**修改后的改变：**

1.information.py文件和imformation\_extract.py文件效果差不多，由原来的整体爬取，更改为按市爬取。

2.增加了爬虫的效率，速度提升大概是之前的6倍左右，

3.增加了一些头文件，和防反爬措施，降低了被反爬机制识别的概率。

**文件使用说明：**

1. information.py直接运行提取主要信息，形成结构化文本，并最后形成一个record2.json的文件。这里只提取了500个网页信息，如有需要请自行添加。
2. listdir3.py直接运行，形成一个UI界面，从record2.json获取文本信息，并在此显示出来。

**注意：**

以上文件的运行环境在anaconda3.5以上的版本，需要另外安装jieba库，如有其它未安装的库，请自行安装。