



python爬虫期中作业报告



内容

CONTENTS



01

项目介绍

02

效果图截图

03

网站设计的思想

1.项目介绍

项目是基于Django的爬虫网站，项目下有三个文件夹：HelloWorld,templates,TestMode（1，2，3）。

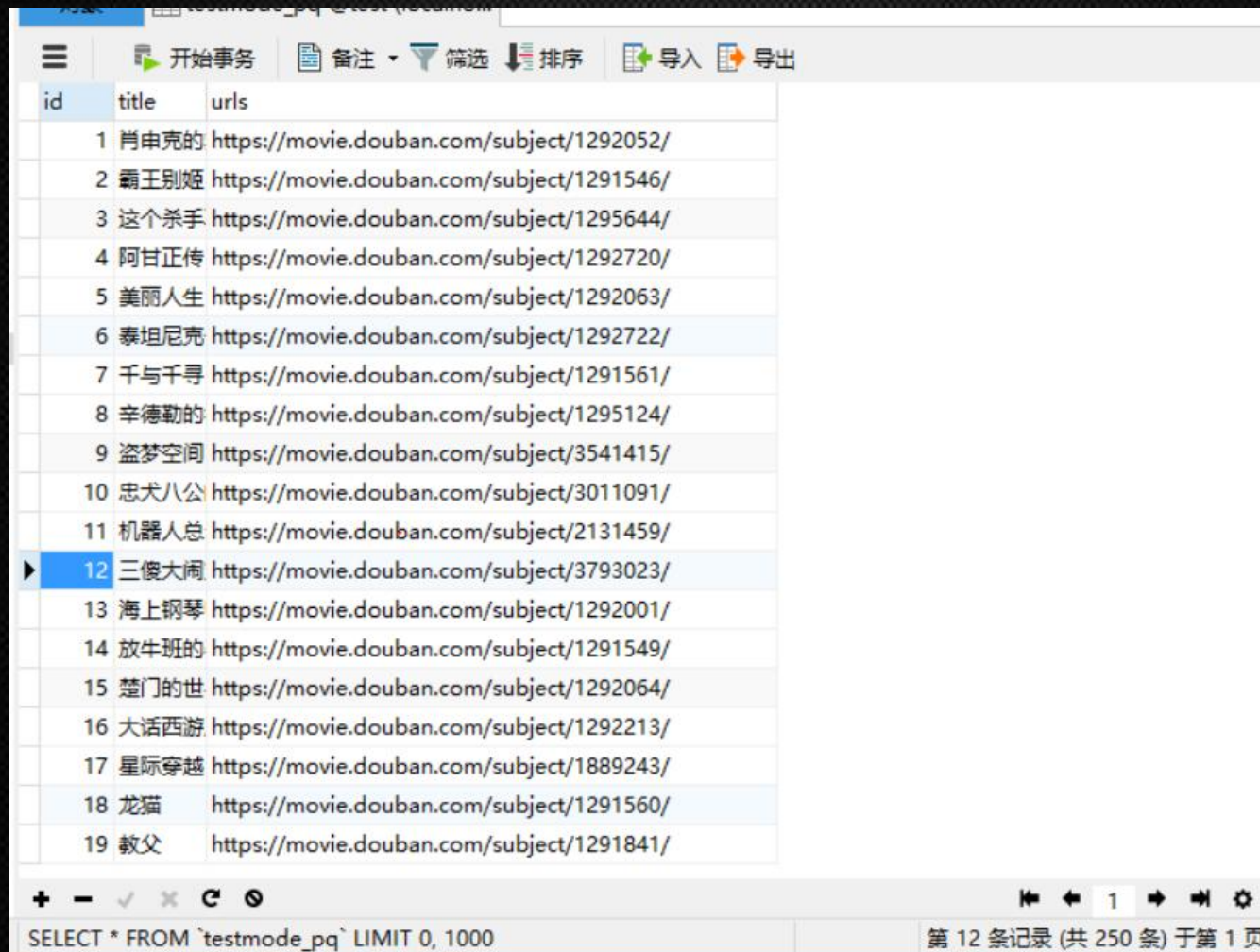
1：存放项目，里面除了Django自动添加的文件，还加入了三个视图：hello.py(首页)，testdb.py(电影)，weather.py(天气)，phone.py(手机)，然后在urls.py文件中绑定了视图函数和url

