

商品评论抓取工具使用

一只尼玛

<https://github.com/hunterhug/taobaocomment>



2016-1-10

目录

前言.....	1
使用说明.....	2
程序结果.....	6
来源分析.....	10

电商商品评论抓取工具使用说明

前言

满足大多数客户的需求，是一个电商经营者必须考虑的问题。客户的意见很重要！电子商务平台不仅给我们提供了商品展示的平台，同时也提供了用户的各种评论数据。利用这些文本数据，我们能够得知用户的感受，商品的优点和缺陷。本课题要达到的效果是实现一个自动化的评论抓取工具，同时提供一种分析评论的思路。

宝贝详情

累计评论 14

成交记录 4145

专享服务

宝贝评论

售后服务评论 110

问大家

全部

图片 (5)

追评 (4)

好评 (14)

中评 (0)

差评 (0)

☒ 有内容的评价

推荐排序



q***e



颜色真的好看，料子还不错，尤其是大翻领显得脸小，非常显瘦。扣子处有一点画笔痕迹，不过看不出，无所谓了！五分好评，再他们价买了好几件衣服了，一如既往的支持！大家看图呗。



2015年12月19日 19:31

颜色分类：紫粉色 尺码：S

有用 (6)

提问



j***7 (匿名)



大衣挺厚的，个人觉得零度以上都没有问题，版型佳，冬天里面穿的多，这个穿外面倒不会显得臃肿了，高个子穿效果应该更好吧，发货能快一点就完美了！



2015年12月21日 13:38

颜色分类：紫粉色 尺码：S

有用 (4)

提问



7***1 (匿名)



质量不错，颜色比较好看的，宽松版型，挺厚实的，关键是我穿上很好看，而且很显肤色哦 系个围巾刚刚好 赞



2015年12月19日 18:38

颜色分类：紫粉色 尺码：L

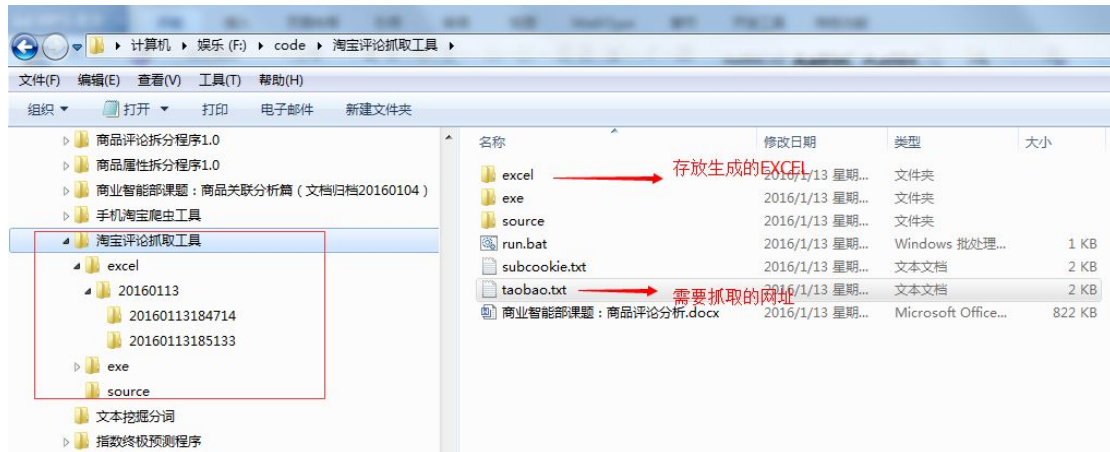
有用 (5)

