**阿里巴巴宝贝详情爬虫程序开发需求**

# 开发背景

主要有如下两方面原因：

1. 选货时制作进货单，需要手动输入标题、链接、主图、颜色、尺码、单价、费用等信息，全部手动操作，耗时费力，消耗了无意义的时间和精力；
2. 宝贝上架时，一件上传宝贝的图片，部分图片不理想，需要对图片进行二次加工处理。此时需要获取原图，之前通过手动提取代码制作html文件，使用迅雷批量下载的方法，同意手动操作频繁，效率低下，浪费时间。

综合上述两方面原因，需要进行此爬虫程序的开发。

# 业务需求

## 单个宝贝的信息抓取和保存

### 进货单报表自动生成

根据制定的宝贝链接，加载宝贝信息，提取制作进货单时需要的信息，根据模板自动生成xlsx文件，将信息自动写入到excel文件中，生成进货单。需要提取的宝贝信息如下：

标题、链接、主图、颜色、尺码、单价、费用

### 宝贝图片下载

自动提取宝贝的所有图片链接，包括主图、颜色图、详情图，去除缩放的图片属性（下载原图），生成只包含宝贝图片链接汇总的html文件，并将所有的图片，下载到本地的宝贝同名文件夹下。

## 批量抓取宝贝信息

给出一个包含宝贝列表的txt文件，一行一个链接，换行符分割，运行程序时，读取此文件内容，依次抓取宝贝的网页信息，完成报表生成和图片下载。