**Python大作业——“Gungnir”比价软件 说明文档**

徐瑾卿 5140379026

樊泽坤 5140379046

计婷 5140379049

楼思琦 5140379052

(项目主页: <https://github.com/odinaryk/Gungnir> )

1. **作业题目**

* 做个比价的软件，输入关键字告诉我淘宝京东一号店等等网站哪家的东西更便宜？

1. **代码运行环境**

* 操作系统：win8/win10；
* Python版本号：python2.7；
* 第三方库名字及其版本号：BS4；PYQT4；Matplotlib；urllib；json；requests；

1. **作业内容概要**

大作业布置的时候正好是双十一刚刚结束，淘宝当天总消费值高达912亿，网购似乎已经成为了人们日常生活中不可缺少的部分。然而，“货比三家”的想法并没有消失，以前是需要跑好多家商场才能比较出哪一家的商品性价比最高，而如今我们需要浏览很多家购物网站来使我们的钱花的最值。

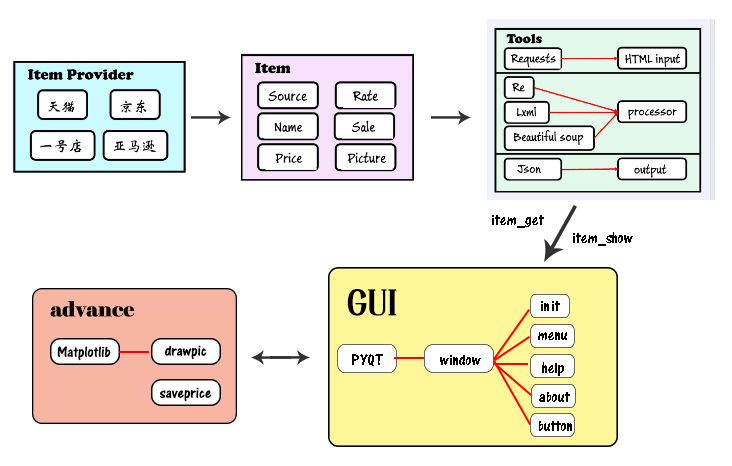
Python语法简洁清晰，具有丰富和强大的库。它常被昵称为胶水语言，能够把用其他语言制作的各种模块（尤其是C/C++）很轻松地联结在一起。而且python在抓取网页信息中是非常方便的，有很多工具可以直接使用，因此常常用python写爬虫。

基于此，我们决定用python写一个比价软件，利用上课提到的BeautifulSoup来抓取网页信息，利用PYQT提供GUI界面设计。同时考虑到往往单一的价格比较不足以满足用户的需求，增加了每个商品的具体信息（销量、评分、图片等），同时利用xxx实现历史价格记录，用户可以通过analyze zone选择自己想要的比较方式，也可以通过价格曲线直观地比较历史价格走势。

软件名为“Gungnir”，Gungnir是北欧神话主神Odin的武器，特点是一经出手永不落空，用以喻作一买即中。体现此比价软件的可靠性。Logo则为程序猿，体现开发者的特点。

1. **设计思路**

* 程序框架



每一个商品被定义为一个item，其中包含这个商品的URL、名称、价格、评分、销量和图片共六个要素。通过对各自网站的源代码解析，利用第三方库（request、re、BS等）直接从网站抓取相关信息提供给item。最后统一将item定义为dict，利用json的dump()实现输出。