提问

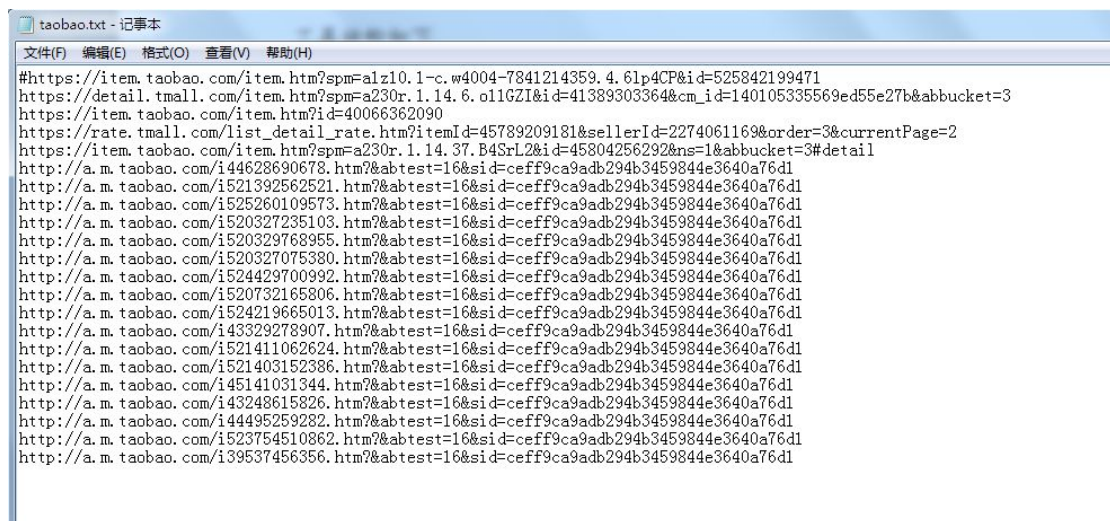
1

使用说明

工具结构如下：



首先在 taobao.txt 中按行写入网址：



网址格式有如下几种（#开头的网址直接被忽视）：

<https://item.taobao.com/item.htm?spm=a1z10.1-c.w4004-7841214359.4.6lp4CP&id=525842199471>

https://detail.tmall.com/item.htm?spm=a230r.1.14.6.o11GZl&id=41389303364&cm_id=14010533556&ed55e27b&abbucket=3

<https://item.taobao.com/item.htm?id=40066362090>

https://rate.tmall.com/list_detail_rate.htm?itemId=45789209181&sellerId=2274061169&order=3¤tPage=2

<https://item.taobao.com/item.htm?spm=a230r.1.14.37.B4Srl2&id=45804256292&ns=1&abbucket=3#detail>

<http://a.m.taobao.com/i44628690678.htm?&abtest=16&sid=ceff9ca9adb294b3459844e3640a76d1>

一般来说，直接从浏览器复制下来的商品链接可以用~~

然后点击 run.bat 或 exe 文件夹中后缀为 exe 的文件。

名称	修改日期	类型	大小
excel	2016/1/13 星期...	文件夹	
exe	2016/1/13 星期...	文件夹	
source	2016/1/13 星期...	文件夹	
run.bat	2016/1/13 星期...	Windows 批处理...	1 KB
subcookie.txt	2016/1/13 星期...	文本文档	2 KB
taobao.txt	2016/1/13 星期...	文本文档	2 KB
商业智能部课题：商品评论分析.docx	2016/1/13 星期...	Microsoft Office...	822 KB

