

基于 Web 的特定领域智能答疑系统的设计与实现

石凤贵

(马鞍山师范高等专科学校 软件工程系 安徽 马鞍山 243041)

摘要: 在人工智能快速发展驱动下,对课程教学及课程教学网站建设提取了新要求。答疑是教学过程中必不可少的环节,那么如何高效、及时的解答学习者提出的问题?传统的答疑是面对面或在线答疑,教师的精力和时间有限,而且有的问题存在重复性和普遍性,学习者的问题往往得不到及时的解答,这将在一定程度上影响学习者的积极性。因此,需要构建一种智能的自动问答系统,及时、准确的向学习者返回问题的答案。对自动问答系统进行了阐述,并基于《编译原理》课程设计和实现了自动问答系统。

关键词: 自动问答; FAQ 库; 相似度计算; 分层架构

中图分类号: TP391.1 文献标志码: B 文章编号: 1008-9446(2021)01-0044-06

DOI:10.13377/j.cnki.jcpc.2021.01.010

Design and Implementation of Web-based Intelligent Question Answering System in Specific Domain

SHI Feng-gui

(Department of Software Engineering, Maanshan Teacher's College, Maanshan 243041, Anhui, China)

Abstract: Driven by the rapid development of artificial intelligence, new requirements have been raised for course teaching and construction of teaching website. Question answering is an essential part of teaching process. How to answer learners' questions efficiently and timely? Traditionally, teachers answer questions face to face or online, which has many disadvantages which are due to the limited energy and time of teachers. Repetitive and universal questions as well as delayed answers to the questions affect learners' enthusiasm to a certain extent. Therefore, it is necessary to build an intelligent automatic question answering system to send learners correct answers in time. Based on Compiling Principle course, this paper not only expounds but also designs and implements the automatic question answering system.

Key words: automatic question answering; FAQ database; similarity calculation; hierarchical architecture

随着互联网技术和多媒体技术的快速发展,传统的教学模式再也不能满足日常的教学。许多课程都构建了在线精品课程教学网站,提供了丰富的教学资源,部分教学网站还提供了如留言板、BBS 等互动版块。技术和社会的发展推动了人才培养方案的改革,教学方式和模式也随着发生了变化,出现了 MOOC(massive open online courses)教学。无论是在线精品课程网站还是 MOOC 均缺乏实时交互功能,问题不能得到及时回复。学生在课程学习中肯定会遇到疑问,如果这些疑问不能达到及时解答将直接影响学生学习的积极性。因此,需要寻求一种能实时智能解答问题的方式——自动问答(Question Answering)。

尽管搜索引擎相当于一本“百科全书”,但是搜索引擎返回的结果是网页结果集,包含了许多噪声,

基金项目: 安徽省教育厅科学研究项目: KJ2017A852; 安徽省高等学校质量工程项目: 2018ylzy160

收稿日期: 2020-07-01

第一作者简介: 石凤贵(1982-),男,安徽宿松人,讲师,硕士,研究方向: 自动问答、人工智能应用, E-mail: 226297089@qq.com。

从中获取有效答案需要进行人工二次筛选,用户体验较差。纵观科学技术发展,每一次变革都伴随着技术的革新,鼠标和键盘打开了 PC 时代的大门,手持终端打开了移动互联时代的大门,语音交互和模式识别带领人类进入人工智能(Artificial Intelligence)时代。答疑是教学中必不可少的一个环节,如何有效提高答疑时效?借助计算机技术和人工智能技术开发一个高效、准确、智能化的自动问答系统。本研究以《编译原理》课程为例,构建基于 Web 的课程智能答疑系统。

1 自动问答

人工智能已上升到国家战略,已经渗透到人们的日常生活、工作和学习。计算机辅助教学的方式还是不能满足个性化教学的需求。智能化教育以学习者为中心,以学习者需求为导向,培养人的思维和创造能力。应用人工智能技术,遵从教学规律,创新教学方式,优化人才培养方案,人工智能技术与教学深度融合,实现个性学习、实施交互、及时问答。从而改变传统教学与学习方式,实现人机结合;改变课程设计,线上线下同步实施,自动答疑解惑。