再通过调用预先设计的接口，调用不同网站信息的模块获取json数据，再从各个文件中读取后获得itemlist。构造自定义类，将item输出到前端GUI中。

除此以外，用一个价格表储存搜索过的价格，再利用第三方库matplotlib将历史价格走势以折线图的方式直观地呈现给用户，让用户对该商品有更多的了解。

* 图形界面设计说明

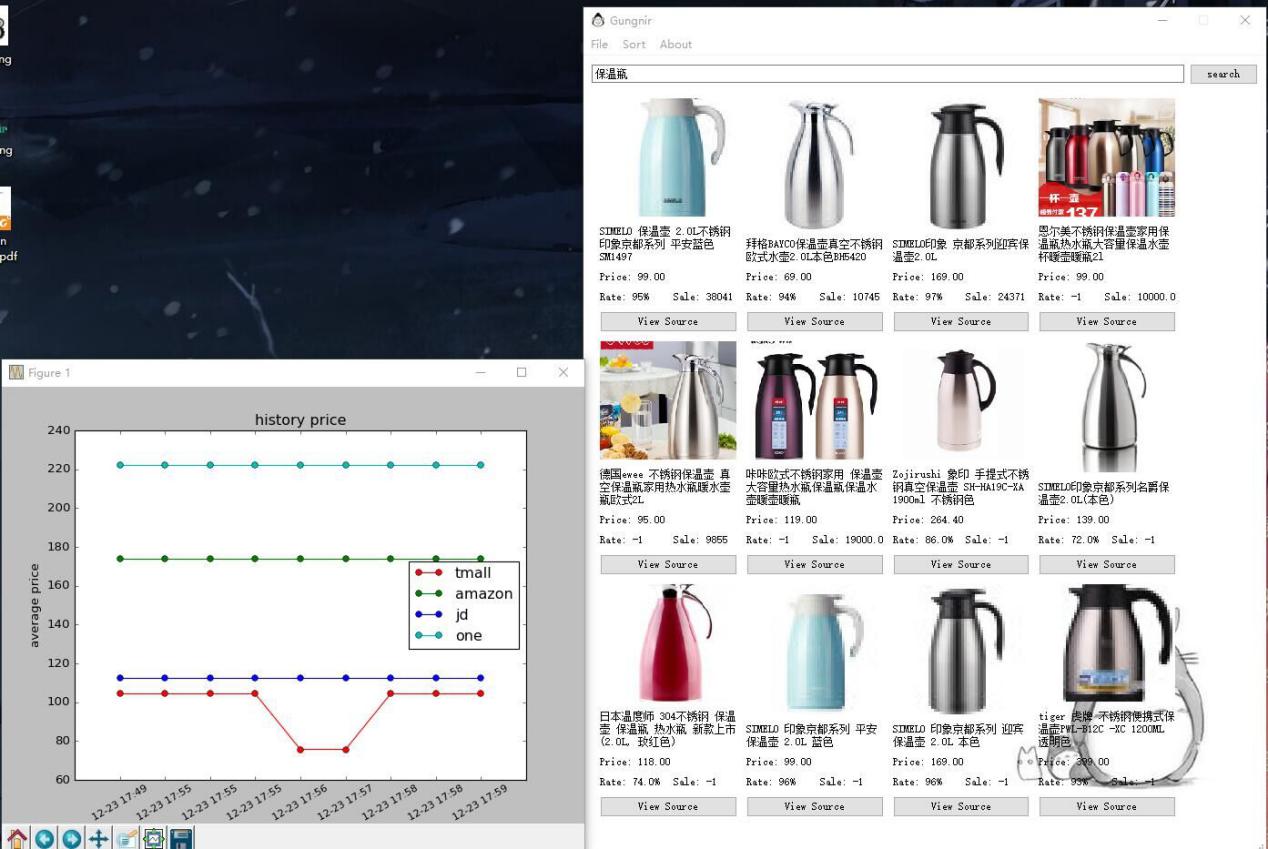
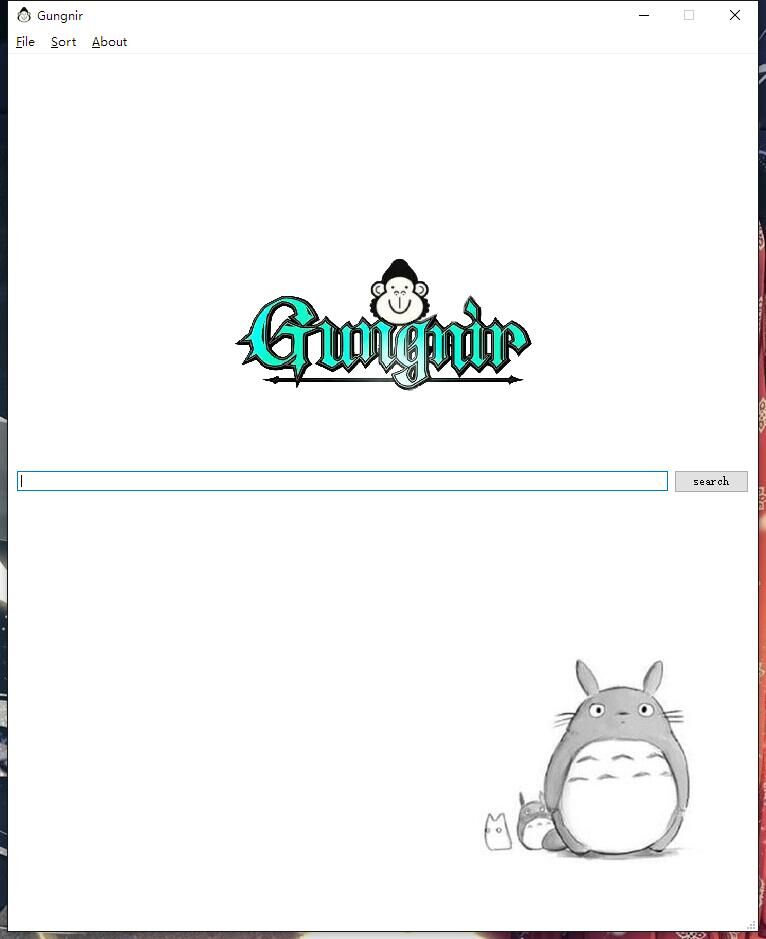
首页采用最简洁的logo+搜索框+菜单栏，使所有功能和特性一览无余，便于操作和使用。界面左上角是Gungnir比价软件的logo和名称，然后是菜单栏。菜单栏中功能如下：