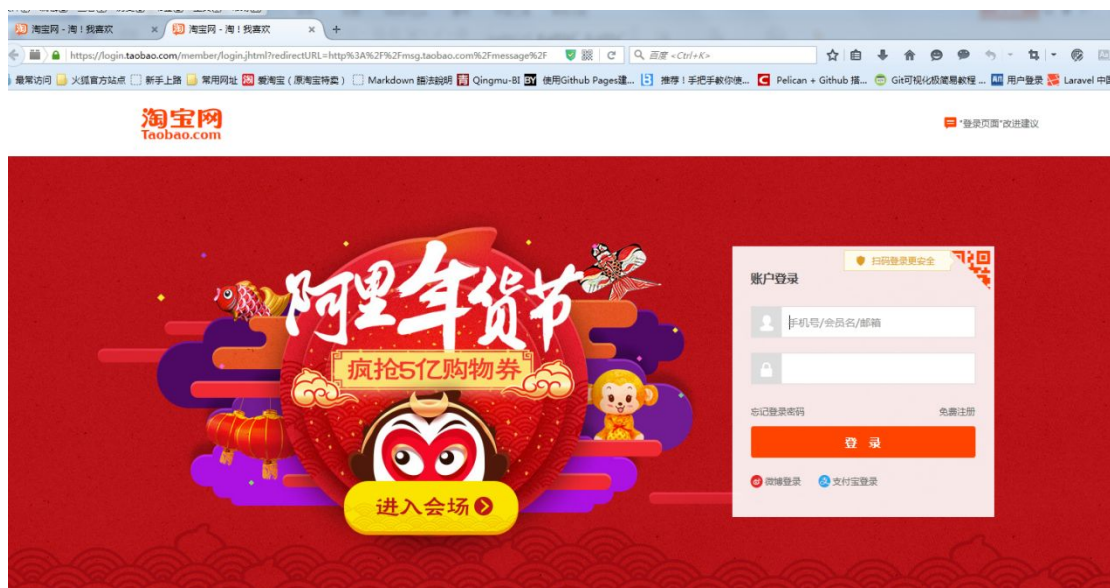
_bz2.pyd	2015/2/24 星期...	PYD 文件	53 KB
_decimal.pyd	2015/2/24 星期...	PYD 文件	178 KB
_elementtree.pyd	2015/2/24 星期...	PYD 文件	129 KB
_hashlib.pyd	2015/2/24 星期...	PYD 文件	765 KB
_lzma.pyd	2015/2/24 星期...	PYD 文件	134 KB
_socket.pyd	2015/2/24 星期...	PYD 文件	46 KB
_ssl.pyd	2015/2/24 星期...	PYD 文件	1,185 KB
cookie.txt	2016/1/13 星期...	文本文档	1 KB
library.zip	2016/1/13 星期...	WinRAR ZIP 压缩...	1,897 KB
PIL_imaging.pyd	2015/12/24 星期...	PYD 文件	947 KB
pyexpat.pyd	2015/2/24 星期...	PYD 文件	130 KB
python34.dll	2015/2/24 星期...	应用程序扩展	2,677 KB
select.pyd	2015/2/24 星期...	PYD 文件	10 KB
taocomment.exe	2015/11/25 星期...	应用程序	1,235 KB
unicodedata.pyd	2015/2/24 星期...	PYD 文件	741 KB

