**哈 尔 滨 理 工 大 学**

**毕业设计开题报告**

**题 目： 基于机器学习的英语单词智能打卡系统**

**院 系：计算机科学与技术学院数据科学与大数据技术系**

**姓 名： 马超**

**指导教师： 杨锦锋**

**系 主 任： 马超**

**2022年10月**

# 一、选题的依据、意义，理论或实际方面的价值

## 1.选题的依据和意义

美国著名词汇学专家诺曼·莱文斯曾说：“词汇是思维的工具，词汇是思想的武器。我国著名语言学家桂诗春研究认为词汇量和语言水平有密切的关系：词汇量大的人不一定语言水平高，但是语言水平高的人都有较大词汇量；他指出词汇量测验和智力测验或其他成就测验的结果具有很高的相关系数。 由此可知，单词记忆是英语学习最基本的要求。学习者必须扩大英语词汇数量才能使听、说、读等英语综合能力水平得到提升。学习者借助单词打卡系统，可以随时随地地进行单词记忆。软件根据学习者的学习记忆情况，综合分析词汇遗忘临界点，并在这个临界点时间到来时智能地提醒学习者，将短时记忆转化为长久记忆，提高学习效率。

本选题“基于机器学习的英语单词智能打卡系统”中的两个关键词为“机器学习”与“智能单词打卡”。

机器学习是一门多学科交叉专业，涵盖概率论知识，统计学知识，近似理论知识和复杂算法知识，使用计算机作为工具并致力于真实实时的模拟人类学习方式，并将现有内容进行知识结构划分来有效提高学习效率。机器学习是研究怎样使用计算机模拟或实现人类学习活动的科学，是人工智能中最具智能特征，最前沿的研究领域之一。本系统在打卡任务生成、打卡评判两部分应用到机器学习、自然语言处理等技术对打卡任务进行处理。

本研究实现了基于移动端开发的系统进行英语词汇的学习，为建立以学习者为中心、打卡学习以及自我控制学习进度的方式探索出一条新途径。其次，本文通过移动端为学习者提供基于具体语境的初级词汇打卡学习，并使用词义之间的聚合将词汇进行分类，有助于学习者构建系统的词汇结构，使学习者能够达到更好地学习词汇的目的。软件推出后，使学习者可以选择不同方式进行英语词汇记忆学习，提高了词汇学习的便捷性和学习记忆效率。应用大数据对抗记忆曲线，从背单词的效率分析出遗忘曲线，再根据曲线规划记忆时间，安排单词出现概率，保证背单词的最佳效率。教师端根据单词遗忘曲线程度安排考核内容，提高学生背单词效率以及教师工作效率。

## 2.理论或实际应用方面的价值

（1）理论意义

随着移动互联网技术的不断发展，基于移动智能终端的移动学习逐渐被广

大学习者所接受，通过基于手机端的词汇学习平台设计与实现的研究，可以进一步明确移动学习的内涵，有助于移动学习理论进一步发展与完善。

（2）实践意义

首先，本研究实现了基于移动端进行英语词汇的学习，为建立以学习者为中心、自我控制学习进度以及个性化学习的方式探索出一条新途径。其次，本文通过移动端为学习者提供基于具体语境的初级词汇学习，并使用词义之间的聚合将词汇进行分类，有助于学习者构建系统的词汇结构，使学习者能够达到更好地学习词汇的目的。最后，本系统主要针对二年级以上大学以下学生，对于中小学生高效率的提升词汇量，具有一定的现实意义与价值。

# 二、国内外研究现状

随着时代的进步，互联网移动网络技术的不断发展和创新，以及教育技术

的改革和深化，各类移动学习方式和平台的涌现、教育培训组织、学习类、工

作类、生活类软件的应用，促使技术人员不断开发、更新换代移动软件，以满

足人们生活、工作和学习需要。英语作为在国际上最主要的交流语言，国内学

生学习和备考的重要学科，教育的策略和学习的方式方法也受到影响，词汇的

学习是学好一门语言的基石，在现代网络发达的社会，在英语词汇的学习过程

中，移动学习成为一种主要的方式。在这样的社会背景下，移动知识学习的改革正在实际的教学教研活动以及相关领域中发生着。

移动学习在教育上的应用早已被教育科研人员所关注，移动语言学习成为移动学习研究的重要内容，2000 年斯坦福大学学习实验室率先启动了一项有关语言学习的移动学习项目。他们通过手机这种移动的方式对西班牙语的学习开展研究，结果证明，学习者通过使用手机的功能可以使得学习语言的效果明显提高。

