本节知识点概述

爬取拉勾职位信息

数据库安装

MongoDB安装

可视化工具

基础使用

实战环节

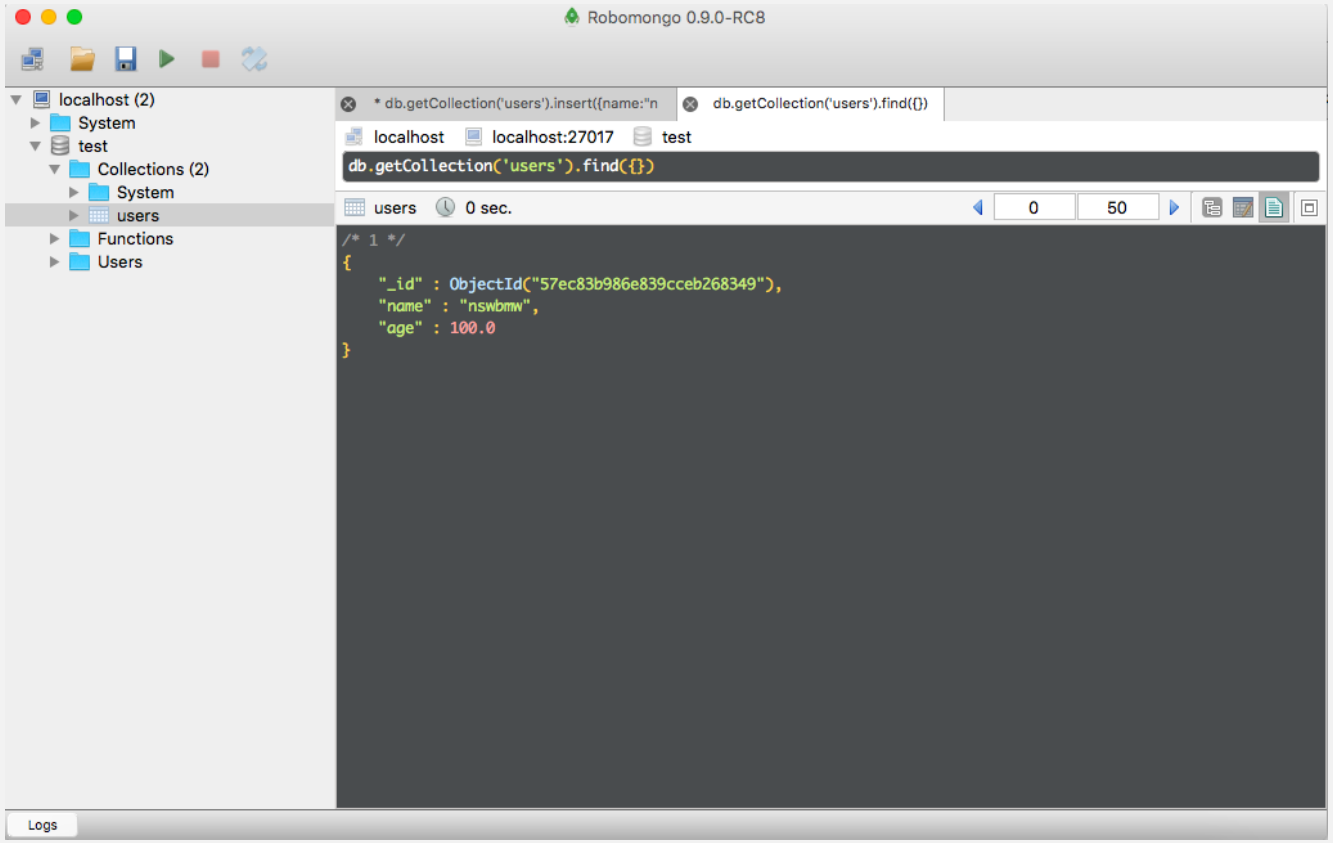
分析请求