自动问答系统是利用自然语言处理技术,通过计算机理解用户提出的问题,然后检索答案并自动将答案返回给用户。自动问答系统可以有效减轻教师的教学负担,帮助学生学习,提供学生学习的积极性。自动问答系统是一种新型的信息检索系统,接受自然语言形式描述的问题,准确返回问题的答案。如“编译与编译过程之间有什么关系?”中文语句结构复杂,识别和处理中文语言是中文问答系统面临的困难与挑战^[1]。

1.1 自动问答系统的发展

20 世纪 60 年代,面向医疗应用的 ELIZA 系统^[2]问世;20 世纪 70 年代,面向 Unix 知识问答的系统 Unix Consultant^[3]问世;后来相继出现了麻省理工学院研发的 START^[4],美国华盛顿大学研发的问答系统 MULDER, Answer Bus^[5]。START 是第一个基于信息检索的开放域自动问答系统,Answer Bus 是支持多种语言的基于 Web 的自动问答系统。基于文中的问答服务系统典型的有中国移动推出的“10086 智能客服机器人”和“灵犀语音助手”^[6],京东商城的“JIMI 客服机器人”^[7],百度推出的“小度”,小米推出的“小爱同学”等。

1.2 基于 Web 的自动问答系统网络结构

基于 Web 的自动问答系统是借助互联网进行访问,用户访问不受时空限制,网络结构如图 1 所示。

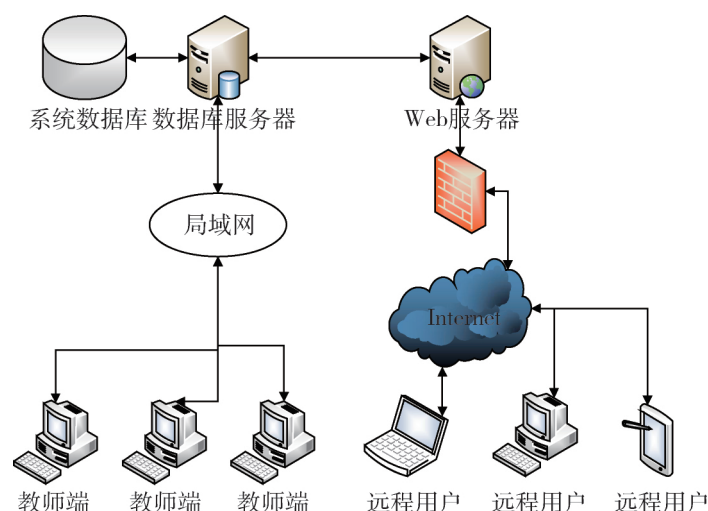


图1 基于Web的自动问答系统网络结构

系统允许学生以自然语言(汉语或英语等)问句的形式提出问题,并及时地将相关的答案返回给学生。如果没有找到相关的答案,则将该问题转发到课程教师的邮箱或提交到留言版和论坛,待解答后返

