

购物比价系统设计文档

3110104862 王刚

一、总体阐述

比价系统其实就是一个简单的搜索引擎，需要大量的网络蜘蛛去采集电商网站的数据，然后汇总到数据中心，然后再推送给用户。

与此同时，我们也要注意，对商品应该有所区别，热门商品，用户关注的商品采集频率高，比如一小时采集一次；冷门商品一天甚至几天更新一次数据也可以。所以这里同样需要蜘蛛具备可以调节采集对象的接口。这样才能充分利用服务器资源。

二、功能需求

基本功能（必须实现）：

1. 提供一个网页，用户输入商品名称后可以查询到主流购物网站的价格，淘宝、亚马逊中国必须实现，其他自由选择。
2. 定时到各网站提取信息或者用户查询时实时到各网站提取信息的实现方案都可以。

高级功能（可选实现）：

1. 实现用户注册、登录功能，用户注册时需要填写必要的信息并验证，如用户名、密码要求在 6 字节以上，email 的格式验证，并保证用户名和 email 在系统中唯一。
2. 用户登录后可以设置感兴趣的商品，系统后台定期到各网站查询，保存在用户网页上
3. 实现一个 Android 或 iphone 客户端软件，实现感兴趣商品的低价推送

三、前端、后端功能

服务器端：

用户基本查询：

用户输入关键字选择相应的搜索方式，可以查询到符合自己要求的产品；

产品价格比较

用户通过查询结果，通过进行筛选获得产品价格比较的信息。

浏览产品详细信息

用户可以点击查询结果，访问店家获得具体商品信息。

管理员中心

通过搜集各大网站上的物品信息，展示给用户。对用户的信息进行管理。

在服务器端实现产品查询、比较、强行浏览，移动终端用户可以通过浏览

器来实现产品查询、比较、想请浏览。

时间特性要求：

要求将结果较快的返回给用户页面。

故障处理要求：

当用户在使用网站进行一系列操作的时候，能够给出一系列合理的提示信息，不能因为操作失误而导致系统错误，或者程序停止运行。

程序运行时，操作故障能够识别并提示，当故障排除后，程序恢复正常运行。

设计目标：

操作简单，易于使用，实用性强，可维护性好。

四、体系设计

搜索引擎一般由搜索模块、索引模块、检索模块以及用户接口四大部分组成。

4.1 搜索模块

1) 爬虫

爬虫即网络蜘蛛，是搜索引擎的搜索模块部分，它是一个可以自动抓取电商网站上物品价格等信息的程序。在搜集信息的过程中，要保证输出结果快，且定期更新结果。

2) 内容提取

需要有不同的插件管理程序对各种格式的网页采取不同的插件处理。搜索引擎的建立，是基于对文本文件对象处理的基础上的，所以在爬虫抓取各种网页后，要把所需要的物品价格等信息提取出来制成文本文件。

4.2 索引模块

索引模块读入信息文本文件信息，进行倒排索引。首先通过分词处理产生索引项，然后整合索引项生成倒排表。倒排表用于根据所印象查找相应的文档，在倒排表中，每个分词都被记录了含有该词的文档的起始地址，同时记录了含有该词的文档树。

4.3 检索模块

检索模块先获取用户查询的请求，然后在索引库里快速地检索出所需要的文档，并且评价查询和文档的相关度，对查询到的信息排序、聚类，最后在网页中生成搜索结果返回。

4.4 用户接口

用户接口用于用户输入查询关键词，显示返回结果，使用户和搜索引擎之间交互的窗口，是集用户请求与搜索引擎反馈为一体的网页。

五、模块划分

1. 比价信息的获取

获取电子商务网站商品信息主要有两种途径：

1) 通过电子商城提供的 api 来获取商品价格信息

现在主流的电商都提供了可以调用的 API 来供开发者调用查询（淘宝、亚马逊等），但是对于开发者而言 API 的调用往往有权限要求或者需要付费使用。

2) 通过爬虫爬取相应产品的价格信息

在服务器端编写网络爬虫定时爬取制定网站数据并更新数据库内容，或者实时根据使用者的输入来爬取商品信息，相比较而言，第一种方式的响应速度更快，但是数据的实时性不够好，并且需要数据库大容量的支持；第二种方式实时性更好，对数据容量要求也不高，但会造成反应比较慢的较差体验。

为取得最佳的用户体验，我设想采用的技术是对于常见商品搜索入库，并定期更新商品信息；而对于在数据库中没有匹配项目的商品，利用电商提供的 API 或者使用网络爬虫来爬取数据，并且在爬取数据之后存入本地数据库中，方便进行下一次查询。

2. 账户管理功能

该功能可进一步细分如下：

1) 用户注册

2) 比价信息定制

3) 密码找回功能

六、技术支持

搭建网站的结构：HTML 与 XHTML 组织网站的内容

美化与布局网站的样式表-CSS 让内容以更用户友好的形式展现出来

客户端脚本语言-JavaScript：在客户端实现一些动态效果、控制浏览器、数据的检测与验证

服务端脚本语言-ASP PHP JSP 等：通过编程在服务器端实现对内容的分析与操作，比如用户登录、购物车等。

数据库与 SQL 语句：数据库用于存储数据，保留用户与网站信息