出现以下画面后，按指引操作



如果中途出错，请检查网址链接是否有错。

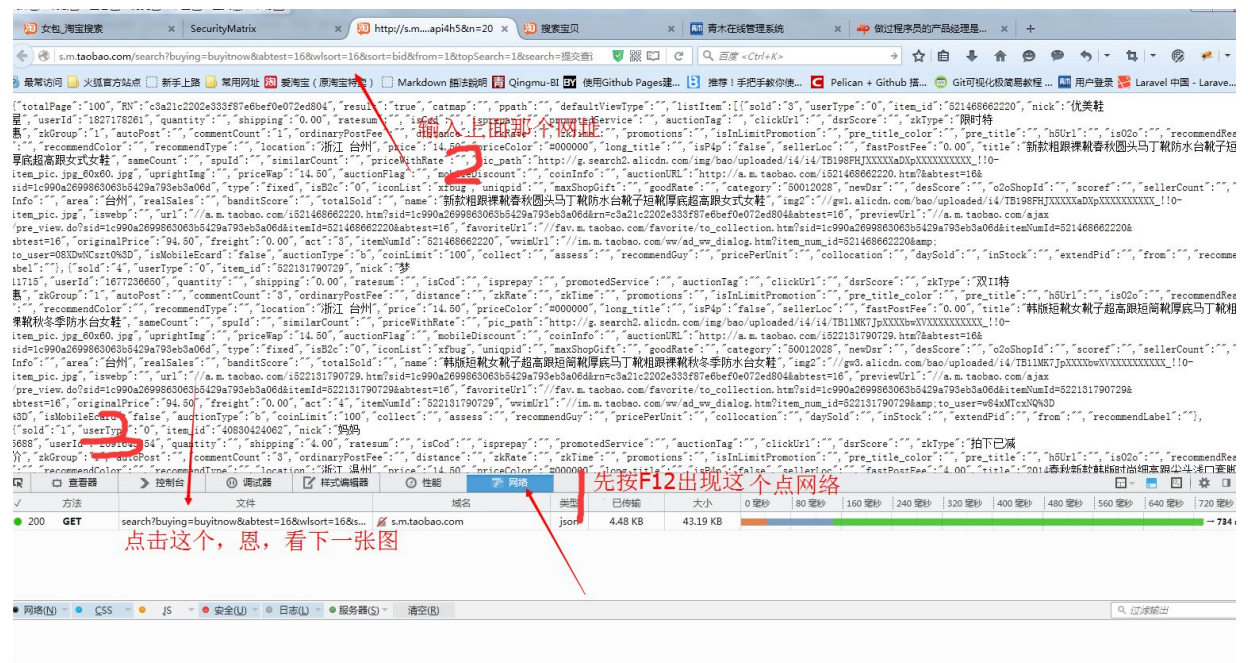
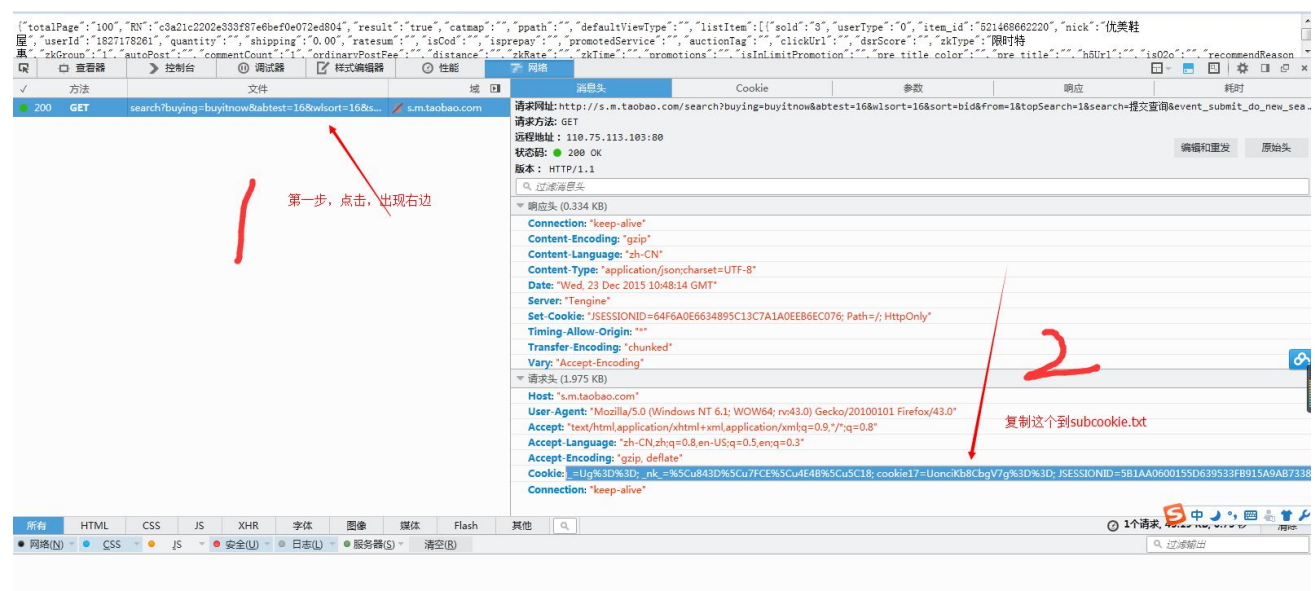
或者先登陆 <https://login.taobao.com>，输入账户密码：



登陆后在浏览器上按 F12，点击网络，然后在地址栏输入：

<https://rate.taobao.com/feedRateList.htm?auctionNumId=40066362090&userNumId=85356923¤tPageNum=5&showContent=1>

