

和系统检索出的答案返回给用户;分析层主要是对用户键入的问题进行处理,包括对问句分析进行预处理,然后通过信息检索计算相似度值确定最佳答案;数据层主要负责对本系统的停用词库、同义词库、Python 知识问答对知识库等的存储。

“Python 程序设计”课程智能答疑系统的基本流程如图 1 所示,包括问句分析、信息检索以及生成答案三个部分^[4]。系统采用 Python 语言开发,利用不同自然语言处理技术分别对三大模块进行分析和处理。Python 语言 jieba 库是常用的分词工具包,其准确率已经达到 95% 以上,用于对问句分析阶段的分词处理和词性标注;过滤停用词是特征词提取的重要流程之一,系统采用哈尔滨工业大学停用词扩展表将句子中助词、语气词、标点、虚词等信息量较低的词汇进行过滤;句子是短语的承载者,句子的语义可由句子的特征词来表达,本系统采用基于统计与语义信息相结合的算法实现特征词的提取。信息检索的功能主要是用预处理后的关键词与知识库进行匹配检索,在候选集中利用改进的基于知网(HowNet)的语义相似度计算从知识库中匹配问题答案^[5],相似度值大于阈值则直接输出,否则转入等待答疑模块等待教师答疑。Mongodb 数据库是基于分布式文件存储的数据库,支持的数据结构松散,适合作为本系统的基于问答对的知识库存储。用户层的交互是基于 Django 的 B/S 模式,这种轻量级前端框架使得前后端交互更加轻便,自助式的后台管理的优点也让教师端管理界面的搭建更便捷,便于将主要精力投入到其他研究和开发中。

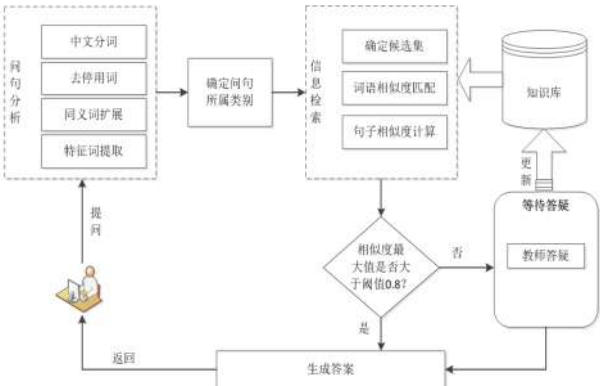


图 1 “Python 程序设计”课程智能答疑系统的基本流程图

2.2 系统主要功能

“Python 程序设计”智能答疑系统主要面向高校 Python 程序设计学习者,系统用户角色包括学生和教师,其中学生主要功能包括登录模块、自主答疑模块、我要答疑和我的问题模块,教师端包括用户管理模块、等待答疑模块以及知识库更新模块^[6]。具体如图 2 所示:

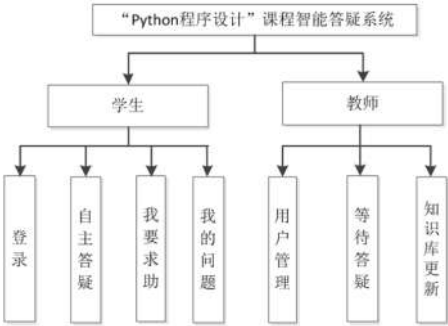


图 2 系统功能模块图

2.2.1 学生端

- 1)登录模块。根据账号和密码登录系统,可用于登记学生登录信息,准确记录学生检索日志。该账号和密码是教师端已为授课学生注册过的账户和密码。
- 2)自主答疑模块。本模块是答疑系统的核心模块,支持学生检索键入的 Python 程序设计相关知识自然语句,按照指定规则与知识库进行匹配,相似度值大于阈值 0.8,直接将准确、详细、简单的答案返回给用户。否则,系统自动提示问题返回失败,询问学生用户是否将问题提交到“我要求助”模块。
- 3)我要求助模块。学生在本模块中可将经自主答疑模块无法正确查询答案,可提交保存到此模块等待教师答疑。
- 4)我的问题模块。本模块供学生查看所提问的问题列表以及教师答疑问题的答案。

2.2.2 教师端

- 1)用户管理模块。开发人员指定唯一系统管理员账号,系统管理员可分配同权限功能教师端账号,但不可消除系统管理员,教师可为学生分配账号密码,也可以拒绝待申请账号和删除已有账号。
- 2)等待答疑模块。教师在等待答疑列表中可以查看学生在“我要求助”问题的列表,教师可以对问题列表中的问题进行回答。本模块有利于教师掌握学生对课程掌握薄弱之处,也利于补充知识库的知识点。
- 3)知识库更新模块。教师在本端可将优质问答定期更新到知识库,利于知识库的补充和完善,使之能有效提高学生对于知识检索匹配结果,让学生在答疑中效率更高,对教师的依赖性更少。

2.3 系统的知识库建设

知识库的全面是系统实现答疑功能的重要保障。选择“百度知道”“百度百科”“知乎”等为知识来源,利用 Scrapy 组件编写的网络爬虫进行数据抓取^[7-8]。共抓取到 Python 基础知识、程序设计、常见问题、Bug 调试信息等 3 万余对,同时结合本校开设的“Python 程序设计”课程的知识整理成问答对共 1600 余对一并存入知识库。教师答疑问题功能利于动态构建知识库,根据学生的我要答疑列表问题,可不断更新问答对知识,以保证知识库的充分与全面。具体的知识库组建模型如图 3 所示。

