

附件 5

湖南省大学生研究性学习和创新性实验计划

项 目 申 报 表

项目名称：局部人工干预的半智能化编程语言初学者导引平台				
学校名称	长沙理工大学			
学生姓名	学 号	专 业	性 别	入 学 年 份
罗丽美	201516080103	软件工程	女	2015
陈泽宇	201516080210	软件工程	男	2015
陈武阳	201516080204	软件工程	女	2015
李思超	201554080322	通信工程	男	2015
赵一霖	201515020123	网络工程	男	2015
指导教师	叶吉祥	职称	教授	
项目所属一级学科	计算机科学与技术	项目科类(理科/文科)	理科	
<p>学生曾经参与科研的情况</p> <p style="padding-left: 20px;">多维度定位策略的移动监互系统创新项目。</p>				
<p>指导教师承担科研课题情况</p> <p>多生理信号融合的中医证候研究，湖南省自然科学基金, 主持。</p> <p>多生理信号融合的中医证候诊断平台研究, 湖南省科技厅科技计划项目，主持发表 CSCD 及 EI 检索科研论文数十篇</p>				

项目研究和实验的目的、内容和要解决的主要问题

项目研究和实验的目的：

随着互联网步步深入人们的生活中，越来越多的人对编程语言产生兴趣，希望能将编程语言应用到自己的生活中，这是互联网时代进化的趋势。但他们在学习过程中常常遇到以下问题：

- (1) 面对庞大知识体系不知如何下手
- (3) 具体语言知识点理解难
- (4) 知识点知识运用难
- (5) 编程问题描述难
- (6) 编译、运行错误提示理解难

而我们前期在各个编程学习网站上进行调查，发现并没有任何网站完善的解决了上述问题，提供的学习路线和学习方法都几乎一致，因此，我们将在本项目中进行改进。

项目研究和实验的内容：

为解决上述研究目的，结合计算机语言初学者的学习特点，本项目将从下面几个方面展开研究：

总体概括：

- 1、通过交互的方式形象地引导初学者了解计算机编程体系，并对一个项目的制作有一个初步的了解。让初学者能够先形成计算机编程体系的思想，以便进一步确定并选取自己学习的方向，不再是盲目地学习一些应用性不强的知识。比如网页分为前台和后台，前台由哪些元素组成显示出网页的界面，后台由哪些逻辑方法构成网页的功能。
- 2、进一步交互式引导初学者选择自己感兴趣的方向。针对这个方向再次进行编程语言的介绍，让初学者先了解语言的应用方式，大致知道一门语言的应用范围是什么，学了之后能有什么用？让初学者一开始能对这门语言感兴趣以及大概了解它的体系。
- 3、创新引导初学者学习编程语言。我们对初学者的主要特色引导在于：以编程语言运用小实例来驱动学习。小实例包括实现简单的游戏、完成某一功能的编写等等。在小实例的学习过程中，初学者能够进一步巩固编程语言基础以及语法语义的掌握。即让初学者能够在每一步的学习中都能形象地了解自己学习的知识点的应用层面，化抽象于形象，也可以增加初学者学习编程语言的兴趣，让初学者大致了解自己学习的程度，即可以写出什么样的应用。

具体内容：

- (1) 对初学者进行智能化引导
 - 1) 引导初学者了解计算机体系

- 2) 引导初学者了解各个编程语言简介
- (2) 新型的编程语言学习方法
 - 1) 通过趣味性的知识表现形式让初学者理解知识点。
 - 2) 实例操作让初学者熟悉编程语言的具体知识点和语法应用。
 - 3) 各个知识点之间的联想处理，实例与知识点间的联想处理。
- (3) 模糊查询
 - 1) 搜索问题的分词处理以及数据库数据的分词处理。
 - 2) 索引表的建立
 - 3) 依照权重匹配结果
- (4) 建立错误库
 - 1) 常见编译运行时错误的解决方案
 - 2) 常见程序 bug 解决方案
- (5) 语音搜索
 - 1) 提供语音搜索功能，实时将语音转化成文本显示
- (6) 在线 OJ 系统
 - 1) 提供在线代码调试

项目研究和实验的要解决的主要问题

- (1) 初学者学习的新型引导方式
- (2) 系统自我学习：自主知识扩充算法
- (3) 在线 OJ 系统的安全问题
- (4) 全文检索技术

国内外研究现状和发展动态

主要编程语言学习平台有：

计蒜客、慕课网、开课吧、北风网、萌码网、极客学院、尚硅谷、java1234、实验楼、51CTO、百度传课、菜鸟教程
Fenby、Codecademy 、 Quara、w3school

