

第一阶段任务分配

第一阶段项目计划：

一、任务规划：

目标描述：

目标网站：

目标数据：

分工：

任务要求：

商品数据要求：

二、前端界面（LC的任务）

三、数据库说明：

四、Github仓库

五、云笔记协作

一、任务规划：

目标描述：

到中期检查前，爬取10000+条关于京东美妆商品的数据

目标网站：

京东商城→美妆个护（地址：<https://channel.jd.com/beauty.html>）

目标数据：

爬取“美妆个护”→“香水彩妆”→三个大类→16个主题下的所有商品数据，如下图：



爬虫任务分工：

除LC外的每个人需要负责抓取4个主题下的所有商品的数据。具体分工见下表：

| 主题 | 负责人 | 主题 | 负责人 |
|-------|-----|--------|-----|
| 女士香 | 张敏华 | 口红 | 刘彩君 |
| 套装 | 张敏华 | 底妆 | 刘彩君 |
| 浓香水 | 张敏华 | 眉眼 | 刘彩君 |
| 香水EDP | 张敏华 | BB霜/CC | 刘彩君 |
| 香精 | 何铭宜 | 粉底 | 林鹏珊 |
| 香体乳 | 何铭宜 | 眼线 | 林鹏珊 |
| 古龙水 | 何铭宜 | 睫毛膏 | 林鹏珊 |
| 固体香水 | 何铭宜 | 腮红 | 林鹏珊 |

任务要求：

要求每位负责人第一周抓取200+条数据，并把爬虫的python脚本上传到Git的项目仓库地址（见第四点），每周六早上10点开会汇报进度。

商品数据要求：

以下面一款Dior口红为例，说明下每件商品要爬取到哪些数据：

1：商品的链接，比如：<https://item.jd.com/1458843.html>

2：商品的图片的链接，比如下面这张图片的链

接：<https://img13.360buyimg.com/n1/jfs/t7714/311/1065518825/67942/5b3e0544/599a45ccNd4aca622.jpg>



PS: (至少要抓取一张图片, 至多抓取5张图片)

3、商品级别, 比如: 迪奥唇蜜

一级大类: 美妆个护

二级大类: 香水彩妆

三级大类: 唇部

美妆个护 > 香水彩妆 > 唇部 > 迪奥 (Dior) > 迪奥唇蜜

4: 商品全称 (or 简述)、价格: 比如:

京东超市 克丽丝汀迪奥DIOR魅惑润唇蜜SPF10 3.5g (又名: 克丽丝汀迪奥魅惑润唇膏 口红 (焕彩) 3.5g)

京东价 **¥269.00** 降价通知

优惠券 **满1000减100**

促销 **限制** 此价格不与套装优惠同时享受

累计评价
4万+

增值业务 **礼品包装**

配送至 北京朝阳区管庄 ∨ 有货

支持 99元免基础运费(20kg内) | 货到付款 | 京准达 | 211限时达 ∨

由 京东 发货, 并提供售后服务。11:10前下单, 预计今天(09月28日)送达。

重 里 0.04kg

选择颜色



白条分期

不分期

¥ 91.01×3期

¥ 46.18×6期

¥ 23.76×12期

¥ 12.55×24期



5: 商品介绍: 需要抓取的有: 品牌、商品名称、商品编号、商品毛重、商品产地、妆效、分类、功效、颜色

| 商品介绍 | 规格与包装 | 售后保障 | 商品评价(4万+) | 本店好评商品 | 加入购物车 |
|------------------------------|-------|--------------|-----------|-------------|-----------|
| 品牌：迪奥 (Dior) | | | | | |
| 商品名称：迪奥唇蜜 | | 商品编号：1458843 | | 商品毛重：35.00g | 商品产地：法国 |
| 妆效：自然 | | 分类：唇膏/口红 | | 功效：丰唇 | 颜色：橘色/橘红色 |
| 更多参数>> | | | | | |

6、商品规格与包装：需要抓取的有：保质期、适用人群

| 商品介绍 | 规格与包装 | 售后保障 | 商品评价(4万+) | 本店好评商品 | 加入购物车 |
|------|-------|-----------|-----------|--------|-------|
| 主体 | | | | | |
| 品牌 | | 迪奥 (Dior) | | | |
| 产地 | | 法国 | | | |
| 保质期 | | 3年 | | | |
| 类别 | | 唇蜜 | | | |
| 颜色 | | 珊瑚 (橘色) | | | |
| 适用人群 | | 通用 | | | |