传统的单词学习工具主要有单词手册、课本、以及词典等纸质资源。这些工具的共同点是学习者在学习的过程中，只能按照固定的单词顺序线性的学习，不能与这些资源进行任何交互，更不能对学习者的学习特征作出相应的调整和反馈。随着移动学习软件的广泛应用，各大主流用用商店上出现了大量的英语单词移动学习软件，对这些软件按功能特点进行分类，可划分为“记单词”、“单词助记”、“记忆管理”三个层次。“背单词”属于最低层次，这种软件仅仅是将纸质的词典电子化；“单词助记”则利用了一些辅助信息或媒体帮助学习者对词汇进行有效记忆；而“记忆管理”层次最高，能够对大量单词记忆之后按照遗忘规律进行精确的复习安排。

国外移动语言学习的研究以美国斯坦福大学建立了全球第一个移动语言学

习实验室为标志。他们首先通过手机对西班牙语的学习进行了研究，其研究结

果表明学习者借助手机的语音信箱功能与E-mail功能够明显地提高学习者的语言学习效果。随后移动语言学习的研究在世界范围内展开。在日本，Thornton和 Houser开展了一项关于词汇教学的实践性研究，研究者们通过手机短信向学习者推送有关课程的学习内容，其课程学习内容推送频率为一天三次且内容简明扼要，研究结果表明通过手机短信进行词汇学习的学习者成绩明显高于采用传统课堂学习的学习者。

随着移动互联网技术的发展与进步，目前，国外对移动语言学习的研究领域更多地注重对移动智能终端开展语言学习方面的研究，尤其是语言听力方面的学习研究更为突出，这也是世界范围内的共同研究趋势。

移动语言学习作为一种不受时间与空间限制的信息化学习方式，不仅能够为学习者制定符合自身的个性化学习方案，而且还可以将传统课堂学习活动延伸到生活当中，与此同时也能够给予学习者及时的学习反馈，对学习者学习效果的提升起到了促进作用。正是由于移动语言学习方式在语言学习方面具有诸多的优点，才被人们视作辅助语言学习的首选学习方式。随着技术的不断进步与成熟，移动语言学习与英语学习之间的联系将越来越紧密，其发展前景将会越来越广阔。

# 三、课题研究的内容及拟采取的方法

## 3.1 系统开发环境

操作系统：windows11

数据库：Mysql关系数据库

使用语言：Python、Vue.js、node.js

使用框架：ElementUI

开发软件：Webstorm、Pycharm、Navicat、Postman

## 3.2 需求分析

#### 3.2.1 功能需求

该毕业设计是开发一个智能单词打卡系统，该系统主要有四个功能模块，分别是注册登录模块、打卡任务模块、遗忘曲线分析模块、班级权限管理模块。用户通过注册获得账号，然后进行登录，可以进行个人信息修改，系统共划分两类身份，分别是学生、教师。在打卡任务模块中主要功能有教师发布打卡任务，查看学生完成情况，班级成员权限管理。在遗忘曲线分析模块中，主要的功能有根据学生单词记忆情况分析出记忆遗忘曲线，根据曲线遗忘程度着重安排单词出现频率，最大程度的提升背单词效率以及单词的记忆程度，教师根据打卡任务完成情况，单词错误率安排个别单词考察，提升工作效率。班级权限管理模块中，学生可以搜索班级号加入班级，退出班级操作。教师可以将学生加入到自己的班级，也可以将学生移出班级。例如教师注册账户后可以创建多个班级，每个班级有唯一的班级号，学生可以搜索班级号加入班级，也可以通过教师添加的方式进入班级。根据需求，制作系统功能模块划分结构图，如图3.1所示：



