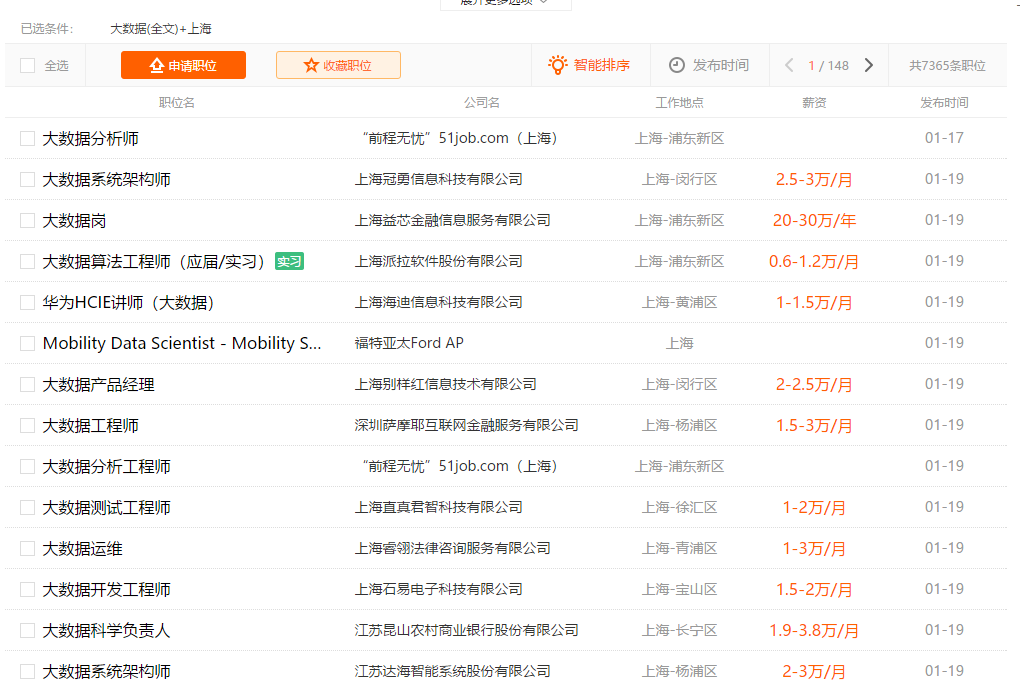
如何用Python找工作

------Python爬虫抓取51Job招聘信息

学习Python也有几个月了，该考虑换工作啦，看了眼招聘网站，就像这样。。。



信息玲琅满目不知该怎么办了，这时候，我想到了Python爬虫，如果能把信息都下载下来，可以随时随地的分析，这样一定可以找到称心如意的工作

一.准备工作 语言：Python3

工具：Pycharm，firefox

技术:requests，BeautifulSoup，re，csv

1. 流程概述 爬取51JOB相关工作第一页，获得页数

然后开始爬虫

获得每一页详细信息链接

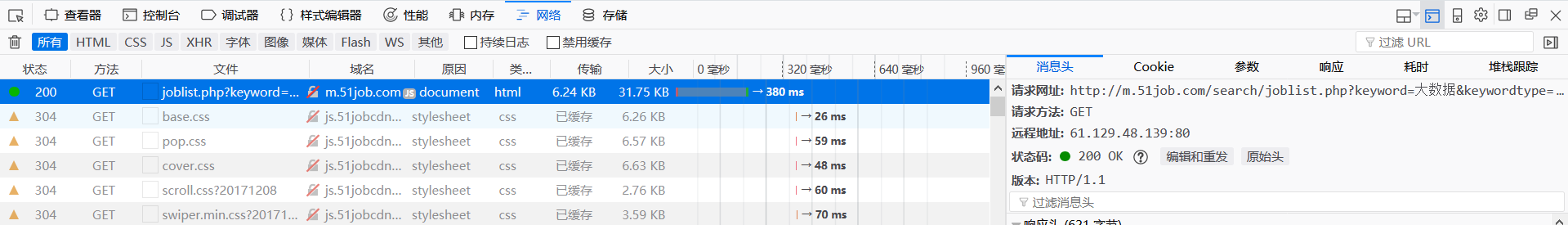
通过链接获取详细信息

存入CSV文件

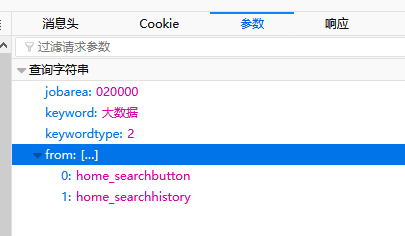
1. 具体实现

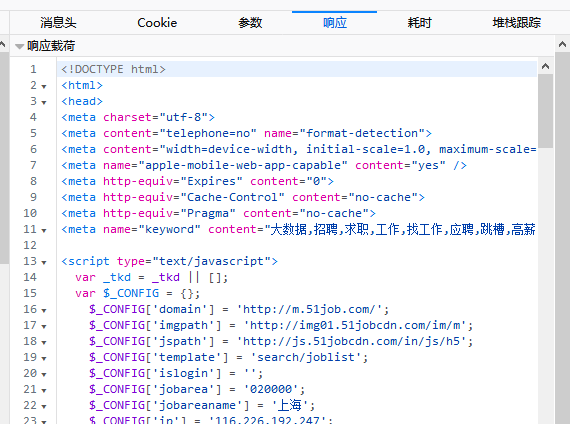
a)使用浏览器前往51JOB打开开发者工具，进行搜索，我们选择http://m.51job.com/,一般来说大型网站的PC端会有反爬虫措施，相对起来移动端会容易点，不过51的PC端也很容易爬，只是不喜欢那个页面风格而已。