动手写爬虫

MongoDB安装

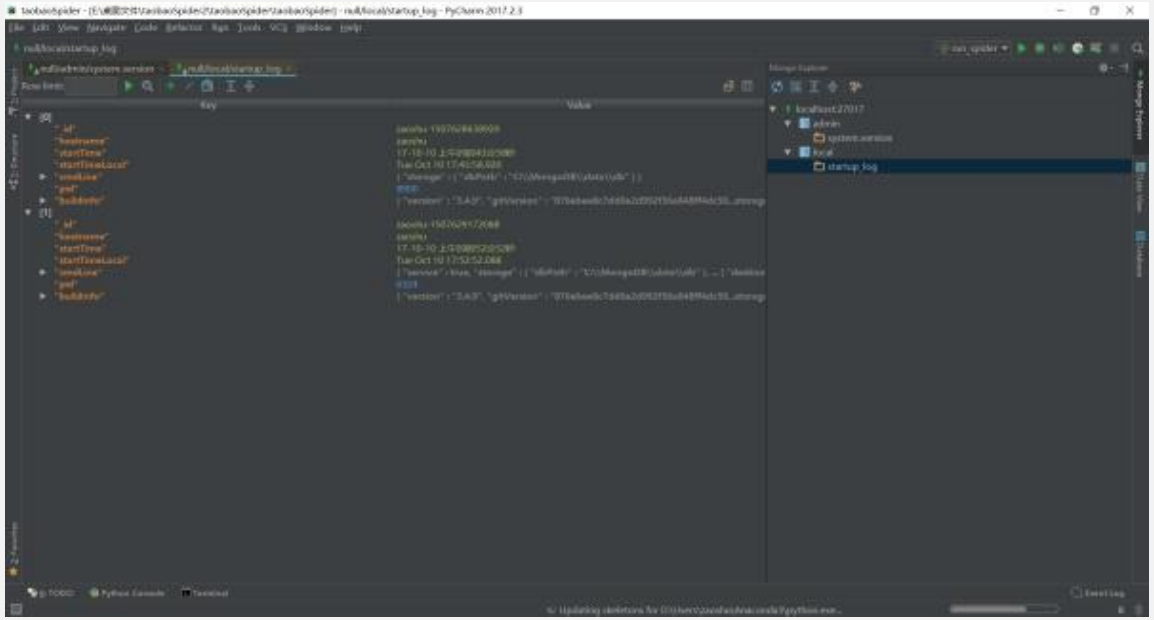
MongoDB 将数据存储为一个文档，数据结构由键值(key=>value)对组成
安装链接：