图 3.1 系统功能模块划分结构图

各个模块功能如下：

1. 注册登录模块

用户通过注册获得账号，通过账号进行登录使用系统，用户登录后要进行身份认证，通过账号密码等信息验证学生或老师的身份，不同的身份所展示的效果不同，不同的身份都可以进行个人信息修改。

1. 打卡任务管理

以教师身份登录单词打卡系统后，可以选择班级发放打卡任务，以学生的身份登录单词打卡系统后，可以对老师发放的单词打卡任务进行完成，任务完成后后台自动评判产生成绩。

1. 遗忘曲线分析模块

以教师身份登录单词打卡系统后，可以学生的班级权限管理，根据学生单词完成情况，重点考察错误率较高，易错的单词，以学生的身份登录单词打卡系统后，单词每隔一段时间会再一次出现，根据完成情况系统分析出遗忘曲线，根据遗忘曲线安排后续单词出现的频率，最大程度上保证背单词的效率以及成功率。

1. 班级权限管理模块

在这个模块中，参与者是老师和学生，教师可以创建班级，添加学生，移出学生。学生可以选择加入班级，，退出班级。事件会同步在一个关系内的用户群体中。

#### 3.2.2 非功能需求

1) 性能需求：用户在软件响应速度、结果精度、运行时资源消耗量等方面的要求。

2) 可靠性需求：用户在软件失效的频率、严重程度、易恢复性，以及故障可预测性等方面的要求。

3) 易用性需求：该系统操作简便，人机界面简单明了，能一目了然的清楚需要做什么，该如何操作。不必借助任何操作手册或相关的系统帮助就可顺利进行各种操作。具有很强的易理解性、易学习性和易操作性。

4) 运行环境约束：用户对软件系统运行环境的要求。

5) 外部接口：用户对待开发软件系统与其他软件系统或硬件设备之间的接口的要求。

6) 安全性需求：该系统中涉及用户的重要信息，不同的用户具有不同的使用权限，只有具有最高权限的系统管理员才可以对系统进行修改，具有一般权限的用户只能读取自己相关信息，不能浏览其他用户信息。系统还要提供方便的手段供系统维护人员进行数据备份以及系统意外崩溃时数据的恢复等工作。