总体优点在于：

- 1、IT 在线教育产品有很好的未来，它将资源最大化利用，并省去了时间空间上的限制。
- 2、资源广，包括各个学校的讲课资源以及各种人员分享的视频资源。
- 3、知识涉及面广，除了计算机编程语言学习之外还包括各种行业类型的知识介绍视频。
- 4、相对于线下教育来说，学习者的成本大大降低。
- 5、部分网站提供了交互式编程的学习模式：主要是提供了在线 OJ 系统，动态给出编译结果。
- 6：大多数网站提供了用户问答板块。
- 7：极少数网站有特色的学习模式：以一个实验为一个学习单元，远程桌面连接云主机，直接提供在线编译环境。
- 8：少数网站的学习模块以一个职业路径为驱动，并提供了相关的学习资料。

总体缺点：

- 1、传统的目录式学习让初学者比较迷茫，难以掌握自己的学习程度与进度。
- 2、枯燥的知识点阅读学习方式，容易降低初学者对编程语言的学习兴趣。
- 3、在线编译器提供的错误提示对于初学者而言难以理解。
- 4、网站并不提供语义化搜索功能，不能针对性的由问题出发引出知识缺陷。

总结：

我们发现要做好一个编程语言学习网站，首先得面向初学者有一个简单生动的初步教学，但进一步，必须得做到教育质量的高品质，且对知识的讲解要深入化才能使学习者真正掌握。最重要的一点：学习的精髓在于提出问题与解决问题，而不是一味灌输知识。即目前的学习网站并没有从初学者角度出发，初学者仍旧感觉不知从何下手。

人工智能被适当的引入到学习平台的研究中，是当前的研究的最新动态。具体体现如下：

一、适度干预对初学者进行学习引导，解决问题描述模糊的难题。

二、通过自学习能力的实现使学习平台具有自我提升的能力。

三、通过描述知识点之间相互关联关系，使平台具有智能化的联想能力：编程语言知识点之间进行连接，知识点与易错点之间进行连接，知识点与应用实例进行连接，建立成横向纵向相互联系的知识体系，这样做到把碎片知识串联，形成整体体系。在学生学习的时候对其主动引导，才能较好掌握。

四、我们提供编译错误查询以及在线 OJ 系统，找到这种错误最可能的原因，也给出相应知识点，解决编译错误提示理解难问题。

本项目学生有关的研究积累和已取得的成绩

罗丽美：获得三等奖学金，英语过四级，获得优秀学生干部称号，获得英语校级竞赛奖，熟悉 C, C++, C#。熟悉 html, css、javascript、bootstrap。

陈泽宇：获得二等奖学金，英语过四级，获得五四优秀个人称号，熟悉 C, C++, C#, MySQL, 能熟练应用 Java 编写项目。

陈武阳：获得一等奖学金，与三好学生称号，英语过四、六级，获英语校级竞赛奖，熟悉 html, css, javascript 语言，能够写出交互式页面以及各种效果。C, C++ 掌握完善，熟悉 Java，动手能力强。

李思超：英语过四级，熟悉 Java, 动手实践能力强，编程能力强。了解 html, css, 熟悉 C, MySQL。

赵一霖：第七届蓝桥杯国家级二等奖、省级一等奖。掌握 JavaSE、JavaWeb 基本应用，掌握 Spring、Spring MVC、Spring Boot 等框架的基本应用。了解 Redis 数据库的原理与实现。了解 MySQL 等关系型数据库的基本优化。了解 Linux 系统的相关操作以及服务器基础运维。

