

# 比价网站开发文档

BS 课程设计

王刚 | 3110104862 | 2014 年 6 月 27 日

# 浙江大学

## 本科实验报告

课程名称： BS 体系设计

---

姓 名： 王刚

---

学 院： 计算机学院

---

系：

---

专 业： 计算机科学与技术

---

学 号： 3110104862

---

指导教师： 胡小军

---

2014 年 6 月 26 日

# 浙江大学实验报告

课程名称： BS 体系结构设计 实验类型： 课程实践与设计类

实验项目名称： 购物比价系统的实现

学生姓名： 王刚 专业： 计算机科学与技术 学号： 3110104862

同组学生姓名： \_\_\_\_\_ 指导老师： 胡晓军

实验地点： 寝室 实验日期： 2014 年    月    日

## 目录

### 目录


1. 数据库设计 .....	3
2. 爬取网站设计 .....	4
2.1 爬取思路 .....	4
2.2 爬取方法实现 .....	4
2.2.1 淘宝网数据爬取 .....	4
2.2.2 亚马逊商品数据爬取 .....	7
3. 界面设计 .....	10
主页布局和风格如下： .....	10
登陆和注册管理 .....	11
搜索框设置 .....	12
个人中心 .....	13
比价结果显示 .....	14
前端界面设计说明 .....	15
4. 功能模块设计和实现 .....	15
4.1 用户注册和登陆 .....	15

4.2 商品信息搜索.....	15
4.3 按价格排序.....	17
4.4 加入个人收藏.....	17
5. 开发心得.....	18

## 1. 数据库设计

数据库中有两张表来记录用户的登陆和浏览信息，一张为 user 表，结构定义如下：

Uid	Username	Password	Regdate
-----	----------	----------	---------



Regdate 为用户注册时间，可以再用户中心中访问到用户注册的信息。

另一张表为 record 表，用于记录用户收藏商品的信息，结构定义如下：

Rid	Rname	Rsrc	Uid	Regdate
-----	-------	------	-----	---------



Uid 为收藏商品的用户的 key, rname 为收藏商品的描述, rsrc 为商品的购买链接, regdate 为用户收藏该商品的时间。

## 2. 爬取网站设计

### 2.1 爬取思路

爬取网站主要有两种途径, 一种是利用网站提供的接口 api 实现较轻松的数据获取, 另一种则是通过分析网页返回的源代码来抓取关键信息。

我在做爬取实现的时候, 首先尝试的思路是通过淘宝提供的开放环境 API 来获得商品数据, 后来发现该方法还需要淘宝的授权, 操作起来比较麻烦。

在询问了几个同学后, 我最终采用了 phpQuery 处理 DOM 文档内容。phpQuery 是基于 php 的服务器端开元项目, 它开元让开发人员轻松处理 DOM 文档内容, 比如获取新闻网站头条信息等。

### 2.2 爬取方法实现

#### 2.2.1 淘宝网数据爬取

我们定义的商品类格式如下:

```
class Item_class{
    public $item_name;//商品名称;
    public $item_price;//商品价格;
    public $item_pic_src;//图片链接;
    public $item_src;//购买链接;
    public $item_source;//来源于淘宝或者亚马逊或者其它网站
    /*
    public function show()      //用来输出商品信息
    {
        print "<p><a href=\"".$this->item_src."" target=\"._blank\".>
        <img src=\"".$this->item_pic_src.""></a></p>";
        print "<h2>".$this->item_price."</h2>";
        print "<h3><a href=\"".$this->item_src."" target=\"._blank\".>
        ".$this->item_name."</a></h3>";
    }
    */
    static function cmp($a,$b)
```

```

{
    if($a->item_price == $b->item_price)
        return 0;
    return $a->item_price > $b->item_price ? 1:-1;
}
public function show()
{
    echo "<span>";
        echo "<form name=\"myform\" method=\"POST\"
action=\"addRecord.php\" target= \"_blank\">";
        echo "<input type=\"image\" src=\"img/cart.png\" width=\"40\" height=\"40\"
value=\"\".$this->item_name.\"!\".$this->item_src.\"\" name=\"info\"
onclick=\"document.myform.submit()\">";
        echo "</form>";
            echo "<a href=\".$this->item_src.\" target= \"_blank\">";
                echo "<img src = \"\".$this->item_pic_src.\">";

            echo "</a>";
        echo "</span>";
        echo "<h3><a href=\".$this->item_src.\" target= \"_blank\">";
        if($this->item_source == "T")
            echo $this->item_price." 淘宝网";
        else
            echo $this->item_price." 亚马逊";

        echo "</a></h3>";
        echo "<p>".$this->item_name."</p>";
        echo "</article>";
    }
}
?>

