**中南民族大学**

**毕业论文(设计)**

**学院： 计算机科学学院**

**专业： 年级：**

**题目:集新闻阅读和电影推荐为一体的微信小程序设计与实现**

**学生姓名：学号：**

**指导教师姓名： 职称:**

**2020年** 5 **月**

**中南民族大学本科毕业论文（设计）原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名： 年 月 日

目录

[摘要： 1](#_Toc39062950)

[Abstract: 1](#_Toc39062951)

[1 引言 2](#_Toc39062952)

[**1.1** **研究背景** 2](#_Toc39062953)

[**1.2** **研究现状** 2](#_Toc39062954)

[**1.2.1** **新闻类微信小程序** 2](#_Toc39062955)

[**1.2.2** **电影推荐研究现状** 2](#_Toc39062956)

[**1.3 论文结构** 3](#_Toc39062957)

[**2 系统分析与设计** 3](#_Toc39062958)

[**2.1 可行性分析** 3](#_Toc39062959)

[**2.2 需求分析** 3](#_Toc39062960)

[**2.3** **功能设计** 4](#_Toc39062961)

[2.4 小程序页面设计及实现 5](#_Toc39062962)

[**2.5 数据采集和处理** 6](#_Toc39062963)

[**2.5.1 用python进行网络爬虫** 6](#_Toc39062964)

[**2.6.2 网络爬虫具体实现** 6](#_Toc39062965)

[4 **总结** 9](#_Toc39062966)

[**4.1全文总结** 9](#_Toc39062967)

[**4.2 不足与展望** 9](#_Toc39062968)

[**致 谢** 9](#_Toc39062969)

[**参考文献** 9](#_Toc39062970)

**集新闻阅读和电影推荐为一体的微信小程序设计与实现**

摘要：

**关键词：**

**Set news reading and movie recommendations as one of the WeChat small program design and implementation**

Abstract:

**Key words**:

# 1 引言

* 1. **研究背景**

随着社会的进步和信息技术的快速发展，各类信息的数据量呈爆炸式增长。一部手机，使人们逐渐从信息闭塞的时代进入了信息便捷的“大数据时代”。人们获取信息的方式从开始的报纸阅读、电视上的新闻联播等变成了从今日头条、人民日报等各类APP上面获取实时热点信息。但是随着技术的进步和生活需求的提高，手机里面下载的APP越来越多，既占内存又增加了暴露隐私的危险。在手机客户端APP几近泛滥的时代，再高端的手机配置也禁不住安装几十个乃至更多的APP。“简洁”已经成为了大多数用户追求的目标。

2017年1月9日，微信小程序正式开放。微信小程序有三大优点：不需要下载安装即可使用；用户“用完即走”，不用关心下载了很多应用；无处不在，随时应用。从用户的角度看来，能够简化按照下载步骤，又能节省手机内存空间。另外，微信小程序对中小型企业非常友好，开发环境和开发逻辑比较简单，能够降低开发成本和推广成本。正是因为这些优点，微信小程序受到了许多人的喜爱。

* 1. **研究现状**
     1. **新闻类微信小程序**

新闻类的微信小程序在内容上均以图片、文字、视频和音频作为主要的内容载体，其中图文资料是新闻承载的主要形式，辅助视频和音频等更为生动的传播形式，形成了有效的内容结构化构建，有效的提升了新闻媒介的信息密度和吸引力。现有的新闻类的微信小程序主要有三方面来源：

1. 以传统新闻媒体为主要承办单位的新闻小程序，如央视网的“新闻联播”、上海广播电视台的“新闻坊+”等。
2. 以第三方门户网站平台为承办单位的新闻小程序。如腾讯的“腾讯新闻”、网易的“网易新闻精选”等。
3. 以其他第三方组织团体为承办单位的新闻小程序，如北京智者天下公司的“知乎热榜”、上海东方网的“东方头条新闻”等

与新闻类其他传播媒介如网站、APP、广告等相比，新闻类微信小程序具有一定的优势，具体而言有以下几个方面：

1. 不需要安装即可使用：微信小程序嵌入在微信之中，不需要在独立的安装便可以实现阅读，避免了因复杂的下载和注册登录等操作造成的客户流失，也避免了由于下载程序等操作而带来的内存忧虑。
2. 基于社交的传播路径：微信小程序依赖微信社交平台进行传播，如将小程序分享给好友、朋友圈以及多人的微信群，会有更多的人看到并使用。
   * 1. **电影推荐研究现状**