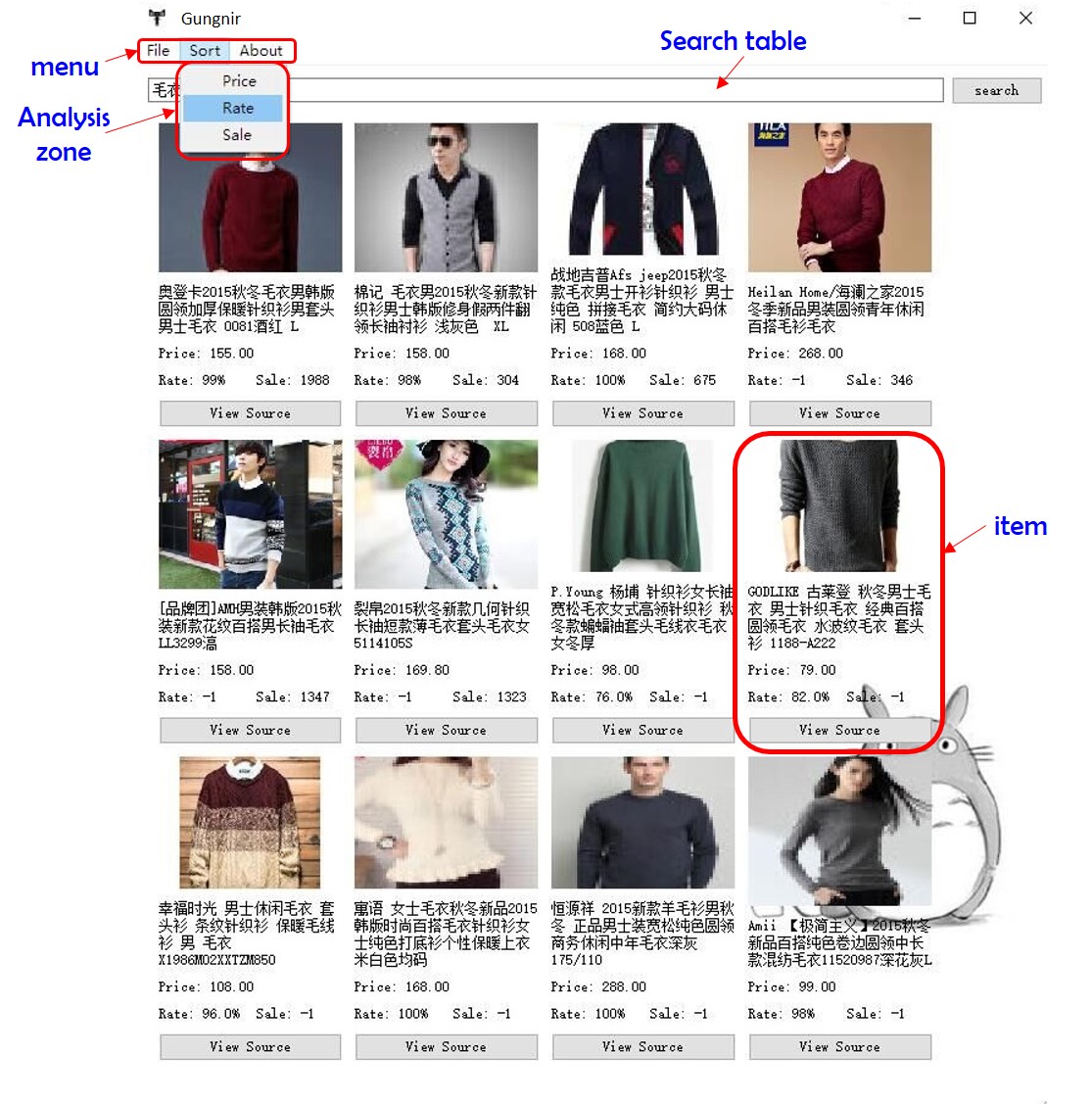
File：save（保存此次查询数据）；analyze（生成历史价格折线图）；quit（退出）