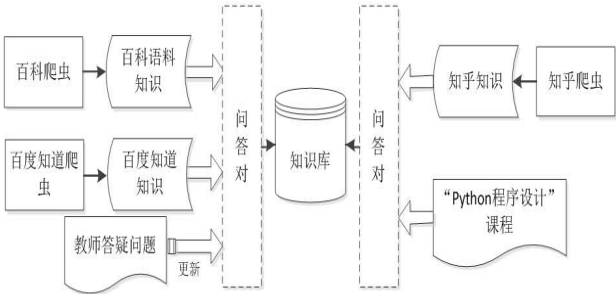


图 3 知识库组建基本模型图

3 实验测评

本系统知识内容主要涵盖了“Python 程序设计”课程概念类、编程常见问题类、Bug 调试类等多类问题,能有效满足学生对 Python 程序设计课程的自主答疑需求。系统答疑问题返回答案准确性是本系统智能性的重要体现,本实验测试用例随机

从互联网上抽取上述 3 类 Python 程序设计知识若干道,来测试答案返回准确率。由于相似度匹配值小于阈值导致无返回结果按照返回答案错误处理,当出现有争议答案时,采用少数服从多数的原则进行评判。评判结果如柱状图所示,蓝色柱状图代表该类实验用例问题总数,橙色柱状图代表该类实验用例正确答案返回总数,综合答疑答案返回正确率达到 85% 以上。

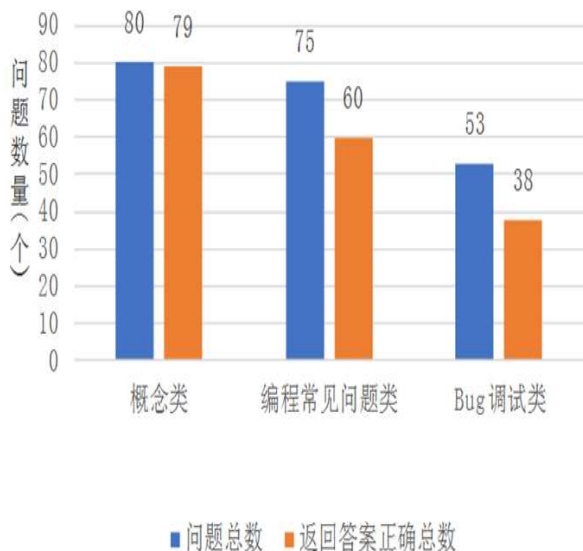


图4 实验用例测试结果图

经实验数据可以看出,概念类的实验结果较后两者好,但正确答案返回问题总数正确率与理想状态仍有差距。这是由于程序设计中编程类与Bug调试类问题灵活性高,知识库的存储不够全面,导致后者较前面准确率差。中文分词等自然语言处理技术对系统准确率也有一定影响。但系统功能目前还在不断完善,教师对学生“我要求助”的问题作答后可不断更新知识库,可以在一定程度上提高系统答疑准确率。同时本系统的研究应用不仅仅局限于“Python程序设计”课程的答疑,我校对

其他课程的知识库也在采集研究开发中。

4 结束语

“Python 程序设计”课程智能答疑系统是以互联网为基础设施的在教育领域的创新发展新成果,不仅能高效答疑 Python 程序设计知识问题,还能充分发挥学生的主观能动性去发现问题、解决问题,潜移默化地将答疑式学习方式融入学生思想当中,这也利于其他学科的学习。同时基于数字化、网络化的答疑平台突破了教学时空和设备的限制,也能避免教师重复劳动,有效减轻教师工作压力。此外,系统的可延展性丰富,在将来的工作中,会不断对系统进行更新和维护,让系统精度更高,涉及课程数目更广,努力让系统在高校中发挥更大的作用。

参考文献:

- [1] 宋海萍, 范丽娟. 论答疑式教学的教育价值及有效策略[J]. 大学教育, 2013(21): 138-140.
- [2] 倪宇虹, 韩秀娟. 试析开放教育中答疑讲解式教学模式[J]. 吉林广播电视大学学报, 2018(3): 5-6.
- [3] 王东升, 王卫民, 王石, 等. 面向限定领域问答系统的自然语言理解方法综述[J]. 计算机科学, 2017, 44(8): 1-8, 41.
- [4] 陈志云, 商月, 钱冬明. 基于知识图谱的智能答疑系统研究[J]. 计算机应用与软件, 2018, 35(2): 178-182.
- [5] 孙润志, 于放. 基于《知网》的词语相似度计算方法[J]. 计算机系统应用, 2015, 24(7): 155-158.
- [6] 张晓莹, 卢卫, 程一舰, 等. 面向慕课的在线SQL自动评测系统及应用[J]. 实验技术与管理, 2018, 35(4): 16-22.
- [7] 韩贝, 马明栋, 王得玉. 基于Scrapy框架的爬虫和反爬虫研究[J]. 计算机技术与发展, 2019(2): 1-5.
- [8] 李倩倩. 限定域问答系统自动建库及检索研究与系统设计实现[D]. 哈尔滨: 哈尔滨工业大学, 2017.

【通联编辑：谢媛媛】