我们经常接触到的电影推荐无外乎是豆瓣电影、猫眼电影以及知乎社区等。豆瓣电影拥有大量的影迷用户和电影数据库，主要提供最新的电影介绍、电影排行榜、豆瓣评分以及用户评论，且包括上映影片的查询和购票服务等，用户可以记录想看、在看和看过的电影，进行电影打分、写电影评论等。从中可以看出，电影推荐主要从两个维度展开，一个是时间轴，按照当前时间段上映的电影热度进行推荐，如“正在热映”、“近期热映”和“电影排行榜”等功能；第二个是以用户为中心，即大众喜好为参照，用户进行推荐，如“本周口碑榜”、“热门豆列”和“受欢迎的影评”等功能模块。猫眼电影和豆瓣电影类似，而知乎社区主要是用户之间的互动，用户围绕着某一感兴趣的话题，如“电影推荐”，分享着彼此的经验和见解，从而进行推荐。

除此以外，还有许多学者把推荐算法进行改进，将其运用到电影推荐上面。如论文考虑到了用户电影评论中的情感信息，将基于物品的协同过滤算法进行改进，综合用户对电影的评分和情感进行推荐。

**1.3 论文结构**

**2 系统分析与设计**

**2.1 可行性分析**

可行性分析是系统开发的第一步，分析该项目是否可以顺利进行下去，它是对项目具体决策的一个系统性的分析和评估。可行性的分析结果决定了这个系统是否可行，是否能够顺利的开发，本文主要对本系统的技术方面的可行性做分析如下：

1. 数据：本文需要电影数据和新闻数据。一般可以通过调用API来获取数据，新闻API有网易新闻API可以调用，电影API有豆瓣电影API。由于现在豆瓣电影API已经不能调用了，可以采取网络爬虫的方法爬取数据。为了数据的统一，新闻数据也采用网络爬虫的方式获取。
2. 微信小程序设计：微信小程序开发平台是由腾讯官方网站提供的——微信开发工具。微信小程序的页面一般由.js文件、.json文件、.wxml文件和.wxss文件构成。.js文件主要是写页面逻辑，用于处理页面的生命周期以及数据处理等；.json文件用于写页面配置信息，如设置页面文件的路径，底部栏tabBar设置以及窗口的颜色设置等；.wxml文件是页面布局文件，在里面可以设置页面的布局方式，添加控件等；.wxss文件是样式文件，主要进行页面的样式设置，进行样式美化等。这些需要前端技术来实现，并且微信小程序开发已经相对比较成熟，有许多优秀的用例可以进行学习。

经过以上2点进行分析得出结论，本系统开发可行。

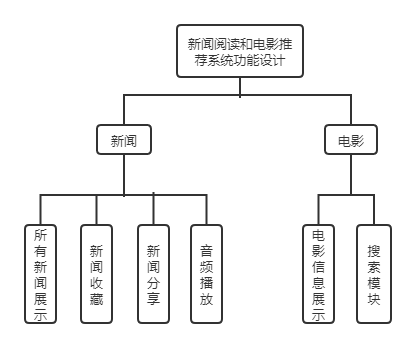
**2.2 需求分析**

需求分析主要包括功能需求分析和非功能需求分析。功能需求分析是分析系统在功能需要实现什么，是规定了开发人员在产品中实现的软件功能，用户运用这些功能来进行操作，满足业务需求。

本系统的整体需求是：通过该应用能够查看新闻信息，并对单个新闻信息可进行分享和收藏，可播放音频；通过电影页面，可查看最近所有的电影信息，可通过电影名字进行搜索，可查看某一部电影的具体信息。