项目的创新点和特色

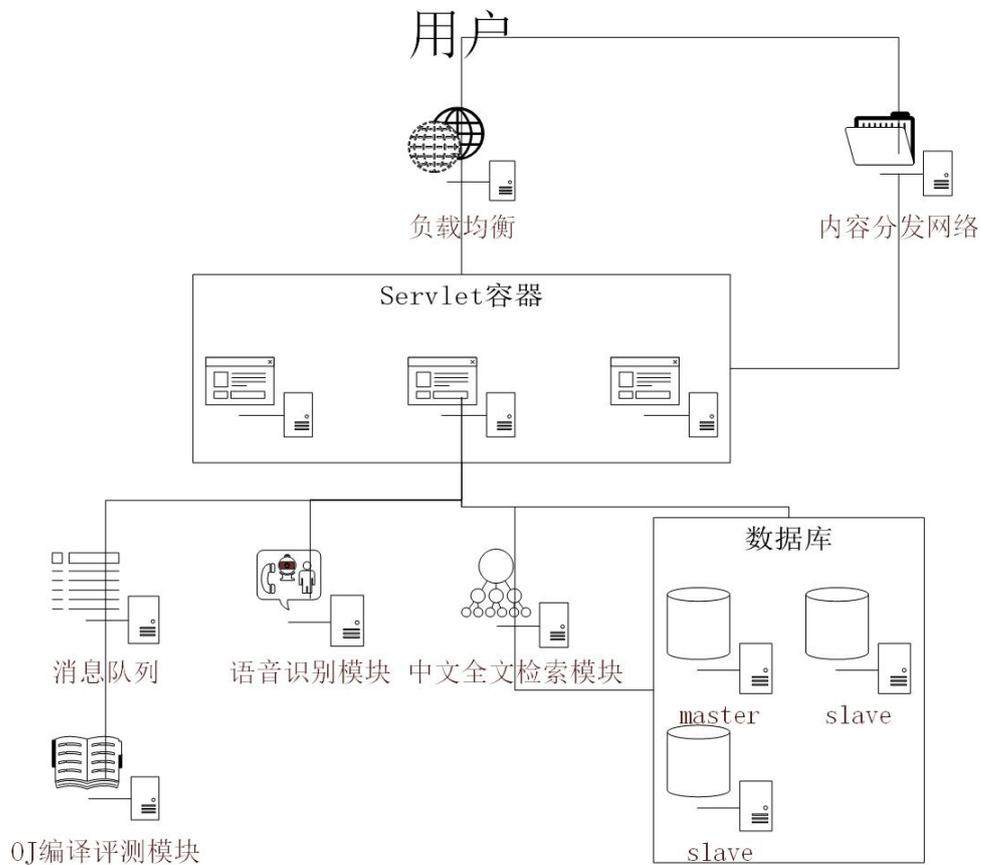
- 1、人工局部干预的智能学习引导方式，可以通过语音搜索相关内容，让初学者从不明白、不清晰、无法描述自己所面对的问题逐步过渡到能清晰的描述并解决问题。
- 2、OJ (Online Judge) 系统的在线编译以及智能的错误提示，用户在编程时可以获得智能的知识点提示，以及口语化的错误提示。
- 3、系统具有自我学习的能力，若用户提出的问题在问题库未找到，则系统将利用搜索引擎对互联网进行搜索，搜索内容将通过人工优化后扩充至知识库。用户使用过程产生的提问、解答、程序代码等都会被系统记录并用于扩充知识库，随着用户的使用，系统知识库将越来越丰富。

项目的技术路线及预期成果

技术路线：

整体项目采用微服务理念，根据业务纵向分割，独立成服务。服务之间互相独立，降低局部故障对整体系统的影响，提高系统的容灾性以及稳定性。每个服务采用 Spring、Spring MVC、Spring Boot、MyBatis 结构构建。

系统架构图如下：



负载均衡：

采用 Nginx 作为负载均衡模块，轻量高效，可以和后端的应用服务协调的契合。将用户请求分散给后端服务，降低单机负载。

内容分发网络：

将静态内容（例如用户头像）缓存至 CDN 服务器，提高用户访问速度，降低服务器负载。

消息队列：

采用 ActiveMQ 作为消息队列模块，将并行请求转化为串行请求，削峰填谷，保证 OJ 服务器负载平稳，运行稳定。

OJ 编译评测模块：

OJ 编译评测模块需要重点处理的问题就是安全问题，由于需要将用户输

入源码编译并运行，这个过程中存在较大的安全隐患。我们将对外部输入代码进行白名单验证，并利用 Linux 系统下成熟的沙箱技术 Apparmor 对运行期的评测程序做细粒度的权限控制。保证被评测程序的运行不影响服务器的安全稳定。