回。通过开发这样一个系统,目的是改革高校课程网站,特别是精品课程网站的在线答疑模式,充分调动学生学习的自主性,提高课程网站本身的资源利用率。

1.3 问答系统架构

问答系统包括问题理解、答案检索、答案返回三个基本过程^[8],本文系统架构如图 2 所示。

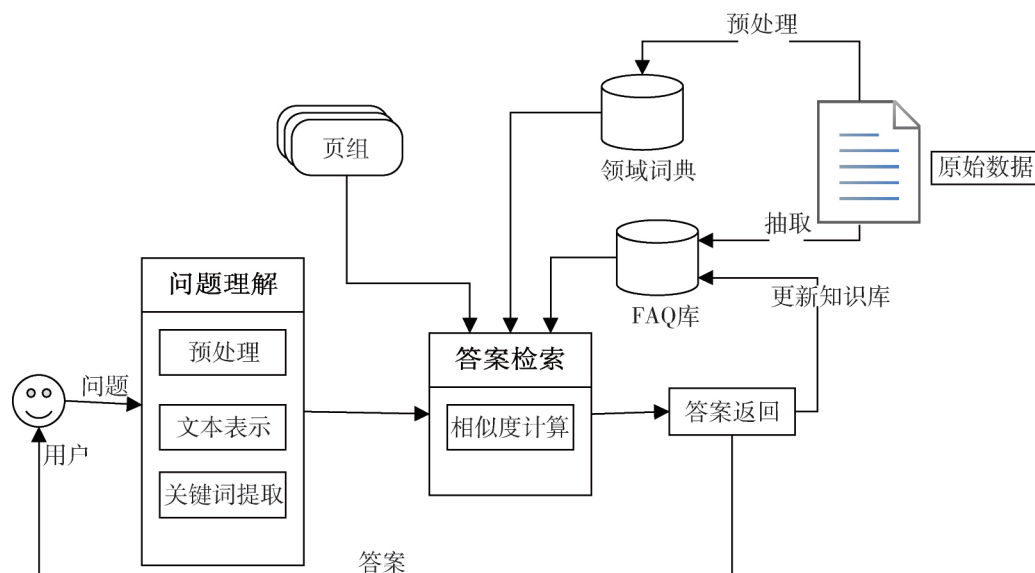


图2 问答系统架构

问句理解主要包括问题的分析理解,将文本转换为可计算的数字化形式,预处理是对问题进行分词和去停用词。答案检索优先在 FAQ(frequently asked questions)库中检索结构化“问题-答案”对,FAQ库中检索不到则从互联网检索。答案返回将问句相似度最高的问题答案直接返回给用户。对于 FAQ库未能解答的问题,根据互联网解答情况或人工解答后自动加入 FAQ库,丰富 FAQ库。

1.4 问句类型识别

自然语言中,问句都有归属的一个类型,对问句类型的判定可以增强对问题语义的理解。早期有人提出了问句分类的标准^[9],Wendy Lehnert 提出概念分类学^[10],包含了 13 个概念类:原因、目标、能力、结果、验证、定义、填充等。不过大多数实际的系统中都认为概念类帮助不大,因此它们大多根据答案类型建立更细的类型学。文中系统针对计算机学科教学领域,定义了 10 种问句类型,它们分别是:定义类、原因类、区别类、描述类、步骤类、点类、时间类、地点类、人物类、数量类,如表 1 所示。

1.5 相似度计算

检索答案是一个计算用户问题与 FAQ库中问题相似度的过程,由未知变成已知。因此,检索到 FAQ库中问题就可以直接返回答案。因此,检索答案检索的是 FAQ库中问题,检索的过程是一个相似度计算的过程。文中相似度计算采用基于问题关键词匹配度计算的相似度计算^[11]。

一个问题中含有一个或多个关键词,但是每个关键词在句子中的地位是不一样的,有的起决定性的作用。该类关键词就是要重点考虑的关键词,它与问题的匹配度为 1。

表 1 问句类型

类型	编号	疑问词举例
定义类/Definition	1	何谓、定义 + 什么、什么是
原因类/Reason	2	为什么、为何
区别类/Relation	3	不同 + 什么
描述类/Description	4	怎么、怎么样、如何
步骤类/Steps	5	步骤 + 什么
点类/Point	6	重点 + 什么
时间类/Time	7	什么时间、什么时候、何时
地点类/Location	8	什么地方、哪里、哪儿
人物类/Person	9	谁、什么人、哪位
数量类/Quantity	10	多少、多高、多大

例如: 编译原理是什么?

