Python爬虫期中作业

1.

静态爬虫：1.引用requests，BeautifulSoup库。 2.创建列表来对爬取的数据进行存储。3.发送URL 请求 获取豆瓣网页URL采用get方法获取 4.用find\_all方法来得到所要爬取的信息。4.将数据存储到列表中 5.最后输出爬取的数据

Django项目：1.创建项目 Movies ，以及APP项目 Movie 2.应用数据库文件sqlite 3.创建超级管理员superuser 4.. 定义模型 models

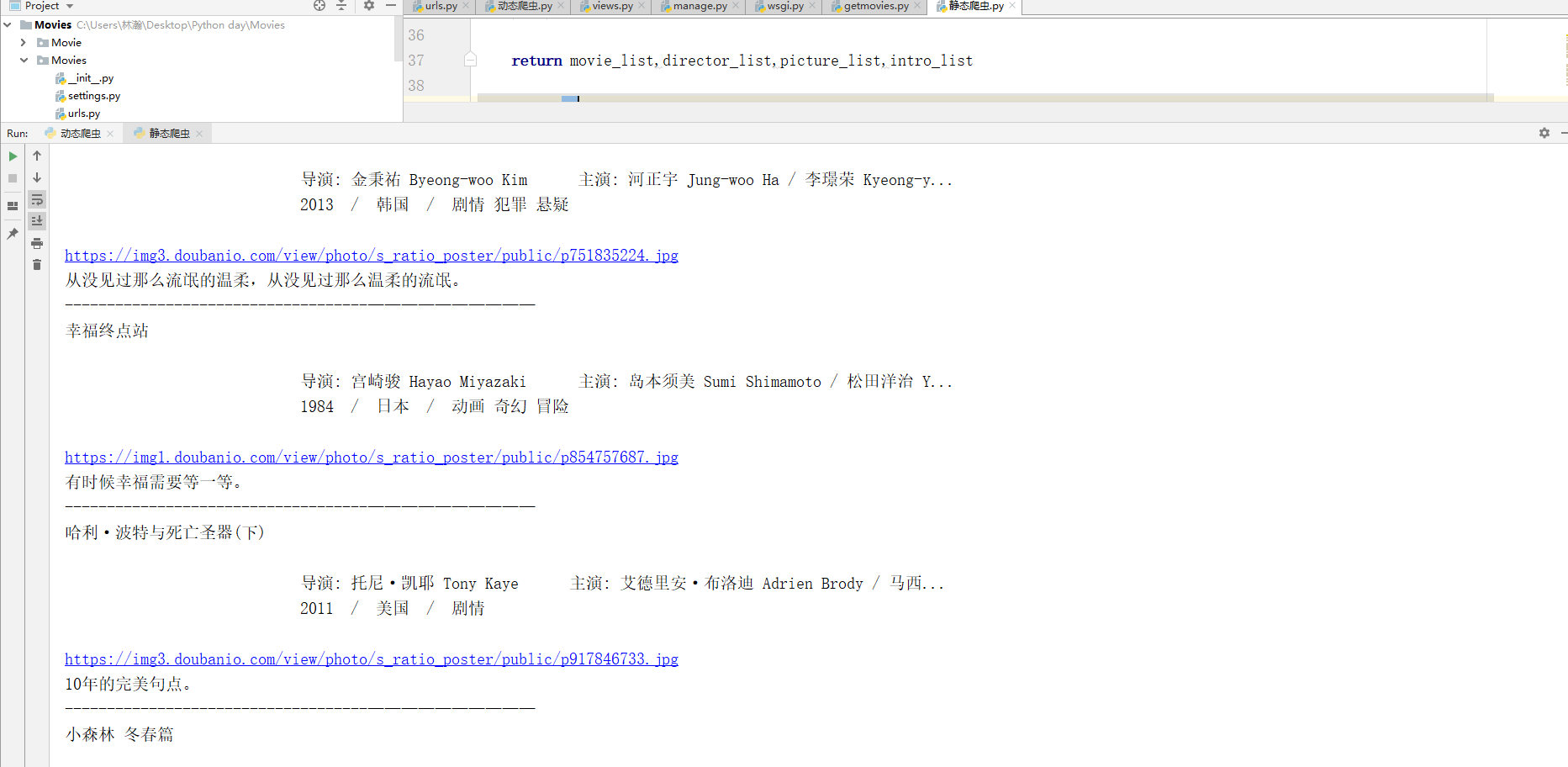
movieType，包含title（电影名），director（详细信息）,img（图片）,和dec（电影简介）。5.将模型models同步到数据库 ，并对admin进行编辑，将应用展现在后台管理界面。 编写views方法 object.get获取记录，将get\_movie 中爬取的电影信息存储到后台数据库中。6.编写templates 中的html模板文件，编写总路由编写一个网址，将后台数据库信息展示在该网址上。