如下图，复制 Cookie。

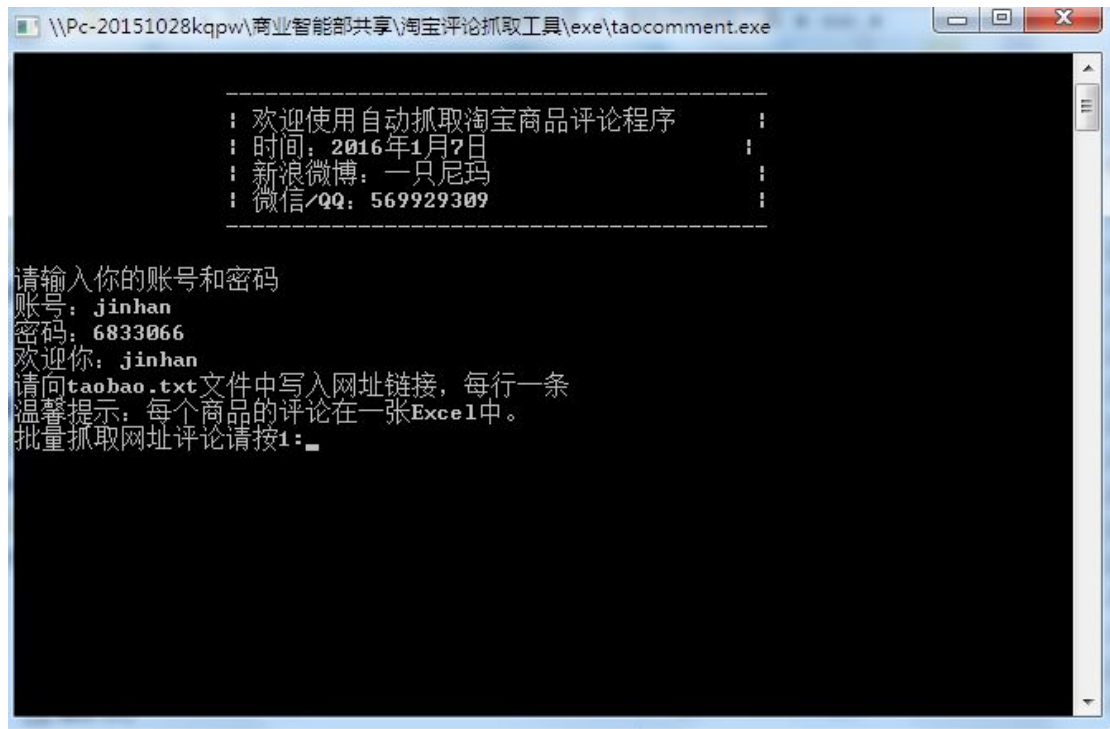


将复制的 Cookie 粘贴到 subcookie.txt 文件中。

就这么简单，这个文件是很重要的，一般有限期一个星期，突破淘宝反爬虫就靠它了。

程序结果

输入密码和账号



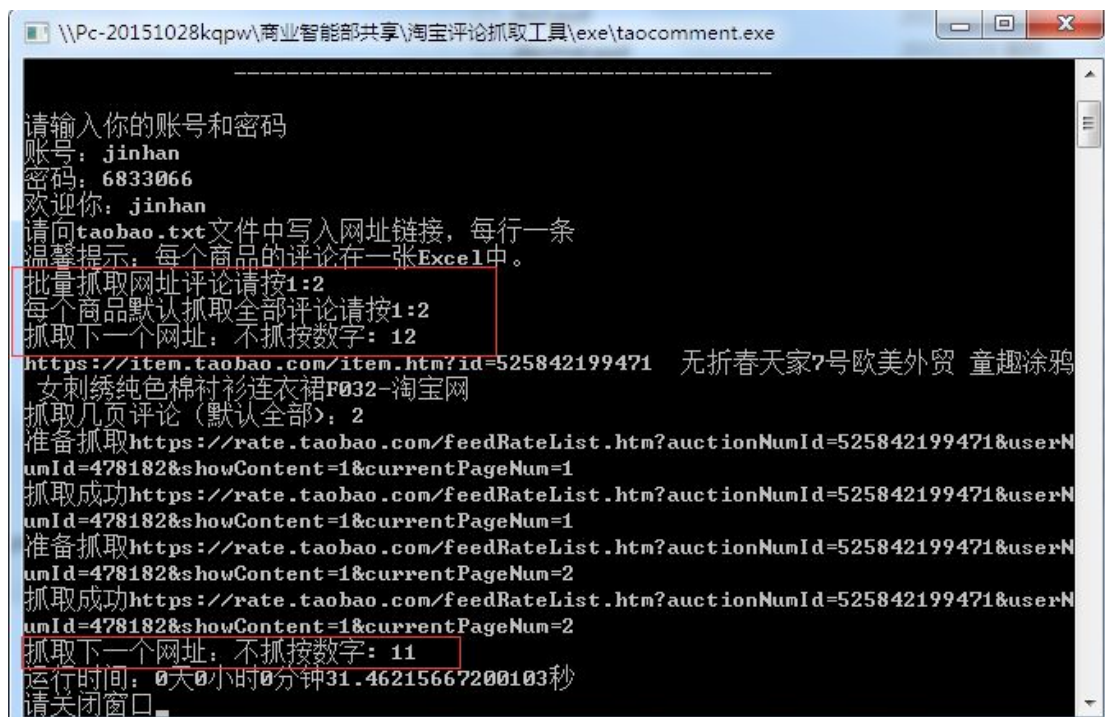
```
\\Pc-20151028kqpw\商业智能部共享\淘宝评论抓取工具\exe\taocomment.exe
: 欢迎使用自动抓取淘宝商品评论程序 :
: 时间: 2016年1月7日 :
: 新浪微博: 一只尼玛 :
: 微信/QQ: 569929309 :
: ~~~~~:
请输入你的账号和密码
账号: jinhan
密码: 6833066
欢迎你: jinhan
请问taobao.txt文件中写入网址链接, 每行一条
