**南京航空航天大学计算机科学与技术学院 大学生科技创新基金项目申报书**

完成比例：杨鹏 55% 张帅 45%

项目名称：基于网络爬虫技术的信息汇总APP

申 报 人：

所在专业：

手 机：

指导教师：

办公电话：

手 机：

计算机科学与技术学院大学生科技创新基金管理办公室

填表日期: 2016 年 09 月 16 日

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 项目名称 | | | 基于网络爬虫技术的信息汇总APP | | | | | | | | | | | | |
| 起止日期 | | |  | | | | | | | | | | | | |
| 申请经费 | | |  | | | | | | | | | | | | |
| 申请者情况 | | | 学 号 | | | 姓 名 | | | 平均绩点 | E-mail | | 电 话 | | | 签　名 |
| 031430330 | | | 杨鹏 | | |  |  | |  | | |  |
| 161410113 | | | 张帅 | | |  |  | |  | | |  |
|  | | |  | | |  |  | |  | | |  |
|  | | |  | | |  |  | |  | | |  |
|  | | |  | | |  |  | |  | | |  |
| 指导老师 | | | 姓　名 | | | 单　位 | | | | | | 职　称 | | | 签　名 |
|  | | |  | | | | | |  | | |  |
| 项目负责人获奖情况 | | | | | | | | | | | | | | | |
| 项目编号 | | | | （评审用，办公室填写） | | | | | | | | | | | |
| 研究内容提要（限200字以内）   1. 基于Android系统的客户端软件。开发在Android系统上运行的客户端软件，为用户提供校园内各大机构和组织的官网新闻整合发布，用学号注册登录后可以进行评论、加星提醒、分享到外部平台以及分类分发等功能，方便用户方便快捷地了解到各种重要消息。 2. 基于Tomcat等技术搭建服务器平台以及MySql数据库搭建稳定的服务器端处理系统，进行网络爬虫处理并实现对客户端功能的支持。   3. 基于面向对象方法的软件系统分析与设计。 | | | | | | | | | | | | | | | |
| 研究的目的和意义（限500-1000字）  由于大学里面有很多消息都是发在不同的机构或组织自己的官网上面，而其中很多网站由于整体的架构在设计的时候年代过于久远，很多仅支持ie6.0进行登录访问，现在一个最严重的问题是目前很多电脑由于操作系统的更新换代windows自带的ie浏览器版本也改变了很多，很多时候会产生网站不兼容当前的浏览器情况导致无法访问，而且解决的方法需要安装一些不必要的软件，给用户带来了很多麻烦。  第二个问题是由于大学校园的组织和机构数目并不少并且访问过于繁琐，很多同学（包括我自己）都没有习惯性地去访问那些网页。网页消息的更新并没有准确时间，而很多消息都关联到很多政策的执行，具有很强的时效性，目前很多通知都是班级的某个负责人在IM软件上进行转发通知，是一种效率比较低的方式。  现在同学们几乎都拥有智能手机并且大多数访问都是通过手机完成，但是很多组织和机构的网页并没有对手机端的浏览器进行适配，导致手机浏览器访问对应的网站体验非常差，所以我们通过在服务器端使用Python对学校内的各种网站（教务处、奥兰、学院网站、校内论坛等）进行爬虫并分好类后把对应的内容通过APP端展现出来，极大的改善用户体验，并且对信息资源进行整合后，对用户来说这是一种能方便快捷的了解到众多消息的方法。而通过加入的评论、分享、分类关注等功能能够有效地提高学生们对通知的了解程度，达到增强宣传效果的作用，而加星提醒功能则是一个点对点的个性化功能，能够很有效的帮助学生对一些关系到自己的重要通知进行时间把握，尽可能地减小遗漏的可能。 | | | | | | | | | | | | | | | |
