1.首先打开redis-server

$redis-server

2.插入评论区链接：

$ cd review\_links

$ python insert.py xxx

xxx用自己要插入年份的review文件，比如要爬1988年的xxx就改成1988\_reviews

3.开启爬取具体评论地址爬虫reviewLinks

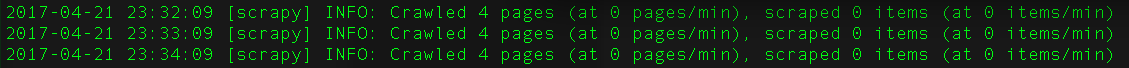
scrapy crawl reviewLinks

4.另开一个终端，开启爬取评论内容的爬虫review

scrapy crawl review

5.等所有数据爬取完毕后

比如出现：



就说明爬取完毕了。

此时执行先cd到本文档所在的目录，如果你之前cd到review\_links，现在cd ..一下。然后执行：

$ python process\_review.py

这时会生成一个csvfile.csv



把csvfile改成爬取年份



里面的内容应该为：



然后大家把这个文件发给我，我给大家分配下一个要爬的年份☺