2：存放网站的显示模板

3：存放数据库模型

2.效果图截图

爬取到的数据库效果：1.豆瓣电影



id	title	urls
1	肖申克的	https://movie.douban.com/subject/1292052/
2	霸王别姬	https://movie.douban.com/subject/1291546/
3	这个杀手	https://movie.douban.com/subject/1295644/
4	阿甘正传	https://movie.douban.com/subject/1292720/
5	美丽人生	https://movie.douban.com/subject/1292063/
6	泰坦尼克	https://movie.douban.com/subject/1292722/
7	千与千寻	https://movie.douban.com/subject/1291561/
8	辛德勒的	https://movie.douban.com/subject/1295124/
9	盗梦空间	https://movie.douban.com/subject/3541415/
10	忠犬八公	https://movie.douban.com/subject/3011091/
11	机器人总	https://movie.douban.com/subject/2131459/
12	三傻大闹	https://movie.douban.com/subject/3793023/
13	海上钢琴	https://movie.douban.com/subject/1292001/
14	放牛班的	https://movie.douban.com/subject/1291549/
15	楚门的世	https://movie.douban.com/subject/1292064/
16	大话西游	https://movie.douban.com/subject/1292213/
17	星际穿越	https://movie.douban.com/subject/1889243/
18	龙猫	https://movie.douban.com/subject/1291560/
19	教父	https://movie.douban.com/subject/1291841/

SELECT * FROM `testmode_pq` LIMIT 0, 1000

第 12 条记录 (共 250 条) 于第 1 页

2.效果图截图

爬取到的数据库效果：2.城市天气

开始事务 备注 筛选 排序 导入 导出				
id	city	date	weather	temp
1	北京	14日 (今天)	晴	23/8℃
2	北京	15日 (明天)	多云	26/11℃
3	北京	16日 (后天)	晴	26/12℃
4	北京	17日 (周三)	多云	28/16℃
5	北京	18日 (周四)	多云	24/12℃
6	北京	19日 (周五)	多云转晴	19/9℃
7	北京	20日 (周六)	多云	21/10℃
8	上海	14日 (今天)	多云	20/7℃
9	上海	15日 (明天)	多云	20/13℃
10	上海	16日 (后天)	小雨转多云	18/13℃
11	上海	17日 (周三)	晴	28/16℃
12	上海	18日 (周四)	多云转阴	29/16℃
13	上海	19日 (周五)	阴	22/17℃
14	上海	20日 (周六)	多云	25/9℃
15	广州	14日 (今天)	中雨转多云	25/19℃
16	广州	15日 (明天)	雷阵雨转大雨	24/20℃
17	广州	16日 (后天)	大雨转阵雨	24/21℃
18	广州	17日 (周三)	多云	27/20℃
19	广州	18日 (周四)	多云转中到大雨	27/20℃

