

基于爬虫的APP批量制作

北京联龙博通电子商务技术有限公司

2016年05月

**修订历史记录**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **日期** | **描述** | **版本** | **作者** | **审核者** |
| 1 | 2016-5-22 | 文档初建 | 1.0 | 雷勋 |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

本文档中所包含的信息属于机密信息，如无有限公司的书面许可，任何人都无权复制或利用。

目录

[1 引言 3](#_Toc451696881)

[2 问题一：开发一款app内容从哪来 4](#_Toc451696882)

[2.1 Jsoup解析器 4](#_Toc451696883)

[2.2 基于jsoup开发的一款android应用----悦读圈 4](#_Toc451696884)

[特别关注-成熟人士的读者文摘（http://www.de99.cn/index.html） 4](#_Toc451696885)

[3 问题二：如何才能进行快速开发 5](#_Toc451696886)

[3.1 项目结构 5](#_Toc451696887)

[4 问题三：怎么盈利 5](#_Toc451696888)

[5 问题四：怎么推广 5](#_Toc451696889)

[6 问题五：其它注意问题 5](#_Toc451696890)

# 引言

很多人都有赚钱的想法，但许多行业的尝试成本太高，程序员就是可以随时行动，不需要做太多投资就可以实实在在的进行尝试，尝试的过程中又会衍生出许多的想法或者机会，关键在于行动。现在就本人曾经尝试过的几款app来跟各位分享下利用爬虫技术开发app的前后经历。

# 问题一：开发一款app内容从哪来

内容问题，首先想到的就是爬虫，因为爬取目标特别明确，所以无需采用爬虫框架，只需要简单的网页解析器即可，本人采用的是jsoup解析器，

## Jsoup解析器

jsoup 是一款Java 的HTML解析器，可直接解析某个URL地址、HTML文本内容。它提供了一套非常省力的API，可通过DOM，CSS以及类似于jQuery的操作方法来取出和操作数据。具体使用可以参考这个链接地址 <http://www.open-open.com/jsoup/>

## 基于jsoup开发的一款android应用----悦读圈

https://www.pgyer.com/enjoyread）

基于此技术开发了一款悦读圈app，内容来源于网站：

### [特别关注-成熟人士的读者文摘](https://www.baidu.com/link?url=EMuPCagveXZCt9-P13ZbzQn42BO4sxza__OQY6pQOr9PQU4Uwrst81ER6kjUwe0s&wd=&eqid=eb5e7c2800358d2c0000000357413c15)（http://www.de99.cn/index.html）

用到的资源：又拍云存储（内容分享时，修改过的html存储），百度应用引擎（主要负责app版本更新）

文章列表：解析html到javabean对象，处理是否有下一页（加载更多）

内容详情：内容详情页的图片需要根据手机屏幕做处理，如果是有多页的文章还需要处理是否有下一页的逻辑

分享内容：需要修改html添加下载链接，或者字体大小处理

正是悦读圈（https://www.pgyer.com/enjoyread）的发布，让我感觉到了可以再多做几款app进行尝试集成广告赚钱，于是如何高效快速开发成为了下一步思考的问题。

# 问题二：如何才能进行快速开发

以爬虫抓取到的网页内容完全可以由自己进行二次整理，然后做成符合自己要求的app，我大致将内容分成两种，一种是纯图片型的，一种是文章带图片型的，于是我的app核心内容模块也是这两种，tips：请各位见谅我app内容的低俗，我想这只是我行进路上的一次小小的尝试，有必要走这一步

## 项目结构

1：资源项目(自定义的布局文件，颜色，字符串，url，主题等）

2：公共库项目（公共模块）

3：核心逻辑项目（公共模块）

4：上层模块项目 （页面风格，包名，签名，activity等的注册）

# 问题三：怎么盈利

# 问题四：怎么推广

# 问题五：其它注意问题