7、商品评价：需要抓取的是：商品评价总数、好评度

| 商品介绍 | 规格与包装 | 售后保障 | 商品评价(4万+) | 本店好评商品 | 加入购物车 |
|---|-------|------|-----------|--------|-------|
| 商品评价 | | | | | |
| 好评度 | | | | | |
| 96% | | | | | |
| 颜色不错(144) 是正品(136) 品牌好(118) 颜色正(114) 东西不错(113) 滋润效果好(112) 润唇效果好(112) 唇膏很好(108) 物流速度快(102) 很好用(91) | | | | | |
| ...<li data-tab="trigger" data-offset="38" data-anchor="#comment" clstag="shangpin keycount product shangpinpingjia_2" class="current"> == \$0"商品评价"<s>(4万+)</s> | | | | | |

二、前端界面（LC的任务）

除了爬取数据，我们还需要尽快敲出导购系统的前端页面，该任务交给卢程负责，任务要求有3点：

- 1、界面尽可能和原型界面保持一致
- 2、一周内需要给出用户登录页面和系统首页
- 3、需要把代码（主要是html）上传到git，保持每天更新代码

三、数据库说明：

我已经在阿里云服务器上新建一个数据库，名为beautyGirls_database，可以远程连接该数据库，连接信息如下图：

主机IP地址：39.108.185.66，用户是root，密码是1234，权限是ALL.



里面已经有了一张商品表，名为**jd_products**，大家要做的事情是：爬取数据后，将数据导入到远程该表中。

数据库商品表格结构以及说明：

```
CREATE TABLE `jb_products` (  
  `id` int NULL AUTO_INCREMENT COMMENT '自增id',  
  `first_category` varchar(30) NULL COMMENT '一级大类，比如美妆个护',  
  `third_category` varchar(30) NULL COMMENT '三级大类，比如唇部',  
  `second_category` varchar(30) NULL COMMENT '二级大类，比如香彩妆护',  
  `img1_address` varchar(255) NULL COMMENT '商品图片的链接地址，至少一个，至多5个',  
  `address` varchar(255) NULL COMMENT '商品链接网址',  
  `img2_address` varchar(255) NULL COMMENT '商品图片的链接地址，至少一个，至多5个',  
  `img3_address` varchar(255) NULL COMMENT '商品图片的链接地址，至少一个，至多5个',  
  `price` decimal(10,2) NULL ,  
  `img4_address` varchar(255) NULL COMMENT '商品图片的链接地址，至少一个，至多5个',  
  `description` varchar(255) NULL COMMENT '商品全称/简述，比如“克丽丝汀迪奥DIOR魅惑润唇蜜SPF10 3.5g（又名：克丽丝汀迪奥魅惑润唇膏 口红（焕彩） 3.5g）”',  
  `number` varchar(30) NULL COMMENT '商品编号，比如1458853',  
  `name` varchar(255) NULL COMMENT '商品名字，比如迪奥唇膏',  
  `brand` varchar(255) NULL COMMENT '品牌，比如迪奥',  
  `img5_address` varchar(255) NULL COMMENT '商品图片的链接地址，至少一个，至多5个',  
  `produce_address` varchar(255) NULL COMMENT '产地，比如法国',  
  `comment_count` varchar(255) NULL COMMENT '商品评价总数，比如：4万+',
```

```
`kg` varchar(255) NULL COMMENT '商品毛重，比如35.00g',  
`good_for_who` varchar(255) NULL COMMENT '适用人群',  
`expiration_date` datetime NULL COMMENT '保质期',  
`color` varchar(255) NULL COMMENT '比如橘黄色',  
`result_effectiveness` varchar(255) NULL COMMENT '功效，比如“丰唇”',  
`category` varchar(255) NULL COMMENT '分类，比如“唇膏/口红”',  
`makeup_effectiveness` varchar(255) NULL COMMENT '妆效',  
`good_comment_percentage` varchar(255) NULL COMMENT '好评度，比如78%',  
`who_handly` varchar(255) NULL COMMENT '负责人，爬取这条记录的人的名字，比如张敏华',  
PRIMARY KEY(`id`)  
)  
;
```

注:表格的字段是我根据上面那个口红的数据设计，其他的商品可以有些额外的属性，如果需要的话，自己往商品表增加属性；也有可能有些商品没有某个属性，比如香水可能没有颜色这个属性，那么这个属性的值请设置为 **null**。

（ps，暂时不要管表的设计问题，我们现阶段的目标是得到数据~~）

四、Github仓库

仓库地址 `git@github.com:ZJane/BeautifulGirls.git`，请大家把自己的ssh公钥发给我，我需要导入到项目中，然后你们才有权限克隆、编辑仓库。

五、云笔记协作

为了方便团队协作，特别是方便文档合作编辑，请大家注册个有道云笔记（可以用网页版，也可以下个桌面版的），然后把你们的账号发给我。