动态爬虫：1.引入selenium库 ，对浏览器搜索关键字，得到查询后的商品列表

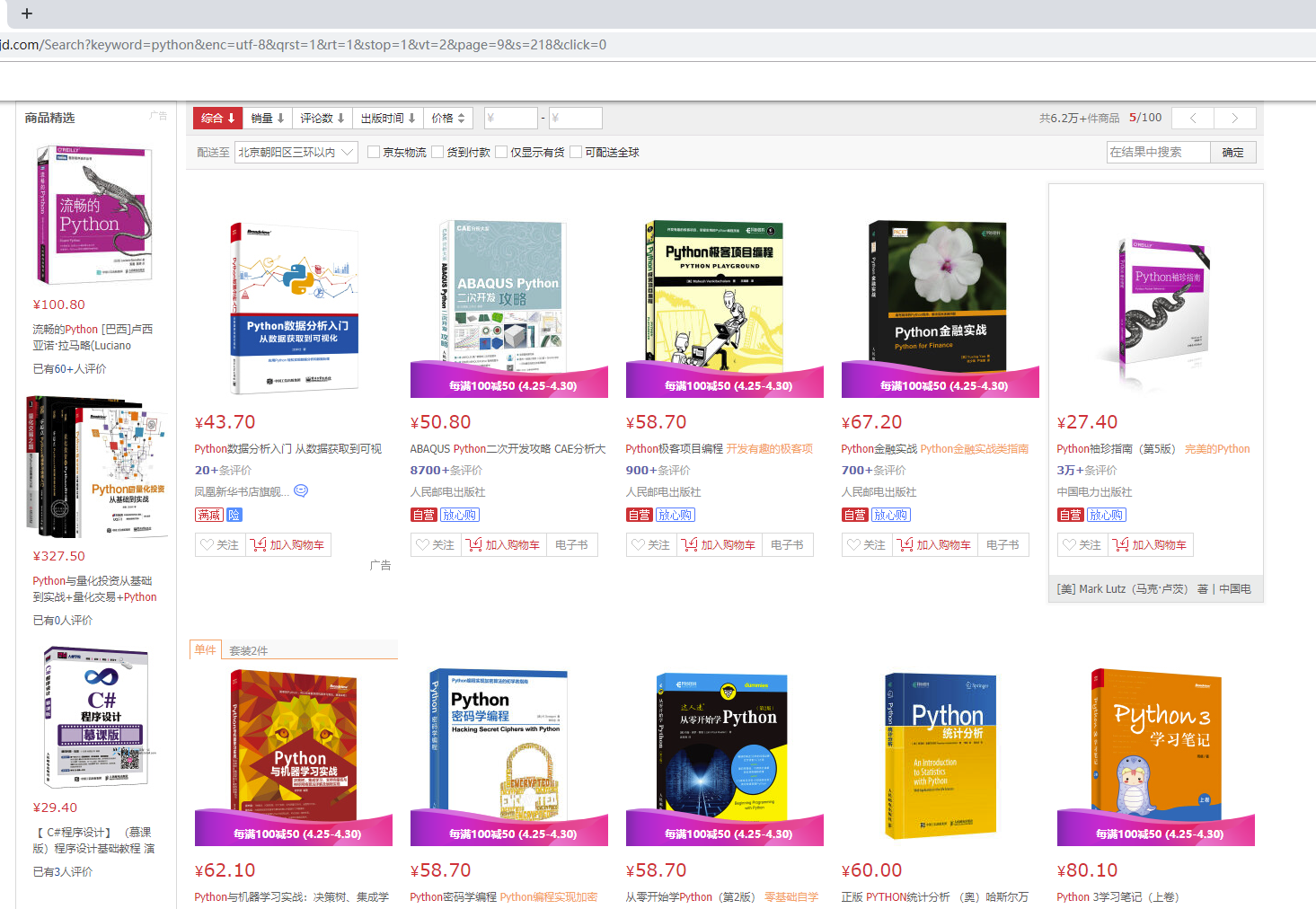
2.分析页码，得到商品的页码数，模拟翻页，下拉，得到翻页之后的商品列表。

