校名-小

# 大学生雏雁计划项目 立项申请书

项目来源：导师科研□ 自主探索类☑ 其他□

项目中文名称： 智能日记--基于自然语言处理和数据分析的在校学生个人记录与管理工具

项目英文名称： Smart Diary--personal record and management tool for school students based on natural language processing and data analysis

项目所属学院： 计算机学院

项目负责人： 李宇豪

联系电话： 15503682713

E-mail： 1461160060@qq.com

指导教师： 崔毅东

E-mail： cyd@bupt.edu.cn

起止年月： 2020年11月 至2021年 4月

填报时间： 2020 年 10 月 26 日

**填写说明**

1、本申请书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要。

2、申请书首页只填写项目负责人

3、“项目所属学院”：如项目负责人属于国际学院，填写“国际学院”；如项目负责人不属于国际学院，填写指导教师所在学院（请见下表）；如指导老师所在学院不在下表中，选择项目负责人所在学院。

|  |  |
| --- | --- |
| **序号** | **学院** |
| 1 | 信息与通信工程学院 |
| 2 | 电子工程学院 |
| 3 | 计算机学院 |
| 4 | 数字媒体与设计艺术学院 |
| 5 | 现代邮政学院 |
| 6 | 网络空间安全学院 |
| 7 | 理学院 |
| 8 | 经济管理学院 |
| 9 | 人文学院 |
| 10 | 马克思主义学院 |
| 11 | 国际学院 |
| 12 | 人工智能学院 |
| 13 | 图书馆 |

4、申请书填写完毕后，须在“北京邮电大学大学生创新创业计划训练平台”开放雏雁计划立项的期限内按时上传并提交。

**一、基本情况**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **项目名称** | 智能日记--基于自然语言处理和数据分析的在校学生个人分析管理工具 | | | | | | | |
| **项目来源** | **导师科研□ 自主探索类☑ 其他□** | | | | | | | |
| **项目负责人** | 李宇豪 | | **学号** | 2019211937 | **所在学院** | 计算机学院 | **手机号** | 15503682713 |
| **专业** | 软件工程 | **班级** | 2019211937 | **邮箱** | [1461160060@qq.com](mailto:1461160060@qq.com) |
| **指导教师** | 崔毅东 | | **工号** | 2010813054 | **所在学院** | **计算机学院** | **手机号** | 13810099495 |
| **职称** | 副教授 | **邮箱** | cyd@bupt.edu.cn |
| **项目性质** | **□硬件 ☑软件 □硬件+软件 □文化创意 □理论研究 □其他** | | | | | | | |
| **项目类别** | **□智能硬件、☑智慧生活、□数字娱乐、□通信网络、□医疗健康 、☑信息服务、☑人工智能、□教育文化、□房产家居、□公益创业、□理论研究、□其他\_\_\_\_\_\_\_\_\_\_\_\_\_** | | | | | | | |
| **项目成员**  **基本信息** | **姓名** | **学院** | | **专业** | **班级** | **学号** | **电话** | **邮箱** |
| 李宇豪 | 计算机学院 | | 软件工程 | 2019211501 | 2019211937 | 15503682713 | [1461160060@qq.com](mailto:1461160060@qq.com) |
| 唐良爽 | 计算机学院 | | 软件工程 | 2019211501 | 2019210010 | 18190326484 | 2809867884@qq.com |
| 陈可淇 | 计算机学院 | | 软件工程 | 2019211502 | 2019211955 | 13518822317 | 1047969038@qq.com |
| 侯瑞海 | 计算机学院 | | 计算机类 | 2019211318 | 2019211919 | 18801056912 | 1764126574@qq.com |
| **团队主要**  **成员介绍** | 李宇豪:  擅长领域:图形交互界面开发,熟练使用Qt,pyqt,Android,对交互产品的一般开发流程及框架有较深的经验;熟悉Java,C++;熟练python中网络爬取,游戏制作,数据可视化(matplotlin,QChart);  创新实践训练:暂无  重要获奖记录:无  唐良爽：  有良好的Java基础，能Java的GUI、多线程、网络通信、数据库等都有较为扎实的基础。并对Python网络爬虫较为熟悉，对Python常见的数据处理库如Pandas，Numpy，Matplotlib等有一定基础。Android开发已经入门。暂未参与过创新创业实践训练，暂未或得重要奖励  陈可淇:  擅长领域:算法设计与实现，有良好的C++,Java基础，熟悉多线程和GUI。有一定的产品设计经验。 创新实践训练:软件学院创新创业卓越人才试点班 重要获奖记录:无  侯瑞海:  擅长领域:算法设计与实现,对深度学习有一定的了解,熟练掌握C++，python编程，数理基础较好,英语水平比较高（可以阅读英语论文）;  创新实践训练:暂无  重要获奖记录:无 | | | | | | | |