```

淘宝网数据编码格式是 gb2312k 编码，亚马逊网站的编码格式为 utf-8 编码格式；为保持最终网页展示效果的一致性，在对淘宝网数据进行爬取时首先对网页进行编码格式的处理。

```

function Transcoding($in_filename)
{
    $out_filename = 'utf8.php';//淘宝网编码格式进行转换

```

```

        $contents = file_get_contents($in_filename);
        $contents = iconv('GBK', 'UTF-8', $contents);
        file_put_contents($out_filename, $contents);
        return $out_filename;
    }

```

获取淘宝网页商品数据的函数：

#### 1. 首先分析淘宝网商品信息页

```

function getTaobao()
{
    $item_array = array();
    if(isset($_POST["name"]))
    {
        $url = 'http://s.taobao.com/search?q='.$_POST["name"];
        echo "<br/>";
        $url = iconv('UTF-8','GBK',$url);
        $phpQuery::newDocumentFile(Transcoding($url));    // 抓取网址
        $arr=pq(".item-box.st-itembox");

        $count = 0;
        foreach($arr as $li){
            $item_temp = new Item_class;
            $item_temp->item_name = pq($li)->find('h3.summary')->text();
            $item_temp->item_price = " ".pq($li)->find('.col.price')->text();
            $item_temp->item_pic_src = pq($li)->find('img')->attr('data-ks-
lazyload');
            $item_temp->item_src = pq($li)->find("a")->attr('href');
            $item_temp->item_source = "T";

            $item_array[] = $item_temp;// 添加商品到数组中
            $count ++;
            if($count > 24)// 爬取商品数量
                break;
        }
        /*
        $url = 'http://s.taobao.com/search?q='.$_POST["name"].'&s=44';// 抓取
        第二页的内容
    */
}

```

```

        echo "<br/>";
        $url = iconv('UTF-8','GBK',$url);
        phpQuery::newDocumentFile(Transcoding($url));    // 抓取网址
        $arr=pq(".item-box.st-itembox");
        foreach($arr as $li){
            $item_temp = new Item_class;
            $item_temp->item_name = pq($li)->find('h3.summary')->text();
            $item_temp->item_price = " ".pq($li)->find('.col.price')->text();
            $item_temp->item_pic_src = pq($li)->find('img')->attr('data-ks-
lazyload');

            $item_temp->item_src = pq($li)->find("a")->attr('href');
            $item_temp->item_source = "T";

            $item_array[] = $item_temp;//添加商品到数组中
        }
    }
    return $item_array;
}

```

### 2.2.2 亚马逊商品数据爬取

GetAmazon()函数

1. 首先获取网页页数信息
2. 一页页爬取商品信息页，并存入数组
3. 返回给上一层使用

```

function getAmazon()
{
    $item_array = array();
    if(isset($_POST['name']))
    {
        //echo$_POST["name"];
        $query_name = $_POST['name'];
        $page_index=1;
        $query_index=0;
        $query_condition="result_";
    }
}

```



```

        $url =
"http://www.amazon.cn/s/&page=".(string)$page_index."&keywords=".$query_name;
        phpQuery::newDocumentFile($url);
        $page_count = intval(pq(".pagenDisabled")->text());

        if($page_count > 2)
        {
            for(;$page_index <= 2;$page_index++)
            {
                $url =
"http://www.amazon.cn/s/&page=".$page_index."&keywords=".$query_name;
                phpQuery::newDocumentFile($url);
                $result_page = pq("#resultsCol");
                for($temp = 0; $temp < 24; $temp++)
                {
                    $item_temp = new Item_class;
                    $temp_id_name =
"#".$query_condition.(string)$query_index;
                    $item_all_info =
pq($result_page)->find($temp_id_name);

                    $item_temp->item_name =
pq($item_all_info)->find('h3')->text();
                    $item_temp->item_price =
pq($item_all_info)->find('.bld.lrg.red')->text();
                    $item_temp->item_pic_src =
pq($item_all_info)->find('img')->attr('src');
                    $item_temp->item_src =
pq($item_all_info)->find('h3')->find('a')->attr('href');
                    $item_temp->item_source = "A";
                    $item_array[] = $item_temp;//添加商品到数组中
                    $query_index++;
                }
            }
        }
        else
        {
            for(;$page_index <= $page_count - 1;$page_index++)
            {
                $url =
"http://www.amazon.cn/s/&page=".$page_index."&keywords=".$query_name;

