爬虫笔记2.0

以下是阅读段学长所写程序的大体思路：

1. 通过观察网页源代码以及由刘老师的指导，我们可以得知这部分信息是原网页通过post的方式来获取的，而数据的格式是json

**网页源代码：**



1. 分析段学长所写程序（分步解读）：

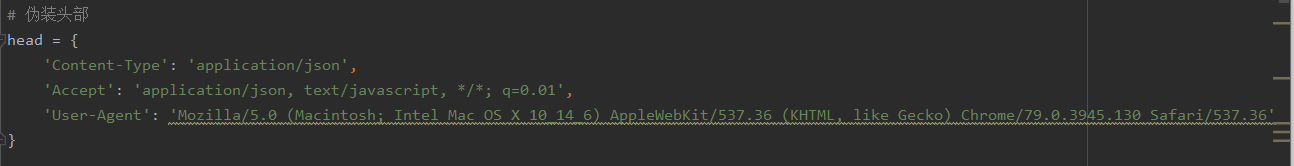
**首先**是库部分（由于我们要用到json、打开文件以及将数据导入到excel表中，所以我们要另外引入三个库：os是操作文件的库 xlwt：操作excel的库 json：用到json）：



其次找网址



隐藏头部（正在研究）：有什么作用？怎么写？



简单爬虫：爬取到的数据为：

原网页源代码：