7) 可保障性需求：用户在软件可配置性、可扩展性、可维护性、可移植性等方面的要求。

## 3.3. 需求模型

1) 学生具备的功能有：登录、注册、个人信息修改、完成打卡任务、加入班级，退出班级。

2) 老师具备的功能有：登录、注册、个人信息修改、发布打卡任务、查看打卡任务完成情况，创建班级，拉取学生，移除学生。

学生老师用例图如图3.2所示：

3）系统用例图如图3.2所示：

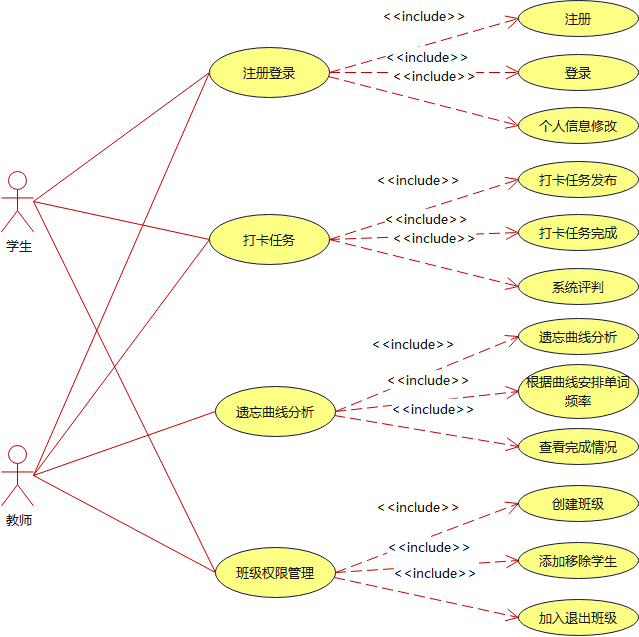


图3.2 学生老师用例图

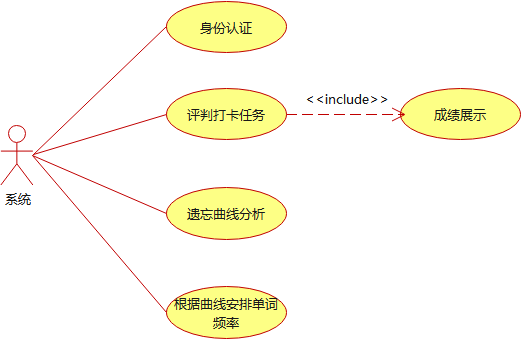


图3.3 系统用例图

# 课题研究中的主要难点及解决办法

## 4.1 现有背单词App功能较完善

互联网技术的不断发展与普及以及教育教学的不断改革与深化，移动互联网技术在教育领域也得到了广泛的实践与应用。随着技术的不断进步，新的教育与学习方式也随之应运而生，以满足现代化人们的生活与学习需求。英语作为重要的交流工具与教育事业中学习的重点学科，其教育与学习的方式同样受到技术发展的影响，无论是科学研究还是工作与生活对于英语方面的人才需求很大，其高水平的英语人才在社会的各个方面需求更是炙手可热。移动语言学习应用于相关学科的教学活动的教育改革正是在这样的社会背景下应运而生的。手机上下载这些App之后，我们便可以利用碎片化的时间进行学习，而不必时刻抱着一本厚厚的单词书，不必每次打开单词书总是abandon，这些软件会根据我们的水平和需求提供相应的单词的练习。市面上现有的背单词软件发展较完善，发展空间比较小。但是现有软件绝大部分都是付费机制，相对于付费，免费使用更能吸引人。而且本系统应用单词打卡模式，教师学生两端相配合。教师负责发布打卡任务，学生可以自行完成，家长也可以起到监督的作用，背单词的任务会完成的更好，更有效果。而且我认为这种打卡鼓励机制可以使我们更动力背单词，在一个人背单词的情况下，坚持的时间也是一个问题，这种鼓励机制和家长老师的双重监督，不会存在中途放弃的情况。而且本系统主要针对于二年及以上大学以下的小中学生，受众群体不同。市面上的软件绝大多是针对于高等教育人群，以及应试人群，少有软件主要针对中小学提出背单词的解决策略。本系统所用鼓励机制以及打卡任务形式双重模式，针对中小学生，会更好的形成约束力，久而久之，背单词形成一种习惯，可以脱离监督。我相信这种打卡鼓励机制会有一定的前景以及市场空间。

## 4.2 对于机器学习与自然语言处理算法的深入研究

艾宾浩斯遗忘曲线在学习和巩固新知识的作用十分显著。目前，大学生大多是使用市面上的英语学习App和小程序或是单词书对单词进行的学习，大部分应用的设计主要将重点放在对英语单词的前期学习上，尽管设计的页面生动活泼，但却忽视了自动提示学习者复习方面的设计。词汇学习的目的应该是牢固地记忆，而不是学会了就放任不管，不加以重复的学习是无效的，学习者在学习过后不注意初学后的及时复习，会导致在学习单词的过程中发生“容易忘”的情况。而艾宾浩斯遗忘曲线就说明了人类在学习新知识后对知识内容的遗忘速度是呈现“先快后慢”的趋势的，所以，应该在学习后对知识进行反复的学习，这样才能使知识在脑海中停留更长时间。将艾宾浩斯遗忘曲线作为理论指导开发英语词汇学习的移动平台是很有必要的，它将自动提醒学习者在合适的时间对单词进行学习和复习，解决“容易忘”的问题，提高学习效率。

艾宾浩斯遗忘曲线描述了人类大脑对新事物[遗忘](https://baike.baidu.com/item/%E9%81%97%E5%BF%98/17196693?fromModule=lemma_inlink" \t "_blank)的规律。人体大脑对新事物遗忘的循序渐进的直观描述，人们可以从遗忘曲线中掌握遗忘规律并加以利用，从而提升自我记忆能力。该曲线对人类记忆认知研究产生了重大影响。所以根据曲线进行单词巧计，得出一次性大脑可以接受的数量，安排复习时间点，对于背单词效率的都是十分重要的。