温馨提示: 每个商品的评论在一张Excel中。
批量抓取网址评论请按1: _
```

选择[批量抓取网址评论](#)按 1, 则不提示抓取下一个网址

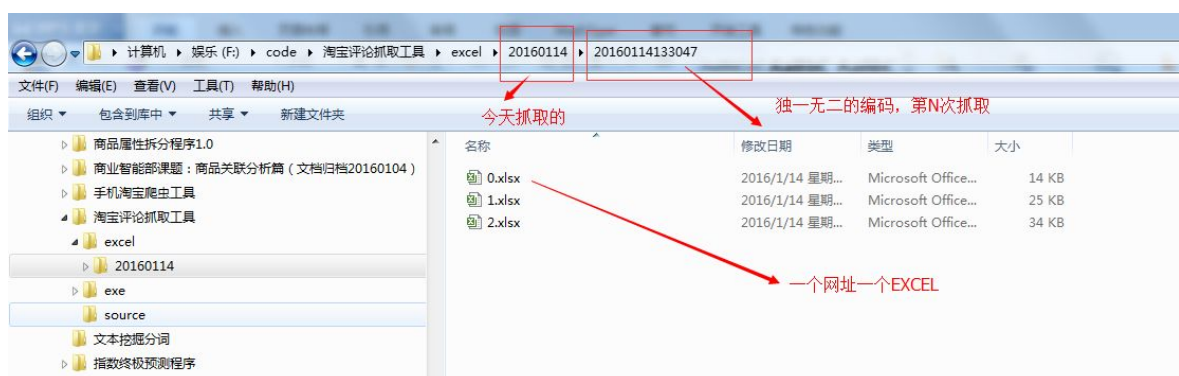


每个商品默认抓取全部评论按 1，则一次性抓取所有评论。按其他键如 2，则提示抓取几页评论（默认），输入数字即可。

选择批量抓取网址评论按 2，则提示抓取下一个网址



生成的 Excel



来源分析

淘宝

商品最短链接：<https://item.taobao.com/item.htm?id=40066362090>

其中 id 为商品唯一编号。



商品评论最短链接：

<https://rate.taobao.com/feedRateList.htm?auctionNumId=40066362090&userNumId=85356923¤tPageNum=5&showContent=1>