SELECT * FROM `testmode_weather` LIMIT 0, 1000

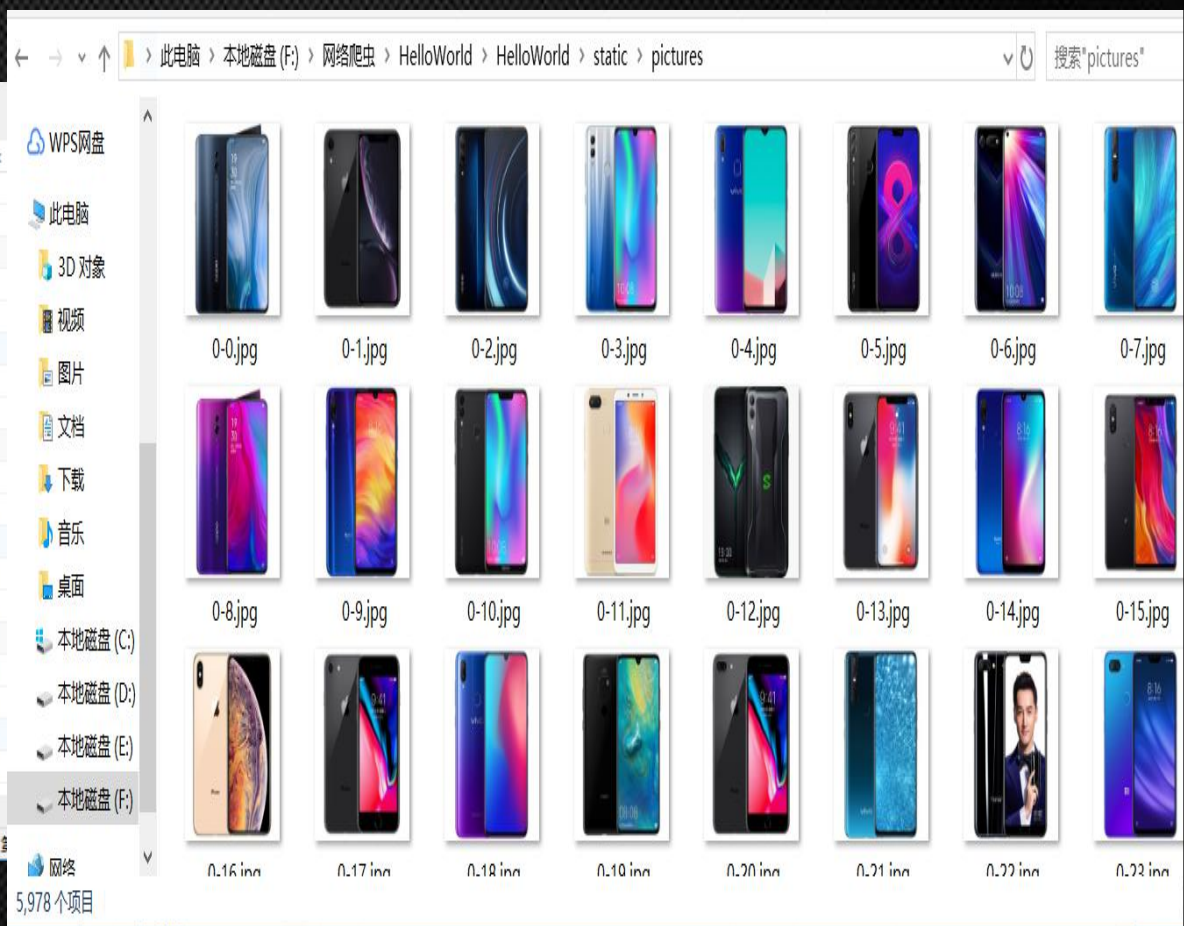
第 1 条记录 (共 35 条) 于第 1 页

2.效果图截图

爬取到的数据库效果：3.京东手机

id	mark	price	img_index
1	OPPO Reno手机 新品 全面屏拍照手机 全网通双卡双待手机 极	2999.00	0-0.jpg
2	Apple iPhone XR (A2108) 128GB 黑色 移动联通电信4G手机	5899.00	0-1.jpg
3	【KPL官方比赛用机】vivo iQOO 44W超快闪充 8GB+128GB	3298.00	0-2.jpg
4	荣耀10青春版 幻彩渐变 2400万AI自拍 全网通版4GB+64GB	1299.00	0-3.jpg
5	vivo U1 水滴全面屏 AI智慧拍照手机 3GB+32GB 极光色	799.00	0-4.jpg
6	荣耀8X 千元屏霸 91%屏占比 2000万AI双摄 4GB+64GB	1299.00	0-5.jpg
7	荣耀V20 胡歌同款 麒麟980芯片 魅眼全视屏 4800万深感相机	2799.00	0-6.jpg
8	vivo X27 8GB+256GB大内存 雀羽蓝 4800万AI三摄全面屏拍照	3598.00	0-7.jpg
9	OPPO Reno 全面屏拍照手机 6G+128G 星云紫 全网通 移动联	2999.00	0-8.jpg
10	小米 红米Redmi Note7 幻彩渐变AI双摄 4GB+64GB 梦幻蓝	1199.00	0-9.jpg
11	荣耀畅玩8C两天一充 莱茵护眼 刘海屏 全网通版4GB+32GB	899.00	0-10.jpg
12	小米 红米6 4GB+64GB 流沙金 全网通4G手机 双卡双待	799.00	0-11.jpg
13	黑鲨游戏手机2 8GB+128GB 暗影黑 骁龙855 Magic Press立	3499.00	0-12.jpg
14	Apple iPhone X (A1865) 64GB 深空灰色 移动联通电信4G手机	6199.00	0-13.jpg
15	小米 红米Redmi 7 幻彩渐变AI双摄 3GB+32GB 梦幻蓝 全网通	799.00	0-14.jpg
16	小米8SE 全面屏智能游戏拍照手机 6GB+64GB 灰色 骁龙710	1599.00	0-15.jpg
17	Apple iPhone XS Max (A2104) 256GB 金色 移动联通电信4G	9699.00	0-16.jpg
18	Apple iPhone 8 (A1863) 64GB 深空灰色 移动联通电信4G手机	3799.00	0-17.jpg
19	vivo Z3 6GB+64GB 极光蓝 性能实力派 全面屏游戏手机 移动	1598.00	0-18.jpg