语音识别模块：

语音识别模块用于解析用户声音，转化为文字再对文档进行搜索。由于语音识别技术过于复杂，我们将调用现有的成熟技术方案实现对音频的解析。语音识别的基本原理如下：

1、发出的语音属于模拟信号，为了对语音信号进行分析和处理，需要进行模数转换

2、数字化后，提取语音信号的声学特征：

1) 人耳的生理特性：

生理 感知

强度 响度

基本频率 基音

频谱形状 音色

开合时间 时间

双耳相位差 位置

2) 根据人耳的生理特征进行声学特征提取：把每一帧波形变成一个多维向量，可以理解为这个向量包含了这帧语音的内容信息。

3、得出语音识别结果

1) 把帧识别成状态

通过“声学模型”里的参数，获取帧和某一状态对应的概率，若干帧数对应一个 状态。

2) 把状态组合成音素。

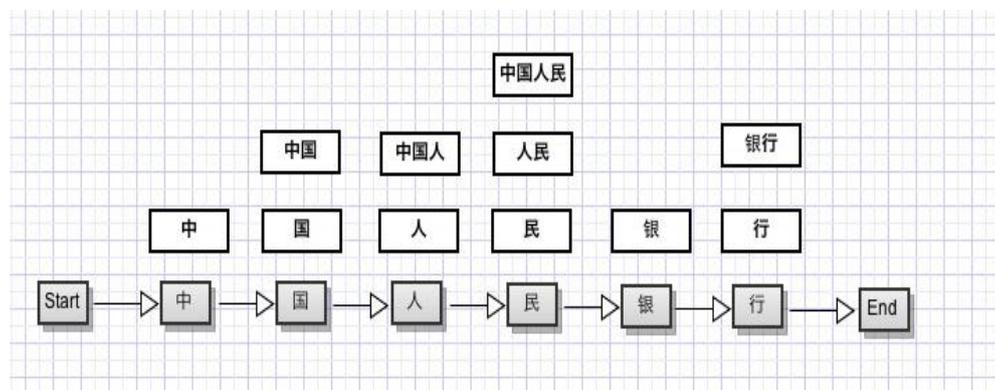
每三个状态组成一个音素。

3) 把音素组合成单词。

中文全文检索模块：

1) 中文分词

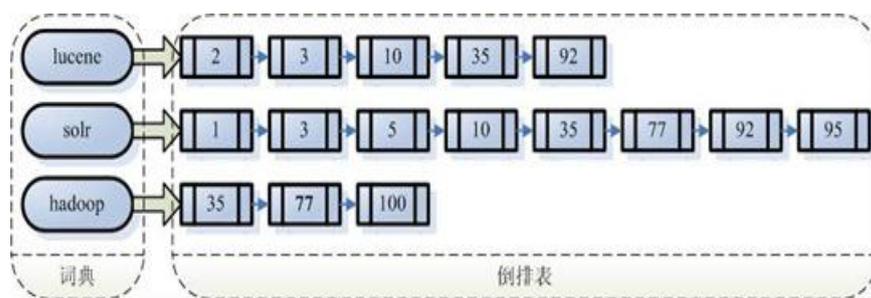
高效准确的中文分词算法是搜索引擎实现的第一步，本程序采用基于 CRF 算法的中文分词实现。



首先对每一个字计算出一个标签属性，某个字根据 CRF 模型提供的模板生成了一系列特征函数，这些函数的输出值乘以

该函数的权值最后求和得出了一个分数。该分数只是“点函数”的得分，还需加上“边函数”的得分。边函数在本分词模型中简化为 $f(s', s)$ ，其中 s' 为前一个字的标签， s 为当前字的标签。于是该边函数就可以用一个 4×4 的矩阵描述，相当于 HMM 中的转移概率。然后通过 BEMS 合并，将分词结果送入词性数据库中查询词性，去除停词 (Stop word) 后进行下一步反向索引。