1. **研究主题（建议1500字）**

**背景分析:**

人的行为习惯有着周期性,相关性等许多规律,对于生活比较单调的大学生更是如此.然而我们却缺少对其准确,科学的认识和管理.

本项目的目标是将自然语言处理(NLP)技术和日记结合起来,同时引入新的表单的记录方式(专门记录一些特征数据项如:学习时长,睡眠时间,锻炼量;),通过移动端的app方式提供这样一款产品.

在用户完成数据输入(即日记文本和表单数据)后,本产品可以首先对表单进行基本的数据分析(包括周期性,概率分布,相关性及趋势等),然后通过NLP进行更进一步的分析,包括:

1.情绪分析----可以将分析出的情绪结果作为一种数据项加入到前面的数据分析;

2.文本推断----在用户以后进行文本输入时给出提示,简化输入;

3.摘要生成----生成一周/月时间内的概要,便于用户浏览;

4.信息抽取----提取用户记录中反应的某些特定的模式(行为,时间之间的关系等);

5.特征词提取--提取具有出现频率等特征的词,并将它们加入到第一步的基本数据分中;

6.信息检索----输入关键词快速查找与之相关的日记内容;

在完成了上面的分析过程后,我们就对用户的行为习惯的规律有了较为充足的了解,然后将这些分析结果通过文本结合可视化的报告反馈给用户,用户(尤其是学生)可以通过这些结果更好的了解自己的学习,娱乐及健康状况,从而能结合自身的规律做出更合适的时间规划和精力安排.从而达到增加学习工作效率的目的.

**用户分析:**

以大学在校本科生及研究生为首要用户角色分析,得以下结果:

1.活动:较高工作量的学习(主要是上课,自习)和娱乐活动(外出游玩,社交媒体,游戏及其他),体育运动(健身,球类运动,跑步)分布在不同时间段

2.态度:认为日记是记录工具,记录生活中情感,日常活动.大部分人不做日记,用来记录日常活动3.或者生活记录的不多.

4.能力:本科教育水平,学习能力较高.

5.动机:大部分都对现有的日记产品不满意及不了解,有提升自我,提高学习效率,让生活更加健康和有条理的需求.一部分愿意了解自身规律.

6.技能:熟练的手机操作能力,大部分熟练理解一些统计知识,能看懂统计图表

**需求分析:**

1.在每天结束前(睡觉)前通过完成表单及文本记录的方式记录一些特征数据

2.可以随时以表格/文本方式查看并修改以往记录的数据.

3.可以随时查看根据记录的数据得到的分析结果

4.在记录文本日记内容时提供文本提示,简化输入

5.输入关键词快速查找与之相关的日记内容;

6.在有需要的时候自定义要记录的特征数据,添加到要完成的表单中

7.产品会在一定的时间间隔以报告的形式给用户反馈其特定时间段内的分析报告,图表,预测及建议

**同类竞品分析:**

1.每日记:

功能点:单纯以文本(会员有图像)的方式记录日记,还有笔记本,记事本,备忘录的功能, 主要功能是记录文本内容.

特点:个性化设置信纸和字体

2.时光手帐:

功能点:记录文本和图片,可以创建多个日记本,外加社区功能,可以访问别人的日记空间.

特点:清新界面设计,丰富图片素材库制作美观的日记,社区功能.

**创新性:**

1.解决方案:将自然语言处理和日记整合在一起,同时结合表单式的新的记录方式,日记不再是一种一维的记录载体,它可以包含更丰富的信息,给用户一种新的个人记录及管理需求的解决方案.

2.用户:发现大学生群体在自我发现和管理方面这方面需求未被满足的群体.