这样信息就一览无遗了，从右边的窗口我们能看到url，参数，以及响应页面



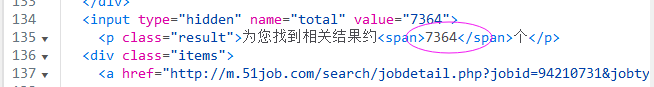


所以我们只要填写参数所有的数据要求，我们便能等到一串HTML代码，这就是我们要爬取的内容了。

b)我们需要先直接请求第一页，获取总数，使用requests获取数据，然后使用bs解析，这两个python库，API都很简单并且都比较轻量，很适合初学者学习使用，利用正则表达式获取总工作数量

数了下大约一页可以显示30条记录，那我们就能得出需要爬的页数





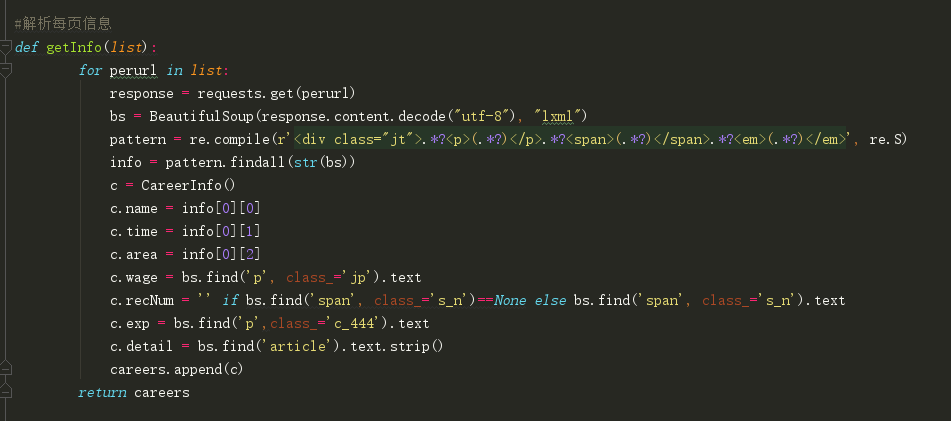
c)从密密麻麻的HTML中我们点开形如“http://m.51job.com/search”这样的链接，工作的详细信息映入我们眼帘，这就是我们数据的来源，我们把获得的链接存入一个列表中，便于我们接下来的调用。HTML我们使用BS4解析，BS4有四种解析器,lxml是需要安装C语言库的，我们可以使用html.parser库，lxml速度快但是在这个项目里并没有多明显。





d)点开链接，我们需要的信息一目了然，从中我们选取我们需要的信息，可以直接使用bs的方法获取，也可以使用正则表达式提取相关段落。特别注意的是有些字段在某些页面是不存在的，比如工作经验，需要判断。