Sort：price（价格低到高排序）；rate（评价高到低排序）；sale（销量高到低排序）

Help：获取关于Gungnir软件的更多信息。

搜索框下方就是各个商品，包括图片、名称、价格、销量和评价五个信息。点击商品可以跳转至各商品原网页了解详情。





1. **代码说明**

* 核心代码

前端：

1. 通过自定义类函数实现。最核心的是Window类，其中包含\_\_init\_\_（构造函数），intiUI（初始化函数），menu（菜单函数），help（帮助函数），about（关于我们函数）以及button（按钮函数）。
2. \_\_inti\_\_函数基本实现了界面的框架。调用initUI函数，initUI函数包含设置窗口大小、GridLayout、grid 间距、背景图片、icon和名称，调用了menu和button函数。
3. QtItemLayout继承QGridLayout，整体排版一系列QWidget的组合，绑定多个QLabel，每个QLabel对应一个数据项。
4. 一个QtItemBox为多个QtItemLayout的组合，其整体作为一个QWidget，以便于被上层QWindow管理。一个QtItemBox默认由12个QtItemLayout为4\*3排布，内部通过一个QGridLayout的addLayout方法对QtItemLayout进行整体排版。

后端：

1. 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，再通过save\_img\_to\_file函数将图片存到本地。
2. Provider负责输出统一格式的json文件。
3. item\_get.py中，统一调用预先设计的get(name)接口，调用jd,amazon,tmall,one这四个模块获取json数据，再从文件读取
4. item\_show.py中，调用item\_get中的get\_items(name)方法，得到itemList，并通过itemList构造QtItemBox
5. gui.py中，捕捉SearchButton的事件，同时读取SearchTable(QLineEdit)的unicode串数据，并调用item\_show中的QtItemBox类构造函数以构造itemBox，并刷新显示
6. advance.py中，Saveprice函数负责保存价格，通过dump和load将新纪录加在到prices list中。Drawpic()负责读取数据，画折线图。

* 每个同学完成的代码部分

徐瑾卿：one.py / advance.py

樊泽坤：item\_get.py / item\_show.py / \_init.py / jd.py

计婷：gui.py / amazon.py

楼思琦：tmall.py

1. **个人贡献**

* 各自负责的工作：

1. 每人负责一个网站抓取商品信息,并将商品信息打包转换成统一的可输出的格式。（一号店：徐瑾卿；京东：樊泽坤；亚马逊：计婷；天猫：楼思琦）
2. 前后端交互及代码整合：樊泽坤；
3. GUI设计：计婷；
4. 高级功能实现（历史价格走势）：徐瑾卿；
5. 文档撰写：楼思琦；

* 该小组四位成员均为电子信息与电气工程学院软件工程专业
* 感想：

在本次大作业过程中，组员们分别学习了requests、BeautifulSoup、PYQT等第三方库的使用，以及熟悉了re的运用及正则表达式书写，python中unicode格式等知识点。同时写爬虫技术进一步提高，对网页信息的抓取有了深入的了解。

不仅如此，在项目的开发设计或者是后期整合中，我们都发现了项目管理和分工、项目架构的具体化的重要性，这些经验对于我们软件工程的学生来说都是非常宝贵的。而且我们以前都是使用C++/C语言写代码，初次尝试python，深刻体会到python相比于C++在开发效率上的优势和语法的自然性，以及不同语言之间各自的优劣点，为今后更为深入地学习编程提供了一定参考。