3.体验:通过自然语言处理的结果加上数据可视化,给用户直观,更进一步的服务.

**可行性:**

1.记录方面:基于android开发,数据库可以实现.

2.分析方面:基于概率论的知识可以完成比如周期性,相关性的分析.基于NLP尤其是贝叶斯分析方法,实现文本挖掘和分析。具体方式是jieba分词后，根据不同词语的分布，结合其出现日期、出现频率等信息，采用贝叶斯公式以及其他相关概率模型分析其中的相关性，从而得出每个词语的特征量，根据特征量再对相应功能进行实现，如人性化提示等。在可视化方面采用MPandroidChart等第三方Android绘图库绘图库进行实现.

**实用性:**

1.带给用户的价值:生活中有很多场景需要记录和比对一些数据,比如健身计划,英语的单词记忆和自习时长都可以丰富用户对自身的了解,如果我们知道了它们的规律就可以更好的做出符合自身习惯的日程表.基于手机这样的移动平台,使用足够简洁易用的交互界面.我们跟考虑到了智能分析带来的私人定制特性(如个性推送)在当下应用产品的宽泛应用.

2.用户的成本:用户需要做的是记录文本内容,同时填一张表(有些项可以自动完成填报),每天总耗时不超过5分钟,没有难度过大或者过于繁琐的操作,免费下载,综合来说成本很小.

1. **研究方法（建议500字左右）**

本项目采用调查法、文献研究法、探索性研究法、实证研究法来解决关键问题，形成不同的解决方案。

1.首先本团队就大学生的时间管理就行调查，调查样本为北邮在校学生，样本大小为50人，采用问卷的形式，获取50人每天时间的安排以及其自我感觉时间安排的合理性。从得到的答案来看，样本中68%同学觉得自己时间安排得不合理，10%的人不关心，22%的人对自己的安排感到比较满意。而本项目的根本目的就在于帮助大学生合理高效地安排和规划时间，所以本项目存在的意义得到了肯定，且本项目具有较强的应用性。

2.在技术实现上采用的是文献研究法和探索性研究法。由于本项目最终形态为具有指定功能的APP，所以实现过程中必须用到一系列软件技术。使用文献研究法，通过调查文献来获得资料，以获得技术支持，从而完成项目的各个功能模块的设计，最终开发出一个功能强大而完整的APP。

3.本项目立志于创新，要与市场上现有的同领域产品有较大区别度，并要比其他产品性能更优、功能更强、创新性更佳。所以本项目会采用探索性研究法，利用已有的样本数据和实现技术，进一步探索、创新新功能、新技术，在理念和技术上实现双重创新。并且在产品的各个研究阶段，都会采用实证研究法来检验产品是否打到了预期效果。具体做法是随机抽取一定数目的同学来体验本产品，并根据其评价进行相应的修正。以上研究方法相结合，便能形成多方位的解决方案。

1. **项目创新点（建议1000字左右）**

**设计亮点：**

1.采用文本结合表单的方式记录日记,将日记从一维载体拓宽为多维,使得日记可以包含更丰富的内涵,实现更多的功能.

2.将NLP和日记文本结合,充分的发挥了日记的文本格式的优势,增加的日记的可能性和潜力(情感分析,文本推断,概要生成,模式提取)

3.日记不仅是生活内容的记录着,在其中更蕴含着人们深层次的行为习惯规律,本产品发现了这一点并将其显示给用户

4.在数据显示方面,除了传统的文本,更增加了可视化这样直观,易懂,简洁的显示方式

**技术亮点：**

1.不仅仅是单纯的Android开发应用，而是采用Andriod与Python的NLP自然语言处理第三方库结合的模式。本项目的最终形态是APP，采用Java语言与Python语言并用进行APP的开发。在Android studio开发工具里安装Chaquopy插件以实现对Python语言的支持，用Java语言编写App的总体框架，用Python编写具体NLP自然语言处理方面的代码。

2.采用的NLP技术包括jieba分词，RNN实现文本生成,贝叶斯方法分析主题,结合上下文**，采用神经网络中的注意力机制，进行情绪分析,结合使用**TextBlob,Gensim,及Java的HanNLP实现信息抽取;