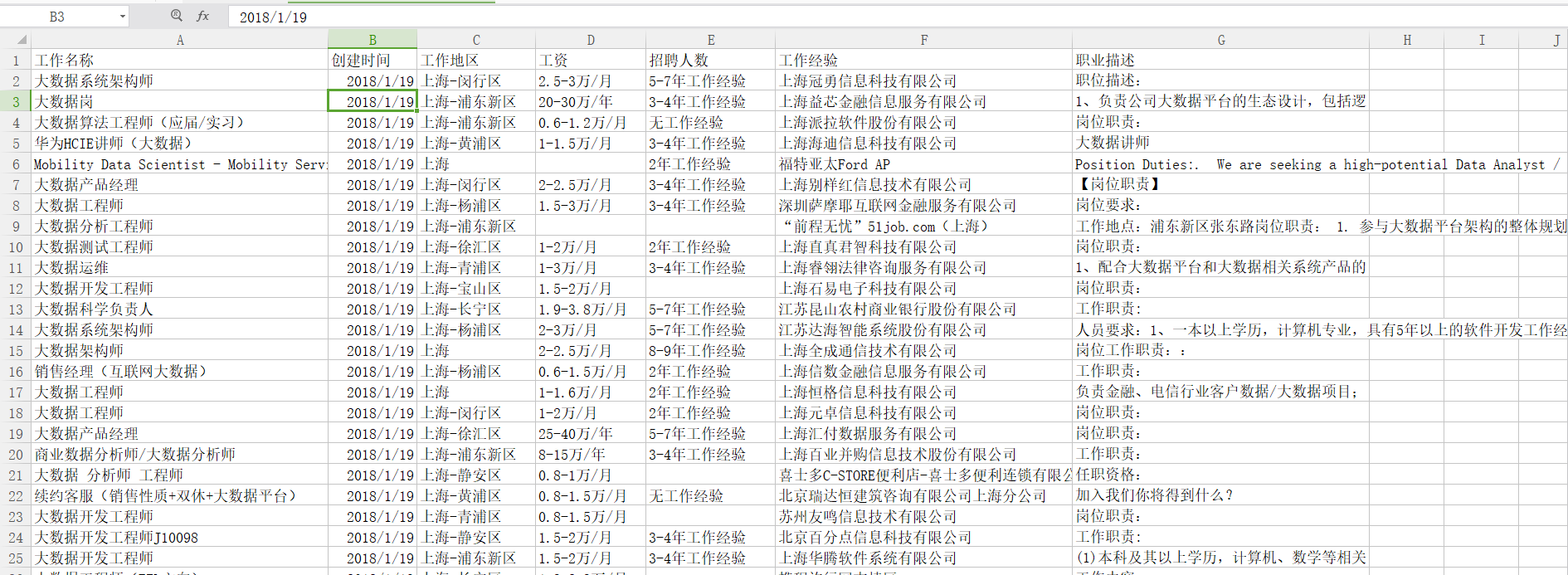




e)最后便是把信息插入到csv文件中，大家也可以使用mysql，mongoDB等工具存放数据。

这里特别注意的是，使用”w”模式写，会有空行，但是使用二进制”wb”模式会报二进制与字符串不匹配，所以在打开时候加入参数newline=’’,会避免空行。





最后完成如上图，可以使用熟悉的Excel软件进行排序操作，随时随地可以查看了。

Python作为现在最流行的一种语言之一，以其编写简单不罗嗦，类库数量庞大而著称。自从学习了python，厚重的JAVA早被我抛到九霄云外了，日常编写点应用类工具什么的真的特别方便。

编写应用并不困难，只要思路理清了，对流程进行分解，按块来编程，最后组装起来，如果错了，直接修改出错的块，并不需要动整个代码。