Python网络爬虫基础实训第七周闯关作业

1. 问题描述：

本周课程主要讲解Python网络爬虫的基础内容。具体有：Python中的正则表达式，网络爬虫基础，以及urllib、urllib3和requests库的使用。本周闯关作业具体如下：

1）.分别使用userlib和requests爬取有道翻译的信息，要求输如英文后获取对应的中文翻译信息：

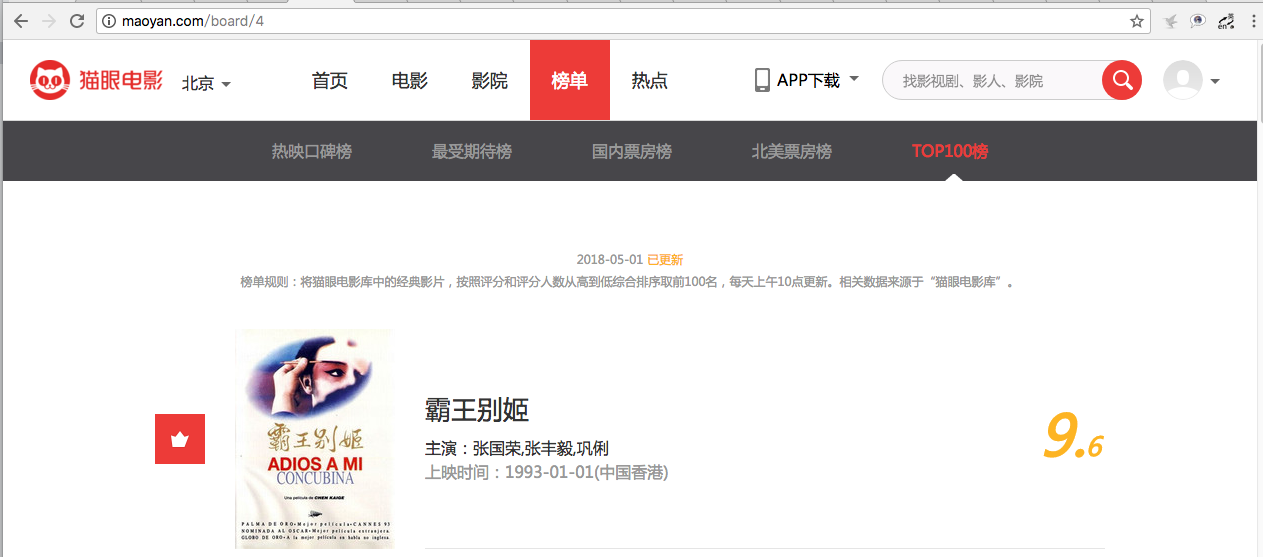


2 ). 分页爬取58同城的租房信息，信息内容要求有：【标题、图片、户型、价格】，并且获取指定页的所有租房信息:

如URL地址：http://bj.58.com/dashanzi/chuzu/pn1/?ClickID=1



3）. 爬取猫眼电影中榜单栏目中TOP100榜的所有电影信息（10页信息全部爬取），字段要求【序号、图片、电影名称、主演、时间、评分】，并将信息写入文件中: 具体参考URL地址：http://maoyan.com/board/4



1. 解题提示：

* 有道翻译信息的爬取可参考本周百度翻译信息爬取案例。
* 58同城的租房信息获取，首先先确定URL地址，编写爬虫程序，要使用正则表达式解析爬取的信息。
* 爬取猫眼电影中榜单栏目中TOP100榜的所有电影信息，首先却动url地址（分页获取），使用正则解析每页的信息，并将信息写入文件。具体代码实现可以参考笔记整理。

1. 批改标准：
   1. 有道翻译信息爬取（30分）
   2. 58同城的租房信息爬取（30分）
   3. 猫眼电影TOP100榜单信息爬取（30分）。
   4. 其他项：(10分)（文件结构清晰，代码整洁，要求适量的注释）。