**聊天辅助软件**

**项**

**目**

**介**

**绍**

**书**

课 程： 软件工程导论

题 目： 聊天辅助软件设计与开发

专 业： 软件工程

班 级： 1602

小组成员：苏碧青 31601382(组长)

陈铭阳 31601386 刘乐威31601400 蓝舒雯31601380

1. **项目背景**
2. **开发背景**

生活中我们经常会碰到这样的问题，在与对面聊天时无法继续新的话题而突然安静或者因为聊天层次不同没有合适的话题而导致沉默。为了解决这样的社交问题，我们决定着手开发这个聊天辅助软件。通过大数据搜索引擎，为你即时推送合适的话题，帮助你打破沉默，轻松掌握话题。

1. **开发工具**

Python：Python是一种解释型、面向对象、动态数据类型的高级程序设计语言。

JavaScript：一种直译式脚本语言，是一种动态类型、弱类型、基于原型的语言，内置支持类型。

Photoshop：PS在线图片编辑器是一个专业的在线ps照片处理软件。

SQL：用于访问和处理数据库的标准的计算机语言。

Git：是一个开源的分布式版本控制系统，可以有效、高速的处理从很小到非常大的项目版本管理。

Project：通用的项目管理工具软件。

Html5：万维网的核心语言、标准通用标记语言下的一个应用超文本标记语言。

1. **项目开发系统环境**

Linux ,Windows XP/7/8/8.1/10

1. **项目投资方**

SE2018春-G11小组

1. **项目需方、用户**

浙江大学城市学院全体师生

1. **项目开发方**

SE2018春-G11小组

1. **支持机构**

浙江大学城市学院计算机与计算科学学院

1. **参考资料**

《Software Engineering》出版社：机械工业出版社 作者：Roger S.Pressman

《SQL编程基础》 出版社: 机械工业出版社作者： Patrick.J.J.

《深入浅出——MySQL开发、管理与应用实例》出版社：清华大学出版社 作者：邹建

《数据库系统概论》 出版社：高等教育出版社 作者：王珊，萨师煊

《Python基础教程》 出版社：人民邮电出版 作者：Magnus Lie Hetland

1. **项目的可行性分析**
2. **项目的目标**

目标实现软件的基础功能、进阶功能和创新功能。

主功能（基础功能）：打开app，用户登录账号以后可以在app界面输入关键字来获取所需信息。搜索得出的信息以弹窗的形式呈现。不感兴趣的信息可以左滑删除，感兴趣的信息可以通过长按信息将信息复制到聊天框发送。

进阶功能：用户提供社交账号（微博，微信等），后台用爬虫搜索账号数据，以此来获得所需信息。

创新功能：退出app后若app账号未退出登录，则以浮窗展示的形式显示在当前屏幕上。若需使用，则直接点开浮窗，输入关键字实施操作。

1. **项目开发难点**

（1）关键技术

一是数据仓库：分布式爬虫系统，分布式数据库（线下备份）；

二是数据挖掘：评论文本挖掘，信息文本挖掘；

三是数据分析：海量数据的清洗、统计分析和建模

（2）技术难点

* + - * 1. 用Python实现一个大数据搜索引擎；
        2. 用python爬虫爬取指定用户微博图片及内容，并进行微博分类及使用习惯分析，生成可视化图表；
        3. 用python爬虫获取热门微博评论并进行情感分析；
        4. 用python在大量数据中根据条件筛选数据；
        5. App兼容性问题；
        6. 前端设计和多平台账号打通；
        7. 分别嵌入新浪微博、微信做第三方授权登录获取到头像、昵称等信息、账号绑定及解绑；
        8. JS实现可缩放、拖动、关闭和最小化的浮动窗口；
        9. 搜索消息以弹窗形式呈现的功能。

1. **项目的需求分析**
2. **项目要求**
3. 用户需求

此软件能登录使用，允许用户输入对方年龄，职业，性别来供给各类谈资八卦新闻材料，也可以在大数据的基础上加上链接私人软件，例如微信、微博，可以快速掌握喜好，并且会定位精确很多。

1. 性能需求

预计app大小为35M

预计搜索内容能在1s内呈现给用户

确保app的安全性，维护用户的隐私，通信加密

1. 可靠性和可用性需求

app在使用时1天内不能超过2次故障，该app能在任意一台安装了安卓系统的手机上使用。

1. 出错处理需求

若系统对环境发生了错误响应，则系统会自动关闭并弹出报错窗口，用户可通过报错窗口向管理员反应出错信息。

1. 接口需求

Ⅰ.用户接口

界面风格：采用图形界面。

界面操作：界面上的每个按钮都是经过精心设计，以求客户使用方便。

界面消息：以弹窗形式显示搜索所得的消息。

Ⅱ.硬件接口

通信协议：HTTP协议

Ⅲ.软件接口

数据库：采用数据库SQL开发。

操作系统：开发系统为Linux。

工具：Python、Photoshop、Git、Css3等。

（6）.约束

在安卓系统中运行

只能在手机上使用该app

网络条件良好

访问量限制在一定范围内

在3个月内完成app的制作，运行，改错。

（7）.逆向需求

无法实现加入敏感字体和语句的屏蔽

（8）.将来可能提出的要求

能够实时捕捉对话内容，实时反馈信息

能提取文字变成语音模式

能识别图片上的文字信息

1. **项目的环境、条件和限制**
2. **项目环境**

资金：资金充足

劳动力：劳动力充足并低廉

价格：项目投入大量人力，少量财力

劳动生产率：高

管理人员水平：高

政府政策：无

顾客需求：暂无

1. **项目条件**

项目周期：3个月

性能规划： 智能辅助聊天

成本预算：

根据杭州社平工资7000元每月计算，则工资支出成本为4\*7000=28000元。软件硬件基本每个人员都具备。小组team building吃饭成本平均每周1-2次，每次100元，则吃饭成本100\*25=2500元。

总成本：28000+2500=30500元。

1. **项目限制**

技术限制：专业技术的缺失

金钱限制：金钱投入有限

时间限制：时间投入有限

1. **项目团队建设**

1.SE2018春-G11小组全体成员

组长：苏碧青

成员：蓝舒雯 陈铭阳 刘乐威

2.项目分工

组织模式：管理层负责决策，各个分组负责执行，分组又分为长期分组和临时分组。长期分组包括核心组、技术组、数据组、文献组、宣传组；临时分组：数据服务项目组、课程服务组等；

各组都有明确的任务分工，几个主要分组的任务如下：

技术组：负责项目的核心技术，包括做爬虫和分布性系统。课题组自己搭建爬虫系统，支持私人定制，不经渲染，提高效率；使用分布式系统完成数据的分布式采集，储存和计算任务。课题组设计及使用的分布式框架考虑到爬虫的针对性，框架成品具有很强的拓展性，只需加一个硬件或模板即可以套用。

文献组：文献组主要看有关大数据的论文，基于这部分与大数据描述有关的文章去做一些描述性的结果。通过基本而全方面的描述看到数据的样子，从而探测数据背后的数据模型。文献组后期会和数据组合并。

数据组：主要用到的技术是机器学习，目前做的是数据挖掘中的文本挖掘。由于在线平台文本数据远多于数字数据，因此需要机器人理解信息分类，做情感分析（从评论），为之后了解人物定位的标准做准备，也可以应用于多平台内容推荐等。同时，未来从技术上支持项目组对线上平台的分析；相比单纯地爬虫和分布式更具有研究意义，影响更深远，对项目发展产生较大优势。