**Python开发文档**

徐瑾卿 5140379026

**功能实现：**

一：抓取一号店商品信息

* 第三方库：bs4、urllib、json、requests,re
* 实现效果：

输入是一个搜索字符串，输出时一个json文件。效果是输入关键词后抓取在一号店上综合排名前三个商品，商品信息包含：名称，详情链接、图片、评分、价格

* 基本思想：

通过观察1号店的网页，发现输入关键词后的网页里可以抓取的信息有商品链接、价格、评分等等。而把抓取的链接作为url进入详情界面可以抓取的信息是名称、图片。价格和评分在详情界面是通过动态加载的，因此我选择在一开始输入关键词的界面一次性抓取前3个商品的价格和评分。

* 遇到的问题与解决方法：

1.拼接url时遇到中文的编码不对，导致url错误，无法抓取信息。原来是中文编码默认用的coding形式是cp936，而1号店的url用的编码用的是utf-8，所以加上encoding=utf-8即可。

2.抓取信息时遇到动态加载的问题在详情页抓取商品的评分和价格时，我使用浏览器的审查元素功能找到了想要信息对应的位置，然而通过bs4解析网页却怎么也无法抓到信息。后来在查看网页源代码中发现这部分是动态加载的，因此选择不在详情面抓取信息。

3.抓取的价格中含有无效字符。解决方法：通过re正则匹配

* 代码架构：

1.全局变量basket是3个goods的list

2. prepare函数负责拼url

3. processor 负责抓取，其中调用了5个帮助函数get\_name, get\_picture, get\_rate, get\_price, get\_source分别抓取5块信息

4. save\_img\_to\_file负责把图片存到本地

5. output负责将goods转成dic，然后输出json

6对外接口是get()

二．保存功能

**功能实现：**

* 第三方库：matplotlib
* 实现效果：

每次点击保存的时候，将每个商铺的平均价格和查询时间通过文件的方式保存下来，通过做折线图来反映价格的变化趋势

* 基本的思想：

把每次的查询记录添加到字典，字典的key是查询的关键字，字典的value是一个二重的list prices，prices里是每次查询的价格list price，price里是1次查询中4个商铺的平均价格以及查询的时间

* 代码架构：

Saveprice函数负责保存价格，dump到文件中。通过load先把文件中的内容读到字典中，如果字典中已经有这个关键字对应的价格纪录，往prices list里加一条记录，如果没有，添加新的键值对

Drawpic()负责读取数据，画折线图

**收获与感想：**

这两块工作让我熟悉了bs4、字符串操作和matplotlib，第一次写python爬虫，花了不少时间研究，很有收获。