3.每日记录数据的规律分析：运用不同的数据分析方法以及数据模型分析其中的相关性，趋势，分布，对比，演变，聚焦，指标等数据规律。用户可以自定义选择要分析的数据项，并且对其进行新的赋值，分配权重或者偏振量函数进行更细致的分析。

**商业模式：**

免费下载，付费获得更丰富的可视化，用这种方式进行商业化尝试。

1. **项目目标及进度安排**

**前期安排及准备:**

2020年11月--2020月12月:Android开发熟练运用

2020年12月--2021年1月:自然语言处理知识的运用,一些数据分析方法

**中期安排进度:**

2020年12月--2021年1月:实现Android交互框架;

2021年:

1月--1月中旬:完成可视化内容;

1月中旬--2月中旬:数据分析功能实现;

2月--3月:完成自然语言处理功能;

**后期安排:**

3月--4月:宣传工作;

**2.关键时间点:**

2021.1.1:

对应任务:完成交互框架;

考核指标:软件的界面齐全,即1.可以实现基本的数据记录功能;2.可以自定义记录项;3.可以通过文本及表格的形式常看并修改以往记录的数据;4.有基本的登陆,密码验证及修改功能;5.可以跳转到可视化及数据分析结果界面

2021.1.15:

对应任务:完成可视化功能:

考核指标:完成了可视化内容,即1.可以通过不同的图表方式显示某个记录项的数据;2.可以将不同的记录项功能显示在同一表格进行对比显示

2021.2.15:

对应任务:完成了数据分析功能:

考核指标:1.软件可以显示记录项的一些规律,比如趋势,分布,周期及相关性;2.用户可以指定分析的数据内容和分析目标,比如分析指定数据项的相关性,指定数据行的某个数据特征

2021.3.1:

对应任务:完成了文本自然语言处理功能:

考核指标:1.可以对记录的文本进行分析,进行特征的提取;2.可以生成词云;3,可以结合之前的数据项分析和待选数据中

2020.4.1:

对应指标:成品交付:

考核指标:1.完成了计划书中的描述;2.可以在手机应用商店中下载并正常使用

**五,已有基础**

1.已有相关研究积累和已取得成绩：

通过搜索市场上已有的日记类App，记录下了每个App的功能，对比发现这些App的功能均局限于写日记，且只能通过用户打字输入方式进行日记记录。而通过问卷形式调研了50个北邮学生，发现其中大部分人对时间管理有需求，所以这正是本项目的立足之处。

2.已具备的研究条件：

组内成员都会Java和Python，并都能较为熟练地应用Java和Python进行实际项目的开发。并且都具有一定的Android基础，通过一定时间的学习能快速掌握Android开发知识并应用于项目的开发。

3.尚缺少的研究条件和解决方法：

组内成员对人工智能相关知识，如贝叶斯算法，NLP自然语言处理等虽有一定了解，但还不是很深入，将人工智能算法应用到实际项目的开发中还不是很熟悉，需要在开发过程中边学边用。

**六,预期成果**

预期成果以Android APP的形式展现。该App不仅具有写日记的功能，还具有数据可视化、文本情感分析、识别手写日记的功能。可以在应用商店中正常下载并使用.

三、成员分工

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 | 项目前期调研学习情况 | 任务分工 | 预期投入精力 | 签字 |
| 李宇豪 | 熟悉Java和Python，精通Qt开发，并对Android有较好的掌握,熟悉GUI开发的一般步骤,对自然语言处理有一定了解 | 编写App框架和界面UI,实现数据可视化,辅助进行数据分析,文本分析,对项目进行管理. | 300小时 | 李宇豪 |
| 唐良爽 | 熟悉Java和Python，对GUI开发有经验，对Android有一定了解 | 编写App框架和界面UI,实现数据可视化,辅助进行数据分析,文本分析 | 300小时 | 唐良爽 |
| 陈可淇 | 熟悉Java和Python,数据结构与算法又比较良好的掌握 | 主要进行进行数据分析,文本分析 | 100小时 | 陈可淇 |
| 候瑞海 | 熟悉Java和Python，对GUI和算法都有很好掌握,熟悉概率论的知识与方法,对深度学习和自然语言处理有良好的了解和运用 | 主要进行进行数据分析,文本分析 | 100小时 | 候瑞海 |
|  |  |  |  |  |

注：包括项目负责人