ELECT * FROM `testmode_phone`



2.效果图截图

页面显示：首页截图



2.效果图截图

页面显示：豆瓣top250电影爬取

首页	豆瓣top250电影	天气爬取	京东手机爬取
豆瓣top250电影 数据爬取			
肖申克的救赎	https://movie.douban.com/subject/1292052/		
霸王别姬	https://movie.douban.com/subject/1291546/		
这个杀手不太冷	https://movie.douban.com/subject/1295644/		
阿甘正传	https://movie.douban.com/subject/1292720/		
美丽人生	https://movie.douban.com/subject/1292063/		
泰坦尼克号	https://movie.douban.com/subject/1292722/		
千与千寻	https://movie.douban.com/subject/1291561/		
辛德勒的名单	https://movie.douban.com/subject/1295124/		
盗梦空间	https://movie.douban.com/subject/3541415/		
忠犬八公的故事	https://movie.douban.com/subject/3011091/		
机器人总动员	https://movie.douban.com/subject/2131459/		
三傻大闹宝莱坞	https://movie.douban.com/subject/3793023/		
海上钢琴师	https://movie.douban.com/subject/1292001/		
放牛班的春天	https://movie.douban.com/subject/1291549/		
楚门的世界	https://movie.douban.com/subject/1292064/		
大话西游之大圣娶亲	https://movie.douban.com/subject/1292213/		
星际穿越	https://movie.douban.com/subject/1889243/		