```

```

        phpQuery::newDocumentFile($url);
        $result_page = pq("#resultsCol");
        for($temp = 0; $temp < 24; $temp ++ )
        {
            $item_temp = new Item_class;
            $temp_id_name =
"#".$query_condition.(string)$query_index;
            $item_all_info=
pq($result_page)->find($temp_id_name);

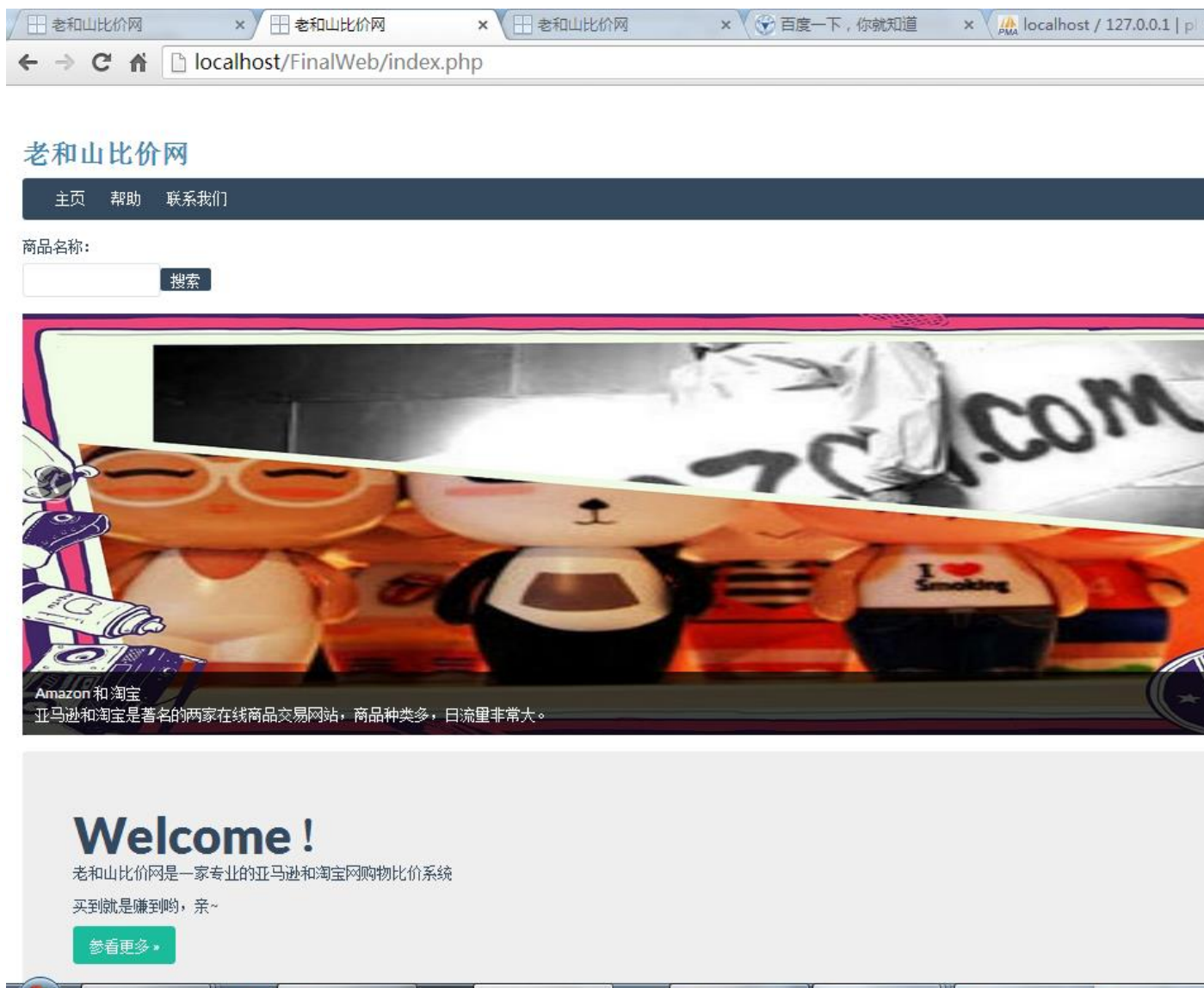
            $item_temp->item_name =
pq($item_all_info)->find('h3')->text();
            $item_temp->item_price =
pq($item_all_info)->find('.bld.lrg.red')->text();
            $item_temp->item_pic_src =
pq($item_all_info)->find('img')->attr('src');
            $item_temp->item_src =
pq($item_all_info)->find('h3')->find('a')->attr('href');
            $item_temp->item_source = "A";
            $item_array[] = $item_temp;//添加商品到数组中
            $query_index ++;
        }
    }
}

return $item_array;