可视化工具



Robomongo

PyCharm插件



1、先安装第三方库

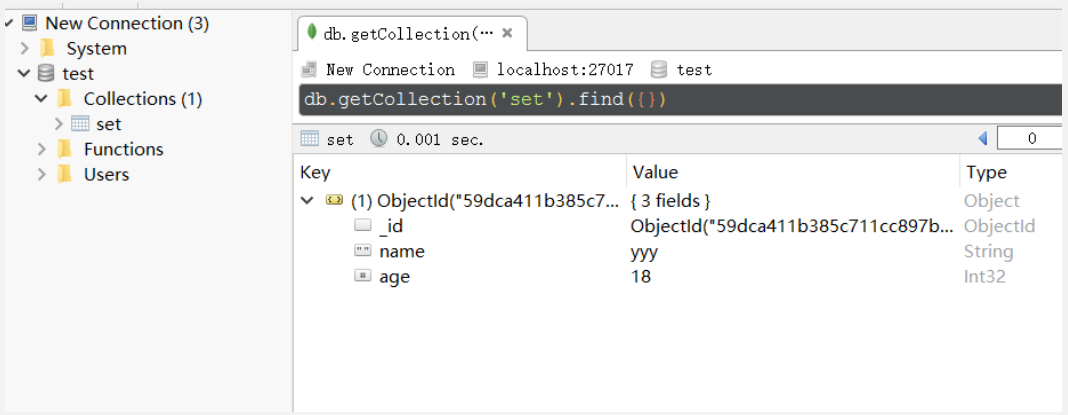
$ pip install pymongo

2、安装PyCharm的MongoDB插件

File——Plugins——mongo

基础使用

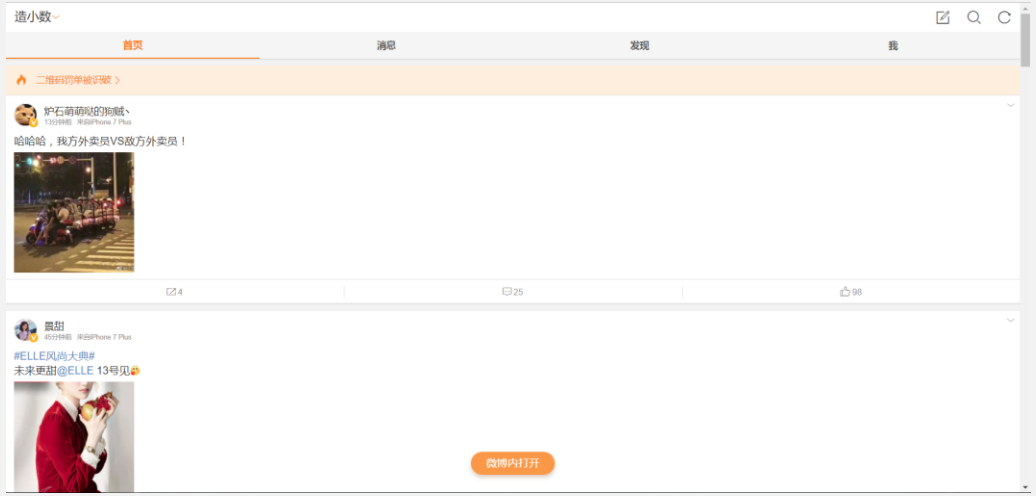
```python
#!/usr/bin/env python
# -*- coding:utf-8 -*-

from pymongo import MongoClient

client = MongoClient()
db = client.test   #连接test数据库，没有则自动创建
my_set = db.set    #使用set集合，没有则自动创建

my_set.insert({"name":"yyy","age":18})
```