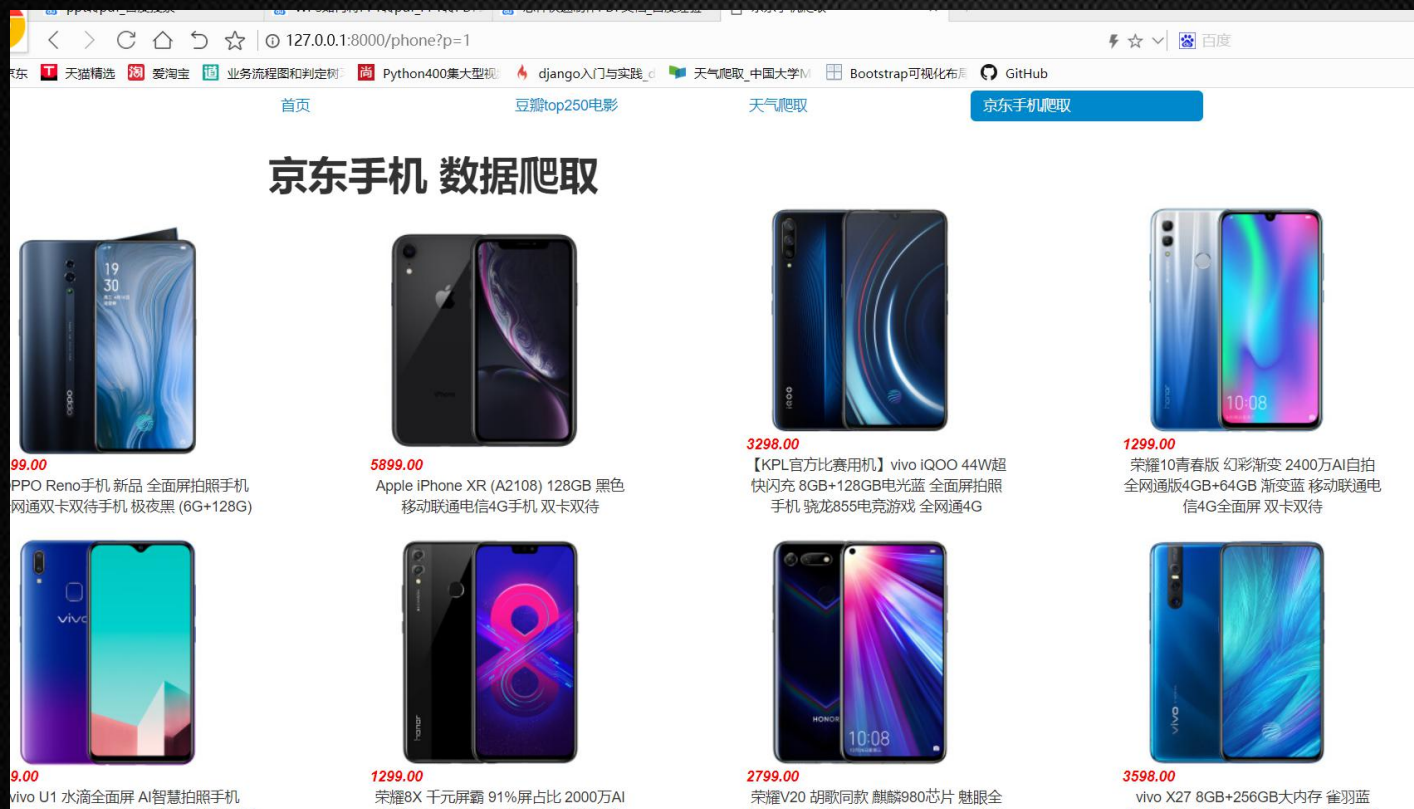
2.效果图截图

页面显示：城市天气爬取

☆ 127.0.0.1:8000/weather			⚡ ☆ ∨ 百度		
业务流程图和判定树			Python400集大型视		
尚			django入门与实践_d		
天气爬取_中国大学M			Bootstrap可视化布局		
GitHub					
首页			豆瓣top250电影		
天气爬取			京东手机爬取		
北京			上海		
14日 (今天)	晴	23/8°C	14日 (今天)	多云	20/7°C
15日 (明天)	多云	26/11°C	15日 (明天)	多云	20/13°C
16日 (后天)	晴	26/12°C	16日 (后天)	小雨转多云	18/13°C
17日 (周三)	多云	28/16°C	17日 (周三)	晴	28/16°C
18日 (周四)	多云	24/12°C	18日 (周四)	多云转阴	29/16°C
19日 (周五)	多云转晴	19/9°C	19日 (周五)	阴	22/17°C
20日 (周六)	多云	21/10°C	20日 (周六)	多云	25/9°C
广州			深圳		
14日 (今天)	中雨转多云	25/19°C	14日 (今天)	阵雨	26/21°C
15日 (明天)	雷阵雨转大雨	24/20°C	15日 (明天)	小雨	23/20°C
16日 (后天)	大雨转阵雨	24/21°C	16日 (后天)	大到暴雨	23/20°C
17日 (周三)	多云	27/20°C	17日 (周三)	阵雨转小雨	26/21°C
18日 (周四)	多云转中到大雨	27/20°C	18日 (周四)	小雨	27/23°C
19日 (周五)	中到大雨	25/22°C	19日 (周五)	小雨转中雨	28/23°C
20日 (周六)	中到大雨转小雨	26/25°C	20日 (周六)	中雨转小雨	27/24°C

2.效果图截图

页面显示：动态爬取-京东手机



上一页 1/100 下一页

跳转

2.效果图截图

淘宝数据爬取：程序运行

```
f:\网络爬虫\实验>cd f:\网络爬虫\实验 && cmd /C "set "PYTHONIOENCODING=UTF-8"
de\extensions\ms-python.python-2019.4.11881\pythonFiles\ptvsd_launcher.py --
obao.py "
```

序号	价格	商品名称
1	45.80	小学生书包男生1-3-4-6年级6-12周岁儿童
2	39.90	迪卡侬双肩包运动背包男女健身包书包儿童学生户外旅行包
3	119.00	kk树书包小学生女孩6-12周岁儿童1-3-6年级女童双肩背包
4	499.00	Fjallraven/北极狐双肩包kanken classic书包女户外旅行
5	129.00	小米双肩包简约休闲多功能书包男女笔记本电脑包时尚潮流
6	258.00	电视剧款JanSport旗舰店官网杰斯伯双肩包时尚女书包背包
7	348.00	爆款anello官方旗舰店日本ins潮风双肩女背包男离家出走
8	199.00	小米 米兔儿童书包 6-12岁男女小学生潮双肩背包幼儿园大
9	79.00	双肩包男士背包大容量旅行包电脑休闲女时尚潮流高中初中
10	109.00	七匹狼商务双肩包男书包中学生女电脑包旅行包休闲男士背
11	148.00	佑一良品男士背包双肩包男韩版大学生书包男时尚潮流大容
12	69.00	巴布豆旗舰店书包1-3年级护脊减负儿童书包男4-6小学生书
13	299.00	BOPAI博牌电脑背包男户外旅行休闲双肩包商务书包出差多
14	49.00	小学生书包6-12周岁 女儿童双肩包 3-5年级女童背包 1-3
15	45.80	儿童书包小学生男童1-3年级6-12周岁4-6年级男孩双肩背包
16	59.80	商务背包男士双肩包韩版潮流旅行包休闲女学生书包简约时
17	168.00	双肩包男书包男士时尚潮流青年休闲简约潮牌旅行背包大学
18	69.00	迪士尼书包小学生男女1-3-4-6年级米奇减负背包儿童书包
19	119.00	巴朗商务双肩包休闲时尚潮流大学生书包15.6寸电脑包男士
20	99.00	米熙休闲运动背包双肩包女书包中学生男韩版时尚大容量旅
21	195.02	国家地理背包女运动户外时尚双肩包男牛津布旅行防水学生