```

### 3. 界面设计

主页布局 and 风格如下：



## 登陆和注册管理

七价网

[帮助](#) [联系我们](#)

登陆/注册

登陆

注册

登陆界面：

# 登陆

用户名

密码

登陆

注册界面：

## 用户注册

用户名

密码

请再次输入密码：

注册

## 搜索框设置

### 老和山比价网

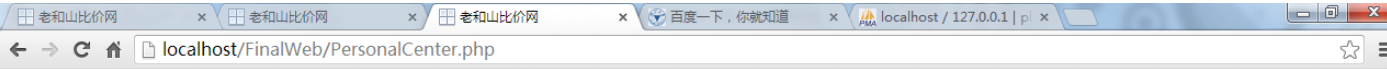
[主页](#) [帮助](#) [联系我们](#)

商品名称：

搜索



个人中心



用户中心

[首页](#)[联系我们](#)[帮助](#)[个人中心](#)

欢迎您，Terry1992!

[个人信息](#)

编号	商品名	收藏时间	购买链接
1	精谛JD-258F多功能料理机（绞肉、豆浆、果汁、果酱、奶昔、刨冰、面膜、干粉等）精谛营养调理机 食物搅拌机 婴儿辅食机 电动果汁机 绞肉机（菜馅、肉馅、包子饺子必备）精谛 /td>	1403878576	<a href="#">详情</a>
2	邻家小厨上海小笼包冷冻方便食品上海特色小吃小笼包子 30个/包 /td>	1403886251	<a href="#">详情</a>

localhost/FinalWeb/PersonalCenter.php#accordion-element-554968

比价结果显示

商品名称:

搜索

价格排序



¥56.00 淘宝网

邻家小厨上海小笼包冷冻方便食品上海特色小吃小笼包子 30个/包



## 前端界面设计说明

用户交互界面的设计和实现，主要利用了 bootstrap 和 layout 的可视化设计软件。

## 4. 功能模块设计和实现

### 4.1 用户注册和登陆

主要原理是，通过 php 对数据库进行访问，实现对数据库中表的查询、插入与修改。

登录：

php 判断输入的用户名或密码是否合法，是则连接数据库，查询输入的用户名是否存在，密码是否相同，是则登录成功！

注册：

所有的注册信息都要判断是否合法，是则连接数据库，检查用户输入的账号是否已被使用，如果没被使用，就可以把信息插入数据库，然后提示成功、跳转页面；注册时通过结合正则表达式对输入的内容进行匹配，若匹配成功，则证明输入格式正确，否则需要重新输入。

### 4.2 商品信息搜索

用户登陆：

```
<html>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
</html>
<?php
session_start();

$conn=mysqli_connect("localhost","test","123456","bs") or die("数据库服务器连接错误：".mysql_error());
//注销登录
if(isset($_GET['action']))
{
```



```

unset($_SESSION['uid']);
unset($_SESSION['username']);
session_destroy();
echo '已成功注销登录，<a href="index.php">点此返回主页</a><br/>';
echo '两秒后页面将自动跳转到首页，如若没有跳转，请点击上文中的链接!';
$url = "index.php";
Header("Refresh:2;url=$url");
exit;
}
//登陆
if(isset($_POST['submit'])){exit('访问不成功');}
$username=htmlspecialchars($_POST['username']);
$password=MD5($_POST['password']);
//包含数据库连接文件
//include('conn.php');
//检测用户名及密码是否正确
$check_query=mysqli_query($conn,"select uid from user where
username='$username' limit 1");
if($result = mysqli_fetch_array($check_query))
{
    //登陆成功
    $_SESSION['username']=$username;
    $_SESSION['uid']=$result['uid'];
    //echo $_SESSION['uid'];
    $check_pass=mysqli_query($conn,"select password from user where
username='$username' limit 1");
    $res = mysqli_fetch_array($check_pass);
    if(!array_diff(array($res['password']),array($password)))
    {echo '用户名和密码正确';echo '<br \>';}
    else
    {echo '密码错误<a href="javascript:history.back(-1);">返回</a>重新输入';exit;}

    echo $username,'欢迎进入<a href="PersonalCenter.php">用户中心</a></br>';
    echo "两秒后自动进入，如若没有进入用户中心请单击上行链接";
    $url = "PersonalCenter.php";
    Header("Refresh:2;url=$url");
    exit;
}else{