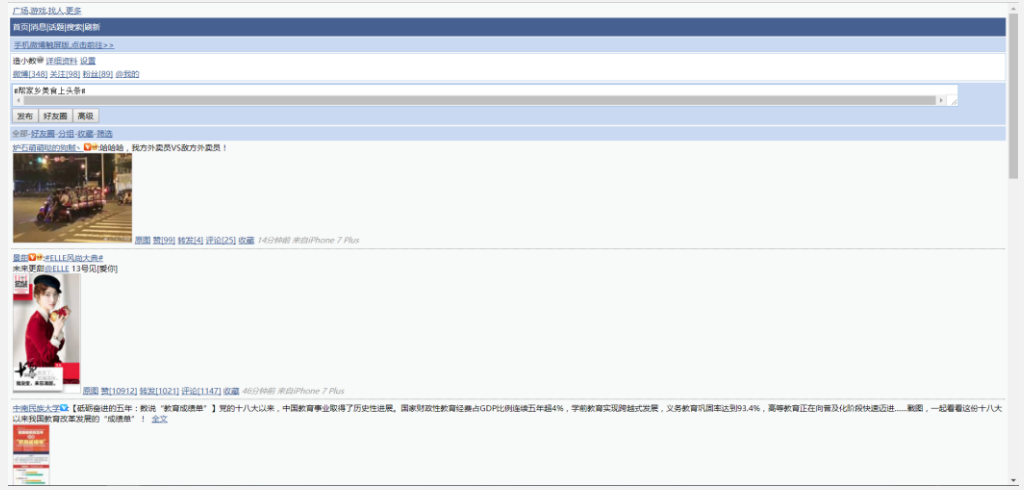


实战环节

https://m.weibo.cn/

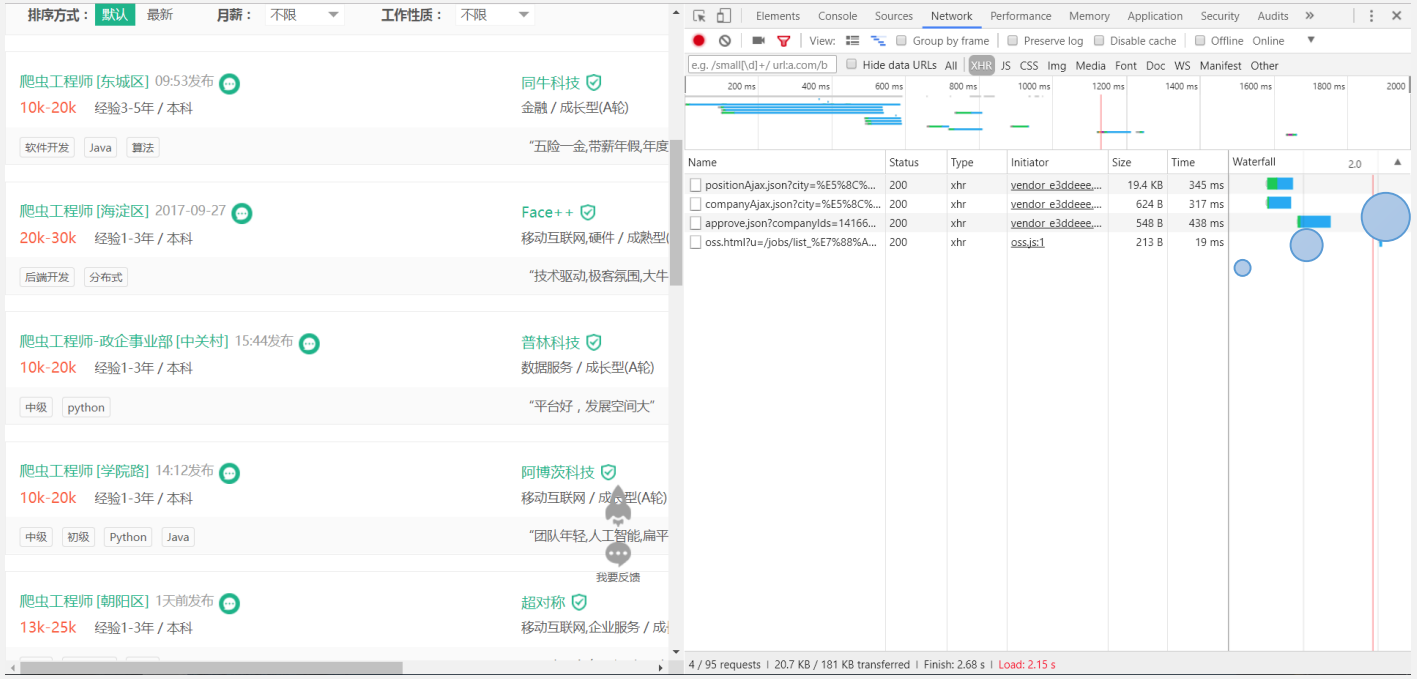
https://weibo.cn/





实战环节

找工作-互联网招聘求职网-拉勾网



寻找请求

动手写爬虫

lagou_spider.py

```python
# -*- coding:utf-8 -*-

import requests
from pymongo import MongoClient
import time
from fake_useragent import UserAgent


ua = UserAgent()

client = MongoClient()
db=client.test
lagou = db.lagou

headers = {
    'User-Agent':ua.random ,
```

```
正在爬取第 17 页的数据...
正在爬取第 18 页的数据...
正在爬取第 19 页的数据...
正在爬取第 20 页的数据...
正在爬取第 21 页的数据...
正在爬取第 22 页的数据...
正在爬取第 23 页的数据...
正在爬取第 24 页的数据...
正在爬取第 25 页的数据...
正在爬取第 26 页的数据...
正在爬取第 27 页的数据...
正在爬取第 28 页的数据...
正在爬取第 29 页的数据...
正在爬取第 30 页的数据...
[Finished in 98.1s]
```

完成作业

- 在电脑上安装MongoDB
- 尝试爬取多个职位数据
- 改进爬虫，爬取详情页数据

爬虫工程师

更多数据科学课程，上DC学院： class.pkbigdata.com

关注 DataCastle

关注造数

造数