首页 豆瓣top250电影 天气爬取 京东手机爬取 淘宝商品

淘宝商品爬取

45.80	小学生书包男生1-3-4-6年级6-12周岁儿童
39.90	迪卡侬双肩包运动背包男女健身包书包儿童学生户外旅行包KIPSTA
88.00	迪士尼书包小学生女童1-3-6三年级公主女孩轻便儿童减负双肩背包5
499.00	Fjallraven/北极狐双肩包kanken classic书包女户外旅行背包23510
129.00	小米双肩包简约休闲多功能书包男女笔记本电脑包时尚潮流旅行背包
258.00	电视剧款JanSport旗舰店官网杰斯伯双肩包时尚女书包背包男大容量
348.00	爆款anello官方旗舰店日本ins潮风双肩女背包男离家出走包书包
199.00	小米 米兔儿童书包 6-12岁男女小学生潮双肩背包幼儿园大容量背包
79.00	双肩包男士背包大容量旅行包电脑休闲女时尚潮流高中初中学生书包
109.00	七匹狼商务双肩包男书包中学生女电脑包旅行包休闲男士背包大容量
148.00	佑一良品男士背包双肩包男韩版大学生书包男时尚潮流大容量旅行包
69.00	巴布豆旗舰店书包1-3年级护脊减负儿童书包男4-6小学生书包轻便
299.00	BOPAI博牌电脑背包男户外旅行休闲双肩包商务书包出差多功能男包
49.00	小学生书包6-12周岁 女儿童双肩包 3-5年级女童背包 1-3年级女孩
45.80	儿童书包小学生男童1-3年级6-12周岁4-6年级男孩双肩背包轻便减负
59.80	商务背包男士双肩包韩版潮流旅行包休闲女学生书包简约时尚电脑包
168.00	双肩包男书包男士时尚潮流青年休闲简约潮牌旅行背包大学生书包

3.网站设计的思想

总体设计：使用bootstrap框架，然后每个页面顶部都有一个导航栏进行页面跳转。

首页设计：进行爬取内容的简单介绍，还有自我介绍。

豆瓣电影页面：在页面里放了一个table，在table放入for循环，循环输出到table的每一行，每一列。

城市天气页面：先在视图函数用objects这个模型管理器的filter()获得所有城市名相同的数据，这样的列表有五个(因为数据库爬了五个城市的天气)，然后将列表及相应的城市名称变成字典，依次加入到一个新的列表中，用于输出。

在页面里放了一个for循环，每一项都会得到一个城市名字和一个列表，将城市名字引用到一个<h1>标签中，然后放入一个table，在table中放入for循环，循环输出子列表到table的每一行，每一列。

3.网站设计的思想

京东手机页面-视图函数：因为用到了分页，所以在phone视图函数中先获得了所有数据的列表，但没有将所有的数据都加入到context中输出，用到了rerequest.GET.get()函数，获得url中的页码参数，然后设置每一页输出60个数据，计算应该输出的数据在列表的什么位置（`list[(页码-1)*60: 页码*60]`），然后将其输出。

还有输出页面中需要的参数:上一页对应的页码参数；下一页对应的页码参数；总页码参数；当前页码参数；

最后在相应位置加入判断，使页码参数不小于1且不大于总页码。

3.网站设计的思想

京东手机页面-页面模板：显示手机部分先放排列，在里面放入循环，循环输出每个及相应内容，设置样式使多个在一行输出。

上一页，下一页用<a>标签，设置相应的href属性。然后加入一个文本框，一个按钮绑定点击按钮事件，用js对输入文本框的值进行判断，如果条件成立就跳转到相应界面。

淘宝商品爬取：直接将爬到的数据显示在网页上。没有加到数据库中。



THANKS