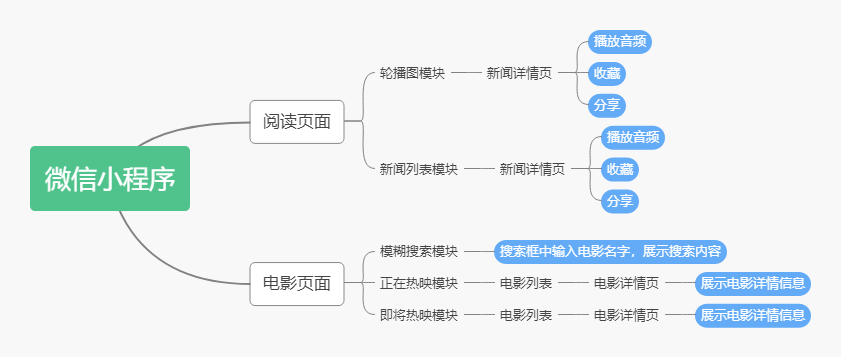
* 1. **功能设计**

系统的功能模块设计如图所示：



1. 所有新闻展示：在一级页面展示所有新闻的标题、来源、发布时间以及新闻图片；在二级页面展示新闻的详细信息。
2. 新闻收藏：点击收藏按钮，可进行新闻收藏，再点击一次，可取消收藏
3. 新闻分享：点击分享按钮，出现弹框，点击确认进行分享，点击取消，则取消分享。
4. 音频播放：点击音频播放按钮可进行音乐播放，再点击可取消播放。
5. 电影信息展示：在一级页面展示电影名字、电影封面和具体评分；在二级页面展示电影详细信息。
6. 搜素模块：在搜索框输入电影名字，可进行搜索。
   1. 小程序页面设计及实现

按照系统的功能模块设计，进行小程序的页面设计。小程序页面主要包括阅读页面和电影页面。首先进入程序时，根据底边栏的选择进入“阅读”还是“电影”，默认进入“阅读”页面。



阅读的一级页面由轮播图模块和新闻列表模块组成，轮播图里面放的是图片，点击图片，可以跳转到对应的新闻详情页。是通过catchtap点击事件来形成的。

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a)新闻一级页面 | (b)新闻详情页 | (c) 播放音频 |
|  |  |  |
| （d）分享页面 | （e）分享列表 | （f）分享到朋友圈 |

**2.5 数据采集和处理**

**2.5.1 用python进行网络爬虫**

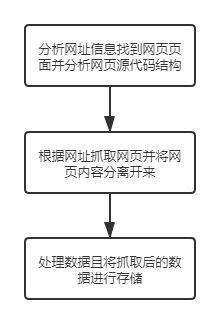
在设计和实现本系统时，需要电影数据信息和新闻数据信息作为支撑。一般情况下，有API接口可以调用，如电影数据有豆瓣电影API可以使用，新闻数据信息有网易新闻API。但是在程序的编写过程中发现，豆瓣电影API现在已经不能使用了，所以不能用使用豆瓣API来获取数据了，故打算使用python进行网络爬虫来获取数据。

网络爬虫是一个从web页面上自动下载网页的程序，即可以按照某种规则从网络上自动爬取用户所需的内容的脚本程序。主要分为连接数据和解析数据两大部分，如果是直接爬取一个页面的话，可以直接根据页面的url获取网页信息，然后再解析数据；若是涉及到多个页面的信息，则需要根据页面的url，进行页面的跳转。一般情况下，第一个页面包含其他网页的入口，爬虫可以通过一个url，链接进入其他网址获取内容，然后解析内容，最后得到所需要的信息数据。目前python语言是最适合用来爬虫了，python语言整合了网络爬虫的一系列库，如requests库、BeautifulSoup4库等，为爬虫提供了便利，能够高效得完成目标数据的获取。

打算爬取豆瓣电影Top250数据作为电影信息，爬取人民日报数据作为新闻信息。爬取电影信息主要需要电影名字、上映时间、电影封面、评分、导演、主演、类型、电影简介；爬取新闻信息主要需要标题、内容、新闻图片、来源、日期。

**2.6.2 网络爬虫具体实现**

用python进行网络爬虫的主要思路如下所示：



爬取电影信息和新闻信息的方式和流程都是一样的，下面以爬取豆瓣电影信息为例进行说明：

1. **分析网页的url,寻找共性**：每个页面显示10部电影，250部电影一共显示了10页。观察前几个页面的url可以很容易的发现其中规律，如第一页到第三页的url如下所示：

第一页：<https://movie.douban.com/top250?start=0&filter=>

第二页：<https://movie.douban.com/top250?start=25&filter=>

第三页：<https://movie.douban.com/top250?start=50&filter=>

发现页面的url规律是：start后面的参数变化为25\*（页数-1）。



1. **进行网页数据请求获取**

网页请求分为2个部分，示意图如下：



* 1. 请求（Requests）：用户向服务器发送请求。请求分为GET和POST，GET是最常见的方式，也是大多数网站使用的方式，一般用于获取或查询信息，响应速度较快；POST方式相比GET，除了查询信息外，多了可以以表单的形式传递参数，因此可以用来修改信息。
  2. 响应（Response）：服务器在接收到来自用户的请求后，会验证请求的有效性，然后向用户发送响应，用户或客户端接收了服务器的响应内容后，将内容展示出来。