每个人的遗忘曲线其实是不一样的。通过机器学习算法，对单词完成情况进行分析，分析出遗忘曲线，结合艾宾浩斯遗忘曲线的遗忘时间点，一次性能背单词的数量，复习的时间单词频率以及时间点的安排，合适的时间对单词进行学习和复习，提高学习效率。

自然语言处理被定义为一个使计算机能够阅读，理解和提取人类所说自然语言含义的人工智能领域。

有了自然语言处理，计算机不仅能够理解自然语言，而且还能通过自然语言对人类做出回应。除了用于人机交互目的之外，NLP技术还广泛用于处理公司的大量数据，而手动处理则需要数月甚至数年的时间。

例如，一个人可能需要花几周的时间来从一本一千页的书中人工提取信息进行总结，有了自然语言处理，类似的文本摘要可以在几分钟内完成。同理，聘用人工读者通读所有用户评论，并编写包含用户对特定产品的意见的报告是花费成本非常高的，然而有了自然语言处理之后，用户意见可以被自动从文本中提取，从而使公司能够制定出组织决策。

自然语言处理在打卡任务完成时会有所涉及。例如文字转语音的开发，打卡需要根据读音默写单词，或者是根据单词进行读音判断对错。以单词的学习与记忆，把词法、句法、句式连起来，实现英语的有效学习。以训练营每日打卡的模式使学生在每日坚持中强化学习和记忆。以最少的经济成本和最少的时间成本获取最好的英语学习效果。培养良好的学习习惯和自我驱动的学习能力。

## 4.3 对于python与机器学习算法学习

对于python，机器学习和自然语言处理这几个技术是比较陌生的，以及python机器学习代码嵌入到node.js中未曾实践过，所以需要一定的时间去学习巩固并应用，采用看官方教程的学习方式进行学习，并结合GitHub开源代码进行学习，早日达到可以进行开发的水准，顺利完成毕业设计。

# 五、进度安排

2021.09.13-2021.09.20:确定论文主题方向，进行论文题目的筛选，及题目最终的确定。并撰写论文任务书。

2021.09.20-2021.09.25:以论文题目为核心，查阅相关咨询系统原理和理论知识，了解国内外单词打卡程序研究现状，在现有基础上，开发单词打卡系统。查阅机器学习可以在单词打卡领域的应用，了解其算法，查阅论文。

2021.09.25-2021.10.20:对已搜集的资料加以整理，进行市场分析，收集需求数据，确立系统内容。通过对当今市场上的单词打卡系统内容与形式整理出相关流程，并分析用户对单词打卡系统的不同需求，论证分析论文的可行性、实际性，将论文题目和大致范围确定下来，进行开题报告。

2021.10.20-2021.10.27:整合已有资料、构筑论文的大纲。

2021.10.27-2021.11.05:概要设计阶段，进行总体设计、模块设计。

2021.11.05-2021.11.20:详细设计。细化单词打卡系统的功能，如视频查看页面、问题显示页的调节字体大小、发布功能、下载功能等，同时设计系统各个接口，确保各接口连接正常。

2021.11.20-2021.12.11:编码设计，本系统主要运用node.js，python编程语言，MySQL作为数据库开发工具，搭建开发环境，并对论文编写过程中所发现的问题，研究其解决方案，推敲整合，并进行修改完善，准备论文中期检查。

2021.12.11-2022.03.05:继续编码，并进行测试。

2022.03.05-2022.04.02:完成论文的初稿部分，向指导老师寻求意见，优化论文的结构，润色语句，修改不当之处，补充不足之处。

2022.04.02-2022.05.21:调试并优化程序，论文资料整合，最终定稿，为最终的答辩做好各方面准备，熟悉论文内容，增强自己对论文内容的把握，进行一定的思维发散，设计论文答辩。

2022.05.21-2022.06.20:准备答辩。在确保程序无误，以及修改后的论文无误后，制作PPT，准备答辩。