没有写过，写下异常处理也能增加爬取到内容的几率，格式也比较标准。

- 状态码判断
 - ``response.status_code``，200是正常，能够输出内容，其余的异常状态码根据情况来定
- 在请求的时候不允许页面自动跳转：
 - ``response = request.get(url, allow_redirects = False)``

之后还有一些新的知识点吧算是：

- 1.代理
从代理网站上面爬取的代理地址，然后在requests方法里面把代理作为一个参数传进去，就可以使用外部代理
`response = requests.get(url, allow_redirects = False, headers = self.headers, proxies = proxies)`
PS.这里的proxies要是字典，格式是{'协议名称': '代理地址'}，如{'http': 'http://171.39.78.71:8123'}
- 2.写入文件
将字符串格式写入txt文件的时候，尤其是网页数据写入，需要在打开文件的时候加上编码格式
如：`with open('{_}微信文章.txt'.format(query), 'a', encoding = 'utf-8') as f`

其实也没啥东西，妈的搞了好几天，有点懒惰了

——by 秦小灵 2018.10.15