技术报告

计算机1502

20154398

贾海峰

1. 如何实现抽取

我选择了zol的手机报价和数码相机的报价，每种有104个网页，从<http://detail.zol.com.cn/cell_phone_index/subcate57_0_list_1_0_1_2_0_1.html>到<http://detail.zol.com.cn/cell_phone_index/subcate57_0_list_1_0_1_2_0_104.html>，从<http://detail.zol.com.cn/digital_camera_index/subcate15_0_list_1_0_1_2_0_1.html>到<http://detail.zol.com.cn/digital_camera_index/subcate15_0_list_1_0_1_2_0_104.html，从这些网页中抽取数据，大约10000>条

先用bs4解析得到的html网页，再用正则匹配得到想要的数据，再用xlwt将数据写入excel中。

1. 解决了哪些问题
2. 我先打算在列表页中把这些手机和相机的详情页保存下来之后再访问得到想要的数据，后来发现网页太多，就直接用了列表页的数据
3. 没有经验，很多都是现学，代码写的也很乱。