```

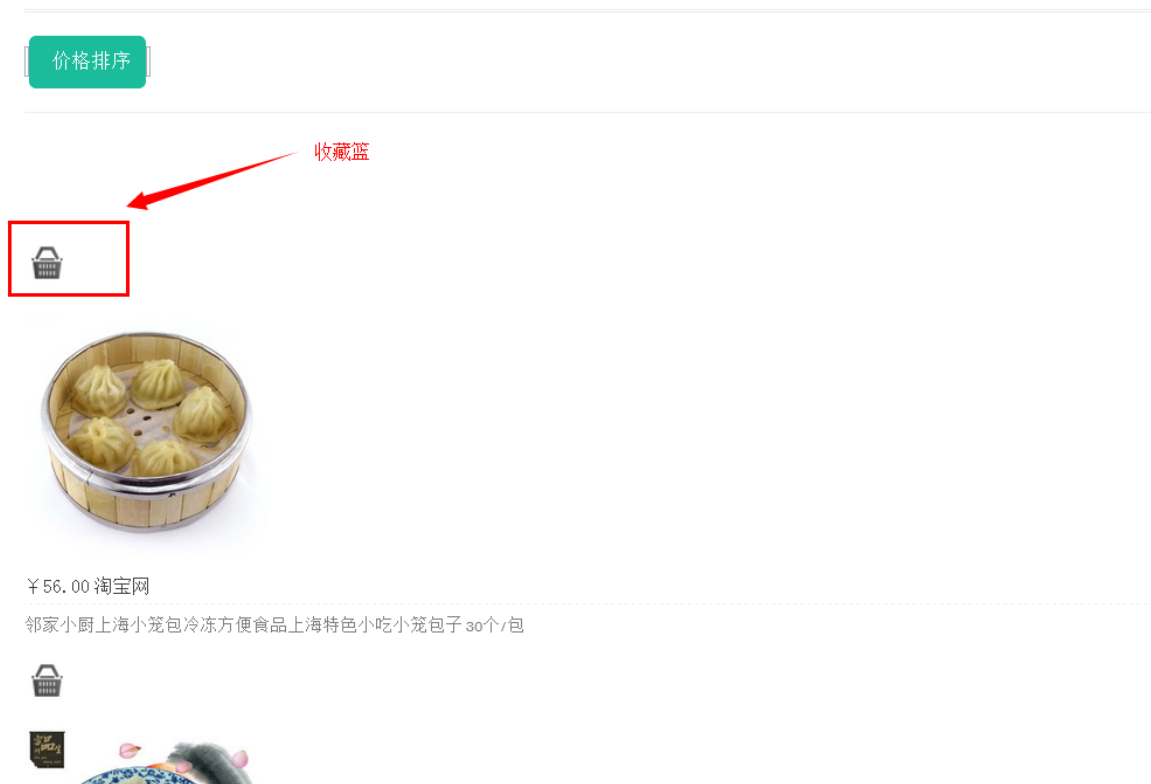
```
exit('登录失败,用户不存在~<a href="javascript:history.back(-1);">返回</a>重试');  
}  
?>
```

### 4.3 按价格排序

点击按价格排序后，提交搜索商品名，得到

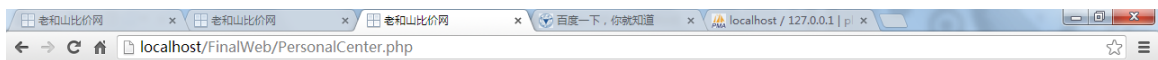
### 4.4 加入个人收藏

如下图为商品搜索结果的显示页：



在商品图片的左上角有一个小篮子图标，点击该图标，商品的信息会 post 到 addRecord.php 页面，该页面会将商品信息存储到与有对应关系的表中。

在个人中心可以看到：



## 用户中心

首页 联系我们 帮助

个人中心

欢迎您, Terry1992!

[个人信息](#)

编号	商品名	收藏时间	购买链接
1	精谛JD-258F多功能料理机(绞肉、豆浆、果汁、果酱、奶昔、刨冰、面膜、干粉等) 精谛营养调理机 食物搅拌机 婴儿辅食机 电动果汁机 绞肉机(菜馅、肉馅、包子饺子必备) 精谛 /td>	1403878576	<a href="#">详情</a>
2	邻家小厨上海小笼包冷冻方便食品上海特色小吃小笼包子 30个/包 /td>	1403886251	<a href="#">详情</a>

## 5. 开发心得

第一次独自开发从未接触过的网站项目，走了很多弯路，但最终总算完成了大作业。通过该次网站开发，我对 BS 软件体系架构有了更深的理解。

王刚

2014-06-27