| 国内外本项目的研究状况（800字左右，附不少于4篇的参考文献）   1. 新闻推荐类APP:国内目前最火的要数今日头条，当然也有网易新闻、Flipboard、搜狐新闻等平台，很多是既拥有网页端也有APP客户端的平台。   今日头条  (今日头条网页端)  今日头条网易新闻  (今日头条APP端) (网易新闻网页端)  今日头条：是一款基于数据挖掘的推荐引擎产品，它为用户推荐有价值的、个性化的信息，提供连接人与信息的新型服务，是国内移动互联网领域成长最快的产品服务之一。  网易新闻:提供极具网易特色的新闻阅读、跟贴盖楼、图片浏览、话题投票、要闻推送、离线阅读、流量提醒等功能。  Flipboard:将Facebook和新浪微博等社交媒体上的内容整合起来以杂志的形式呈现给用户阅读。[1]  可见这些APP的特点都非常明显，有走个性推荐的，有走评论交流的、也有走定向整合的路线，他们都很具有特色的抓住了用户习惯，可以提供很强的用户粘性，都做到了新闻推荐类APP里面的比较大的分量。但是由于这些都是商业化公司，并且为了扩大用户群，会加入许多很杂乱的新闻消息，而且很多意义不大，甚至有标题党的行为存在。我们希望自己做的APP专注于学校新闻的整合和分类推广，而且由于具有公益性质，并且从官网爬取新闻的途径能够保证新闻的准确性和精准性，更像一种学习生活工具而不是娱乐软件。而且通过学号注册的方式，能够避免恶意用户的进入，创造一个干净实用的软件。  2.《四川大学资讯整合网站的研究与设计》：随着新媒体的飞速发展与逐渐成熟，新媒体这一媒体形式凭借其便利、便捷的特点获得了人们的认可和关注。由于新媒体平台的数量增长，人们已经逐渐迷失在浩瀚的网络中，无法在有限的时间内及时、准确地获得想要的信息，所以对新媒体进行有效的资源整合成了解决这一问题的首要途径。本文以四川大学教务处网站、四川大学团委网站、四川大学学生工作处网站为研究背景，结合计算科学技术，提出了新媒体资源整合的方法，使广大的学生群体在有限的课余时间内准确、高效地了解校内外的资讯。[2]  网络爬虫原理、  3.网络爬虫原理：网络爬虫是一个自动提取网页的程序，它为搜索引擎从万维网上下载网页，是搜索引擎的重要组成。传统爬虫从一个或若干初始网页的URL开始，获得初始网页上的URL，在抓取网页的过程中，不断从当前页面上抽取新的URL放入队列,直到满足系统的一定停止条件。[3]  4.微信JS-SDK是微信公众平台面向网页开发者提供的基于微信内的网页开发工具包。通过使用微信JS-SDK，网页开发者可借助微信高效地使用拍照、选图、语音、位置等手机系统的能力，同时可以直接使用微信分享、扫一扫、卡券、支付等微信特有的能力，为微信用户提供更优质的网页体验。[4]  参考文献：  [1]《“今日头条”的竞争优势与传播学解读》，湖南大学新闻传播与影视艺术学院 蔡永波  [2]《基于网络爬虫的四川大学资讯整合网站的研究与设计》，科协论坛（下半月），2013年07期，四川大学软件学院 刘晓阳  [3]《基于模板化网络爬虫技术的Web网页信息抽取》，电子科技大学 2012年乔峰  [4]微信JSSDK概述  <https://mp.weixin.qq.com/wiki/7/aaa137b55fb2e0456bf8dd9148dd613f.html#.E6.A6.82.E8.BF.B0> | | | | | | | | | | | | | | | |
| 研究内容、研究方案（研究方法、技术路线）（1000字左右）   1. 运用面向对象的软件开发方法，在Android平台进行信息平台的开发。首先要行面向对象的分析（OOA――Object Oriented Analysis），其任务是了解问题域所涉及的对象、对象间的关系和作用（即操作），然后构造问题的对象模型，力争该模型能真实地反映出所要解决的"实质问题"。在这一过程中，抽象是最本质、最重要的方法。针对不同的问题性质选择不同的抽象层次，过简或过繁都会影响到对问题的本质属性的了解和解决。 2. Android客户端的开发：Android以Java为编程语言，使接口到功能，都有层出不穷的变化。Android客户端以Activity为基础，可以通过不同Activity进行不同功能。而Activity间的参数传递可通过Bundle进行。Android的界面布局可通过xml进行，android提供了各种功能丰富的控件，可以完成各种界面。因为Android支持Http协议，故可以使用Http协议和服务器进行通讯。、 3. 分享SDK的使用：分享SDK可在用户将重要的消息分享给微信或者QQ上的好友，或者朋友圈、QQ空间等平台，这样便于信息的传播。只需在编写APP的时候利用对应的SDK调用API即可实现这些功能，但是却提供了很大的便利。