

李晓华第二次培训任务记录

2018.10.08

晚上收到任务，开始 google 教程，安装 Python 开发环境,Python3.7

2018.10.09

- a. 开始在网上查找相关资料，开始了对 Python 的自学。
- b. 找同学借阅爬虫相关书籍，开始阅读 1~20 页。

出现的问题：照着网上教程做的时候会发现实际操作出来的结果会和上面的不一致，装的软件总是出现各种问题。

解决方法：多参考一些网上教程，把很大的问题拆分成几个小问题，再一个个百度。软件崩溃实在解决不了的就换个版本。

2018.10.10

由于当天课比较多，加上又有其他的事，所以只是在晚上回来之后在手机上查了几页资料。

2018.10.11

- a. 在前一天的基础上继续学习 Python,并在网上查找相关的实例代码。

出现问题：

看了一些 python 基础知识后，发现完成任务还是有点难，对不了解爬虫的相关知识

解决方案：

百度上查找爬虫的相关实例，文章。

2018.10.12—2018.10.13

这两天也是和 10 号一样,由于有其他的事拖不开身,只是在晚上回寝室后在手机上教程。

2018.10.14

- 1. 最后一天，开始写爬虫

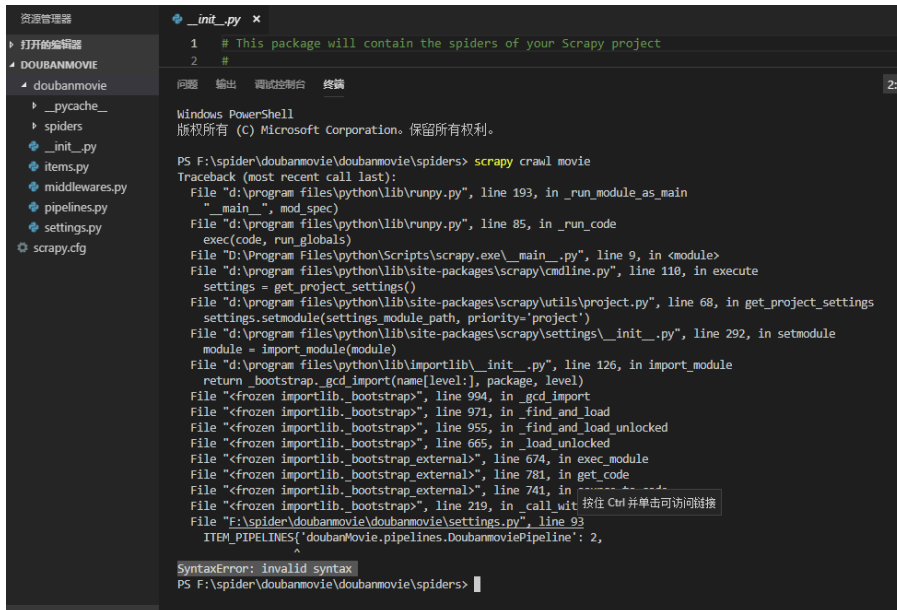
发现 Python3.7 不方便，重新弄了 3.6，配置好了 pip 和 Scrapy 模块后开始上手。

出现问题：

- a. 照着教程来时，最后 setting 文件里，设置 USER_AGENT 时，由于教程里的环境是 linux 系统，而自己用的是 linux，出现问题。

解决方案：在网上查找 USER_AGENT 的设置方法。

- b. 最后运行时，运行失败，错误代码为：SyntaxError: invalid syntax
图片如下：



```
1 # This package will contain the spiders of your Scrapy project
2 #
3 # See https://docs.scrapy.org/en/latest/topics/spiders.html for more
4 #
5 # You can extend this module by overriding the methods defined here.
6
7 import scrapy
8
9 class DoubanMovieSpider(scrapy.Spider):
10     name = 'doubanmovie'
11     allowed_domains = ['douban.com']
12     start_urls = ['http://douban.com']
13
14     def parse(self, response):
15         # Parse the response here
16         # Example:
17         # items = []
18         # for url in response.xpath('//a[@href="/movie/"]'):
19         #     item = DoubanMovieItem()
20         #     item['url'] = url.get('href')
21         #     items.append(item)
22         # return items
23
24     def __init__(self, *args, **kwargs):
25         super(DoubanMovieSpider, self).__init__(*args, **kwargs)
26
27     def __del__(self):
28         # Clean up resources here
29
30     def __str__(self):
31         return 'DoubanMovieSpider'
32
33     def __repr__(self):
34         return 'DoubanMovieSpider'
35
36     def __getitem__(self, item):
37         return self.__dict__[item]
38
39     def __setitem__(self, key, value):
40         self.__dict__[key] = value
41
42     def __len__(self):
43         return len(self.__dict__)
44
45     def __contains__(self, item):
46         return item in self.__dict__
47
48     def __iter__(self):
49         return iter(self.__dict__)
50
51     def __call__(self):
52         return self
53
54     def __getattr__(self, name):
55         return getattr(self.__dict__, name)
56
57     def __setattr__(self, name, value):
58         setattr(self.__dict__, name, value)
59
60     def __delattr__(self, name):
61         delattr(self.__dict__, name)
62
63     def __copy__(self):
64         return self.__dict__
65
66     def __deepcopy__(self, memo):
67         return self.__dict__
68
69     def __cmp__(self, other):
70         return self.__dict__ < other.__dict__
71
72     def __hash__(self):
73         return hash(self.__dict__)
74
75     def __eq__(self, other):
76         return self.__dict__ == other.__dict__
77
78     def __neq__(self, other):
79         return self.__dict__ != other.__dict__
80
81     def __lt__(self, other):
82         return self.__dict__ < other.__dict__
83
84     def __le__(self, other):
85         return self.__dict__ <= other.__dict__
86
87     def __gt__(self, other):
88         return self.__dict__ > other.__dict__
89
90     def __ge__(self, other):
91         return self.__dict__ >= other.__dict__
92
93     ITEM_PIPELINES = {'doubanmovie.pipelines.DoubanMoviePipeline': 2,
```

解决方案：在网上找错误代码的原因，尝试几次，还是失败。

2.上传任务文档和代码到个人的 github