2) 反向索引的建立



将分词后得到的词建立一个字典，按自然顺序排序，合并相同的词，成为文档倒排链表。便于搜索引擎根据关键字快速查询文章。

3) 索引索引

对用户输入的查询语句进行自然语言分析，识别单词关键字（此处和中文分词基本相同）形成自然语言语法树，根据关键词在反向索引表中查询文档。

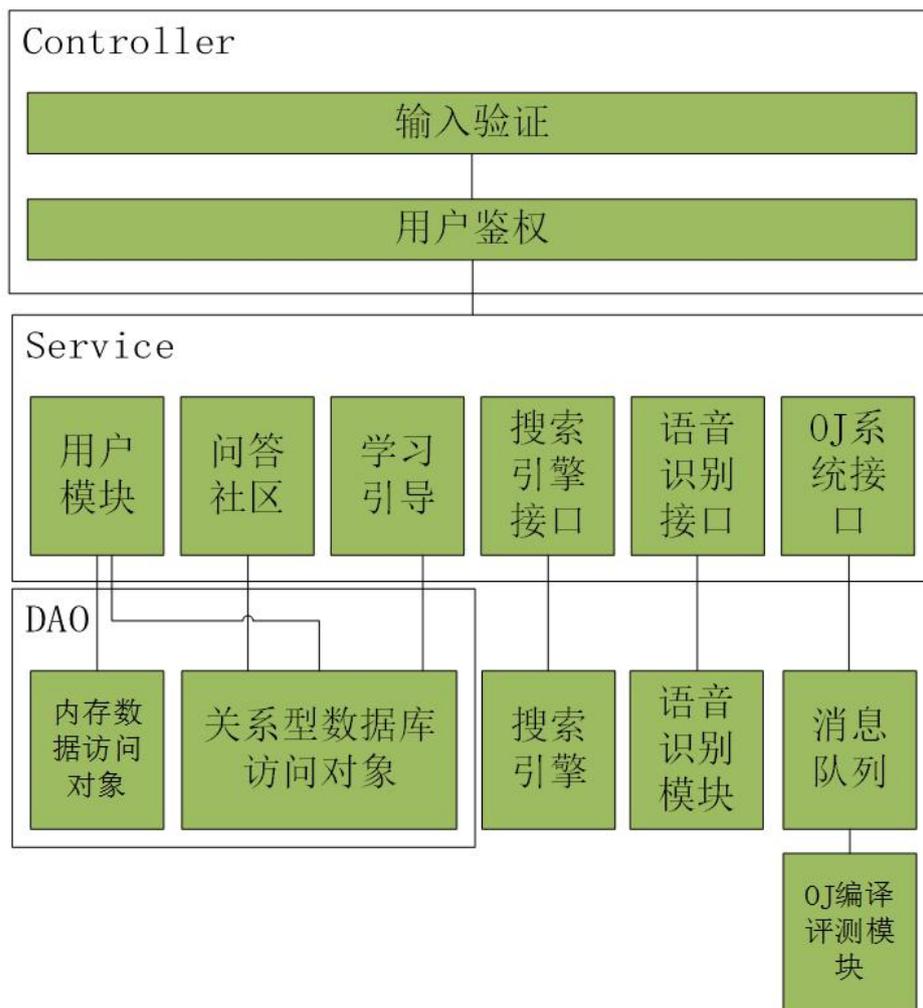
4) 整合结果文档

根据自然语言语法树对查询到的文档进行合并、差等操作。通过关键词词频等参数计算文档相关度，并排序显示给用户。

数据库服务器：

关系型数据库服务器将采用 MySQL Server，NoSQL 服务器将采用 Redis。我们将用户状态信息（例如 session）储存在 Redis 中，保证应用服务器无状态性，易于平滑扩展。

应用结构层次如下：



我们将应用进行横向拆分，形成 Controller，Service，DAO 结构。

Controller 层负责与用户交互收发信息，验证输入数据合法性，按照一定规则将请求分发给 Service。Service 层拆分处理请求。DAO 层负责与数据库交互。我们采用微服务理念，将复杂而逻辑依赖少的模块（例如搜索引擎、语音识别、OJ 测评）独立出来，以微服务的形式存在，服务与服务之间采用 RPC 协议通讯，每个服务专人负责，保证系统稳定性，降低单服务复杂度。

预期成果:

- 1、完成《基于 java web 完成智能化编程语言初学者导引平台系统》软件设计。
- 2、对整个项目过程进行总结，做一份项目报告。
- 3、申请软件著作权。

年度目标和工作内容（分年度写）

一：2017 年第一季度

- 1、对当前现有的编程语言学习平台进行深入的调查。分析各个平台优缺点。寻求合适的解决方案。
- 2、根据市场调研的结果对系统需求进行完善。
- 3、查阅相关材料，分析技术可行性，拟定技术实现方案。

二：2017 年第二季度

- 1、规划系统结构，按需求设计服务接口。
- 2、设计数据库。
- 3、前后端同步开发，搭建开发框架。

三：2017 年第三季度

- 1、解决语音识别、全文搜索等技术问题，实现其基本功能。
- 2、设计并实现一个安全可靠的 OJ 评测系统

四：2018 年第一季度

- 1、完成业务逻辑的基本开发，实现用户、问答社区、学习引导等模块。
- 2、完成前后端接口对接。
- 3、将语音识别、全文搜索、OJ 评测等服务接入系统。

五：2018 年第二季度

- 1、对系统进行集成调试，前后端联调，初步完善系统功能。
- 2、完善系统内容。

六：2018 年第三季度

- 1、对系统进行内部测试，邀请一部分用户体验系统，进一步完善系统功能。

注：每个年度的第四季度为项目进度调整期

指导教师意见

该项目紧跟语言学习平台研发的热点,适当的将人工智能引入到计算机编程语言学习平台的研发中,试图解决语言学习者在学习过程中长期存在的难题:知识点学习应用编程难、程序编译错误提示理解、程序纠错难、对需查询的问题描叙难,瞄准学习平台的难点问题,值得鼓励提倡。项目的研发成功会具有很好的应用价值。

同意申报

签字: 叶吉祥 日期: 2017年4月17日

注: 本表栏空不够可另附纸张