用微信分享举个实现的例子：   wx.onMenuShareAppMessage({  title: '', // 分享标题  desc: '', // 分享描述  link: '', // 分享链接  imgUrl: '', // 分享图标  type: '', // 分享类型,music、video或link，不填默认为link  dataUrl: '', // 如果type是music或video，则要提供数据链接，默认为空  success: function () {  // 用户确认分享后执行的回调函数  },  cancel: function () {  // 用户取消分享后执行的回调函数  }  });   1. 爬虫系统：在服务器端，先利用Python将爬虫实现并且生成新的文件，存储在服务器上。通过Tomcat将最终的文件分发出去。 2. 评论信息的存储：使用MySql数据库进行用户信息和评论信息的存储。 3. 具体UI的设计：由于最终还是一个消息分发平台，我们可以参考今日头条、网易新闻等在网页布局和展示以及评论栏样式以及操作按钮等的设计，尽量实现简洁易用的特点。 | | | | | | | | | | | | | | | |
| 技术指标   1. 客户端：拥有良好易用的用户界面，完善的功能，实现校园网站新闻内容的整合，并支持预期的评论、加星提醒、分享到外部功能。 2. 服务器:能够支持Tomcat、Python和MySql中正常运行。 | | | | | | | | | | | | | | | |
| 研究计划及预期成果 | | | | | | | | | | | | | | | |
| 主要研究阶段  （起止时间） | | | | | 阶段预期成果 | | | | | | | | | 成果形式 | |
| 前期 |  | | | |  | | | | | | | | |  | |
| 中期 |  | | | |  | | | | | | | | |  | |
| 后期 |  | | | |  | | | | | | | | |  | |
| 最终研究成果 | | | | | 最终成果名称 | | | | | | | | | 最终成果形式 | |
| 提示：科技论文、技术报告、软件著作权、专利等 | | | | | | | | | 系统应用软件及网站 | |
| **经费预算** | | | | | | | | | | | | | | | |
| 序号 | | 预算项目 | | | | | | 预算金额（元） | | | 备注 | | | | |
| 1 | | 图书资料费 | | | | | |  | | |  | | | | |
| 2 | | 交通费 | | | | | |  | | |  | | | | |
| 3 | | 调研及学术交流费 | | | | | |  | | |  | | | | |
| 4 | | 论文版面费 | | | | | |  | | |  | | | | |
| 5 | | 印刷费 | | | | | |  | | |  | | | | |
| 6 | | 上机机时 | | | | | |  | | |  | | | | |
| 7 | | 元器件 | | | | | |  | | |  | | | | |
| 8 | | 耗材 | | | | | |  | | |  | | | | |
| 9 | | 其他（注明项目） | | | | | |  | | |  | | | | |
| 合计 | |  | | | | | |  | | |  | | | | |
| 学院辅导员意见  辅导员签字： 年 月 日 | | | | | | | | | | | | | | | |
| 指导教师意见  指导教师签字： 年 月 日 | | | | | | | | | | | | | | | |
| 基金专家评审组意见  组长签字： 年 月 日 | | | | | | | | | | | | | | | |
| 基金专家评审组成员 | | | | | | | | | | | | | | | |
| 姓名 | | | | | | | 职 称 | | | | | | 签字 | | |
|  | | | | | | |  | | | | | |  | | |
|  | | | | | | |  | | | | | |  | | |
|  | | | | | | |  | | | | | |  | | |
|  | | | | | | |  | | | | | |  | | |
|  | | | | | | |  | | | | | |  | | |
| 学生科技创新基金管理办公室意见  主管领导签字： 年 月 日 | | | | | | | | | | | | | | | |