其中 auctionNumId 为商品 id，userNumId 是知名 id，currentPageNum 是评论所在页数，pageSize 为每页评论数，默认不可改。页数不限制，超页时数据为空。

知名 id 抓取首页后可以得到，拼接后可以抓取评论。

<meta name="microscope-data"

content="pageId=205517766;prototypeId=2;siteId=7;shopId=36736293;useri

d=8535692

<meta name="data-spm" content="2013"/>

<meta name="microscope-data" content="pageId=205517766;prototypeId=2;siteId=7;shopId=36736293;userid=85356923;">

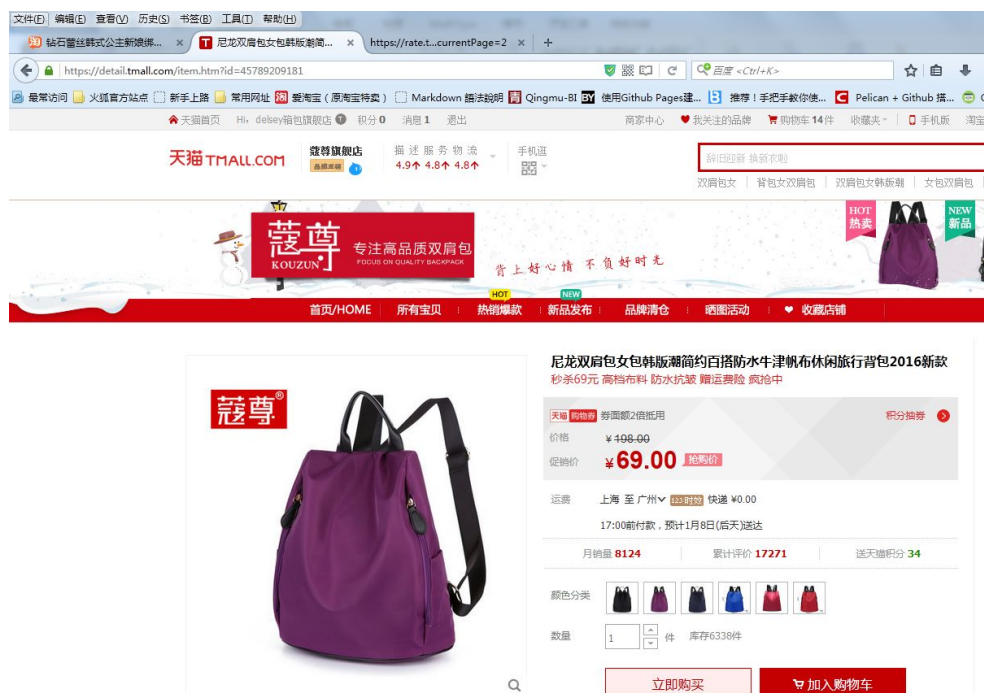
<title>钻石蕾丝韩式公主新娘绑带一字肩婚纱礼服2015冬季新款定制3267-淘宝网</title>

<link rel="shortcut icon" href="//img.alicdn.com/favicon.ico" type="image/x-icon" /><link rel="stylesheet" href="//g

天猫

商品最短链接：<https://detail.tmall.com/item.htm?id=45789209181>

其中 id 为商品唯一编号。



商品评论最短链接：

https://rate.tmall.com/list_detail_rate.htm?itemId=45789209181&sellerId=2274061169&order=3¤tPage=2

其中 itemId 为商品 id，sellerId 为不知名 id，currentPage 为当前评论页数，其他不可改。页数只能到 99 页

不知名 id 抓取首页后可以得到，拼接后可以抓取评论。

<meta name="microscope-data"

content="pageId=1178302750;prototypeId=2;siteId=2; shopId=114139460;

userid=2274061169;">

```
<meta name="spm-id" content="a220o" />
  <meta name="referrer" content="always">
    <meta name="ahot-aplus" content="1" />
  <meta name="microscope-data" content="pageId=1178302750;prototypeId=2;siteId=2; shopId=114139460; userid=2274061169;">
```