另外，现在许多网站都有反爬虫机制，若单独使用requests进行请求，则会返回requests[418]。因此需要进行请求头的伪装。

1. **数据解析**

请求到了网页信息之后，需要对网页信息进行解析，得到所需要的有价值的信息。Python的网页解析器主要有四类：正则表达式，使用第三方库re，进行模糊匹配；xpath,使用第三方库from lxml import etree;BeautifulSoup4,为结构化解析；pyquer解析器，相当于用python实现jQuery，可用于解析HTML文件。因为网上使用BeautifulSoup4进行解析的例子较多，故本文选择BeautifulSoup4进行页面的解析，以便出现bug时，进行学习。以爬取第一个页面为例，展示实验的请求网页和解析网页部分的代码：

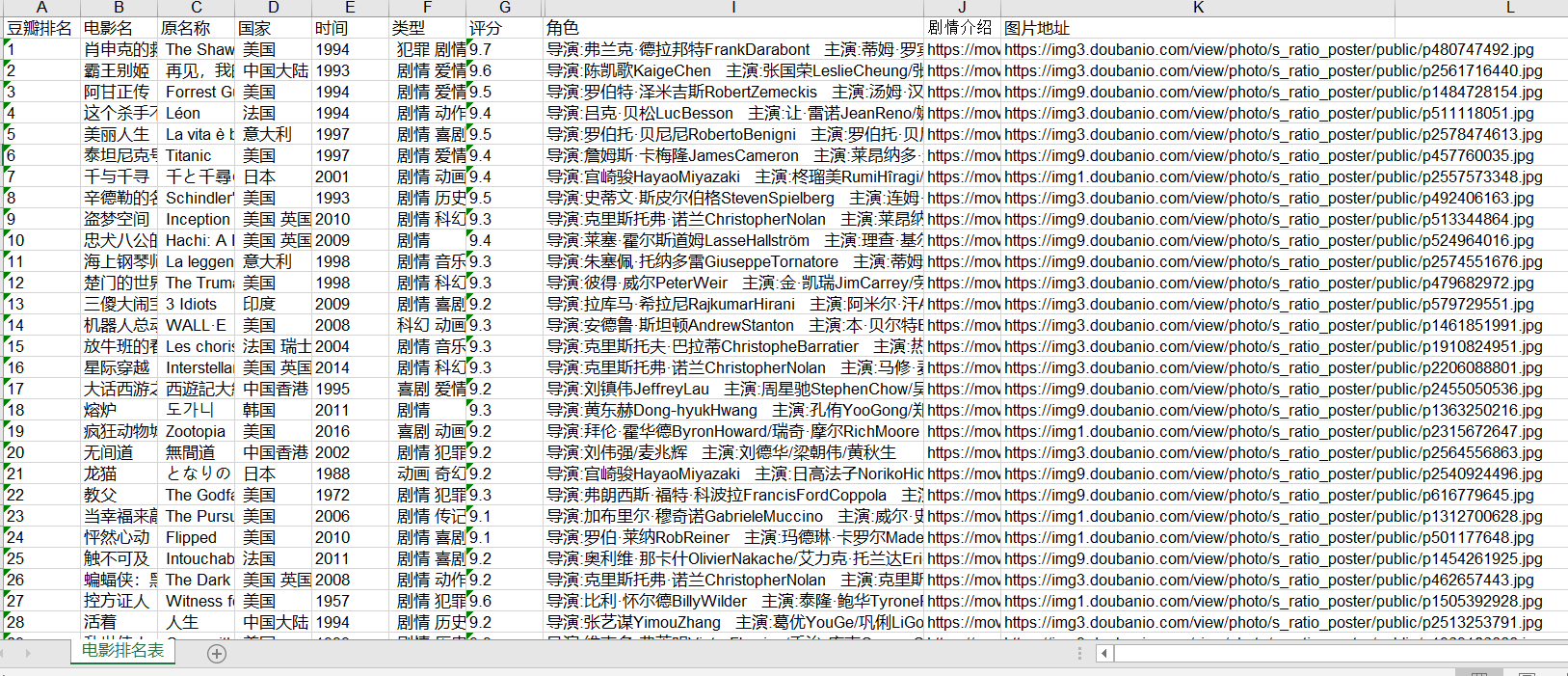


对每一个电影详情页面的解析如下：先确定每一个数据对应的元素及class名称，再通过find()或find\_all()方法进行标签的定位，然后选取每一个字段数据。部分伪代码如下：



1. **数据存储**

将爬取的电影信息以csv的格式进行存储如下：



4 **总结**

**4.1全文总结**

**4.2 不足与展望**

**致 谢**

**参考文献**