

Django 期中实验报告

姓名：毛人杰

学号：16219111421

班级：16 计算机四班

一、实验目的：

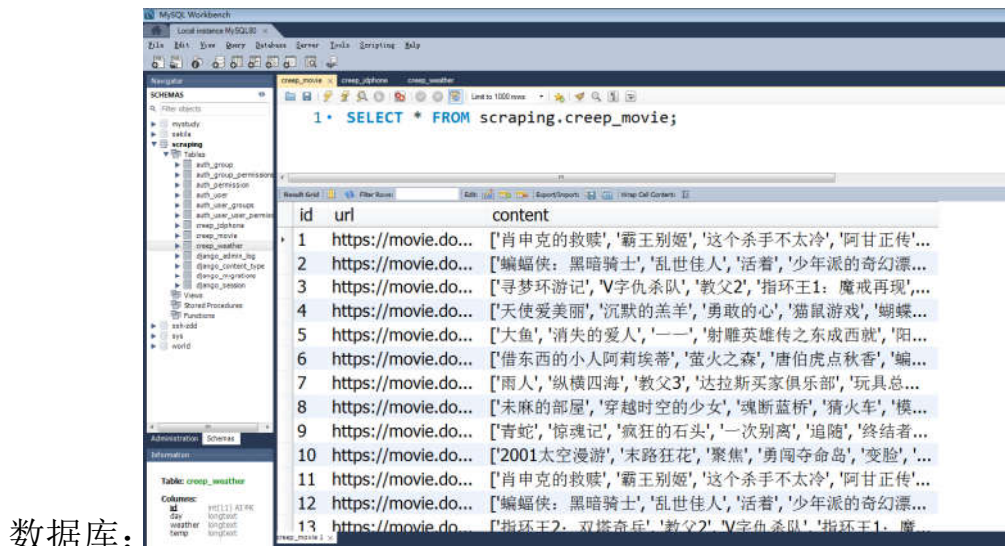
1. 掌握静态爬虫，爬取豆瓣电影
2. 掌握动态爬虫，爬取京东搜索手机界面
3. 掌握将爬取的数据写入指定数据库
4. 掌握 Django 管理网站，连接数据库
5. 用 bootstrap 框架写网站

二、实验环境：

Python、Django、bootstrap、VS Code

三、实验内容：

1. 主页面



数据库：

3. 天气爬虫页面



The screenshot shows the MySQL Workbench interface. The left sidebar displays the 'SCHEMAS' tree with 'scraping' selected. The main window shows a query: `1. SELECT * FROM scraping.creep_jdphone;`. The 'Result Grid' displays the following data:

id	item_name	item_price	item_page
55	魅族 Note9 全面屏游戏拍照 4GB+64GB 幻黑 全网通移动联...	1398.00	Page 1
56	华为 HUAWEI P20 Pro 全面屏徕卡三摄游戏 6GB+128GB 亮...	3688.00	Page 1
57	华为 HUAWEI P30 超感光徕卡三摄麒麟980AI智能芯片全面...	4288.00	Page 1
58	OPPO K1 光感屏幕指纹 水滴屏拍照 4G+64G 摩卡红 全网...	1399.00	Page 1
59	华为 HUAWEI Mate20X 麒麟980芯片全面屏超微距影像超大...	4489.00	Page 1
60	飞利浦 PHILIPS E258S 宝石蓝 直板按键 移动/联通2G 老人 ...	166.00	Page 1
61	-----	-----	-----
62	小米 红米Redmi 红米7 智能老人 梦幻蓝 全网通4G+64G	939.00	Page 2
63	飞利浦 (PHILIPS) E163K 陨石黑 移动联通2G直板按键老人 ...	128.00	Page 2
64	诺基亚 NOKIA X71 6GB+64GB 太空黑 蔡司认证 后置三摄 全...	2199.00	Page 2
65	华为 (HUAWEI) 荣耀9i (荣耀直供 限时抢购) 幻夜黑 全...	1198.00	Page 2
66	小米MIX2S 全面屏游戏拍照 6GB+128GB 黑色 骁龙845处理...	2699.00	Page 2
67	小辣椒 红辣椒7X 4+64GB 学生智能 美颜双摄 微O多开 人脸...	689.00	Page 2

四、实验思考:

1. 三个爬虫界面都是用 etree 写的,之后可以尝试用 beautifulsoup 试试

2. 京东搜索手机界面,动态爬虫,也是看的网上的教学,把一张页面分为前 30 个商品和后 30 个商品。前 30 个商品用正常爬虫就能爬取,后 30 个商品是去找下拉界面,界面发送的请求,再去获取我们需要的请求头,来模拟浏览器。但是在找后 30 个商品的 url 规律的时候,不知道怎么才找的准确。

https://search.jd.com/s_new.php?keyword=%E6%89%8B%E6%9C%BA&enc=utf-8&qrst=1&rt=1&stop=1&vt=2&cid2=653&cid3=655&page=2&s=27&scrolling=y&log_id=

这里的 s 的值每次打开页面都是不一样的数值,虽然说把前 10 个页面爬出来了,但是爬的应该不是很准确。

3. 如果能每天自动刷新就好了,不然页面显示的都是我上次写入到数据库的数据,要看最新的还要重新执行一遍。

4.