3.利用lxml中的etree方法分析源码，用Xpath获取商品的各项信息数据，并输出。

2.效果截图：  
静态爬虫：



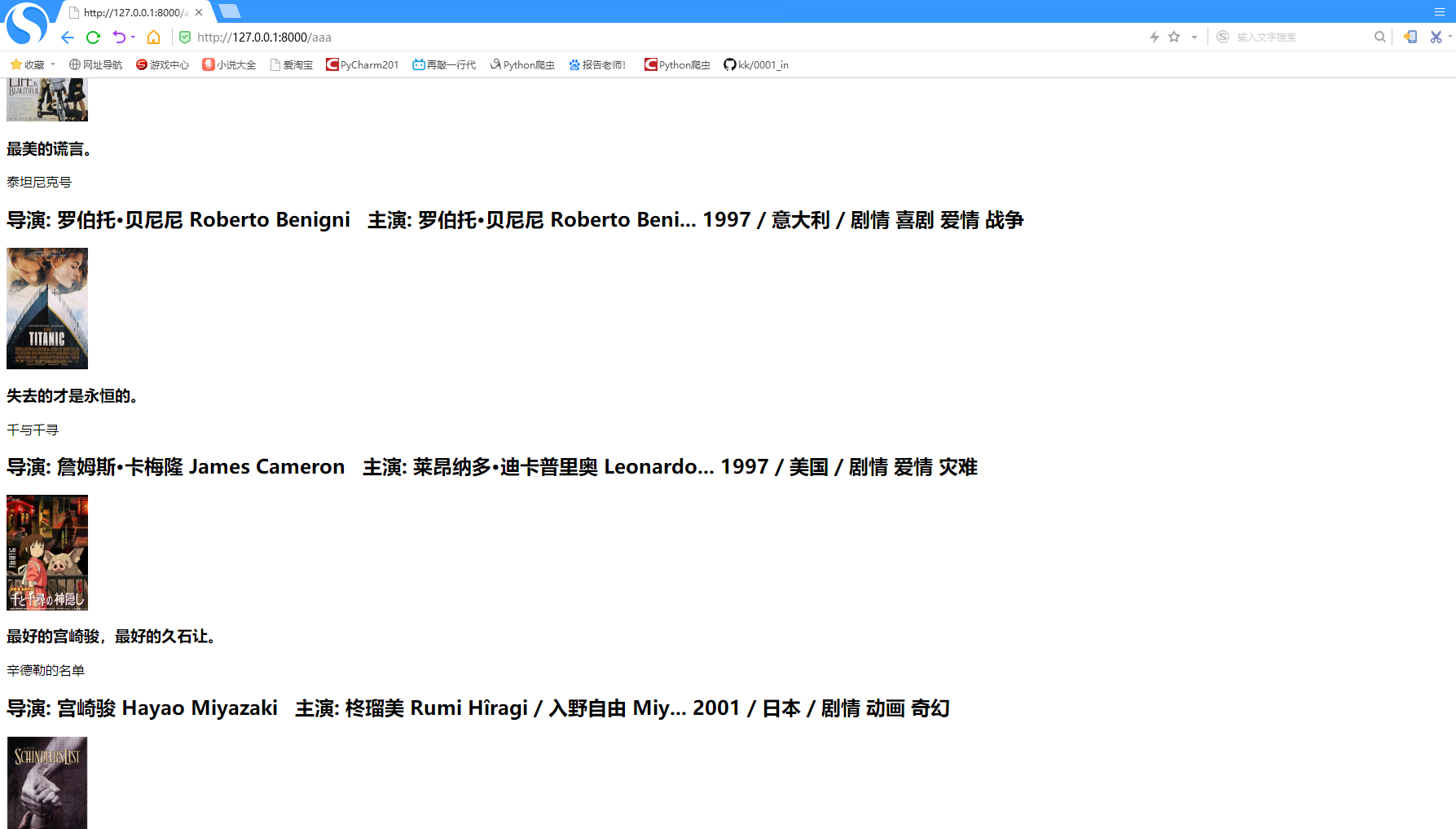
动态爬虫：





静态爬虫显示到Django网站上：

网页效果图：



Django 默认的数据库：