问题	关键词	匹配度	类型
编译原理是什么?	编译原理	1	概念
编译原理是什么?	编译	0.5	概念

起决定性作用的关键词,它与该问题的匹配度为 1,不起决定性的关键词与问题的匹配度为 0.5。

采用这种策略,可以实现简单的语义分析。例如:

“编译原理是什么?”与“什么是编译原理?”是同一个问题。如果用户输入的是这两个问题,则检索到的都是同一个问题。

2 问答系统设计

2.1 系统模块结构

本系统对用户问题的解答策略为:

策略 1: 利用 FAQ 库和问句匹配技术实现常见问题的快速回答;

策略 2: 对于策略 1 无法解答的问题,采 Email、在线答疑等方式来实现。

根据以上解答策略,设计了本智能答疑系统,模块结构如图 3 所示。

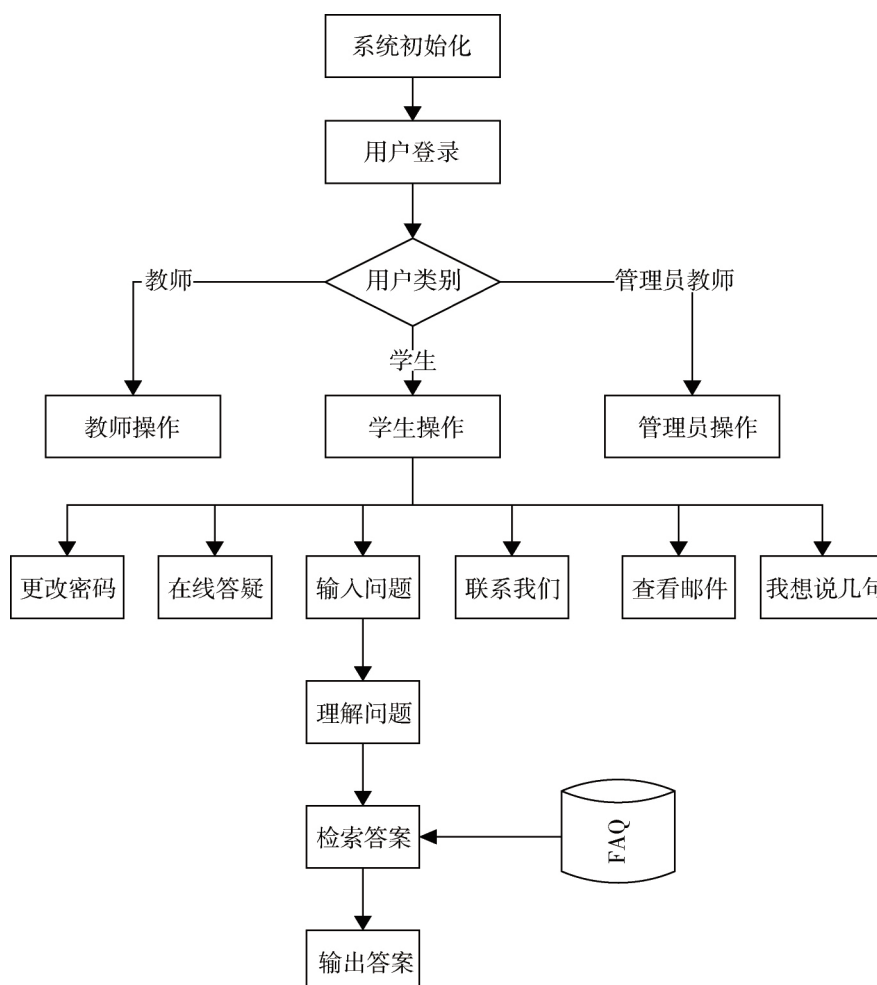


图3 系统模块结构

本系统各主要模块的功能如下:

1) 问句理解分析模块

本模块的功能是通过对用户问句进行分词和问句类型判断、问句关键词提取,从而完成对用户问题语义的理解,然后把语义分析结果表示成中间语言形式。

2) FAQ 库问句匹配模块

FAQ 库中保存的是用户常问的问题答案对,通过计算用户问的问题与 FAQ 库中的问题的句子语义相似度,查找与用户问题相似度较高的问题,若有,则把这些问题对应的答案从 FAQ 库中返回,若无,则同通过其他途径来实现。检索 FAQ 库可以实现快速答疑,省去了复杂的计算处理,提高了答疑效果。

文中重点讨论问句理解分析模块和 FAQ 库问句匹配模块。

2.2 问题匹配算法设计

问题匹配算法流程如图 4 所示。

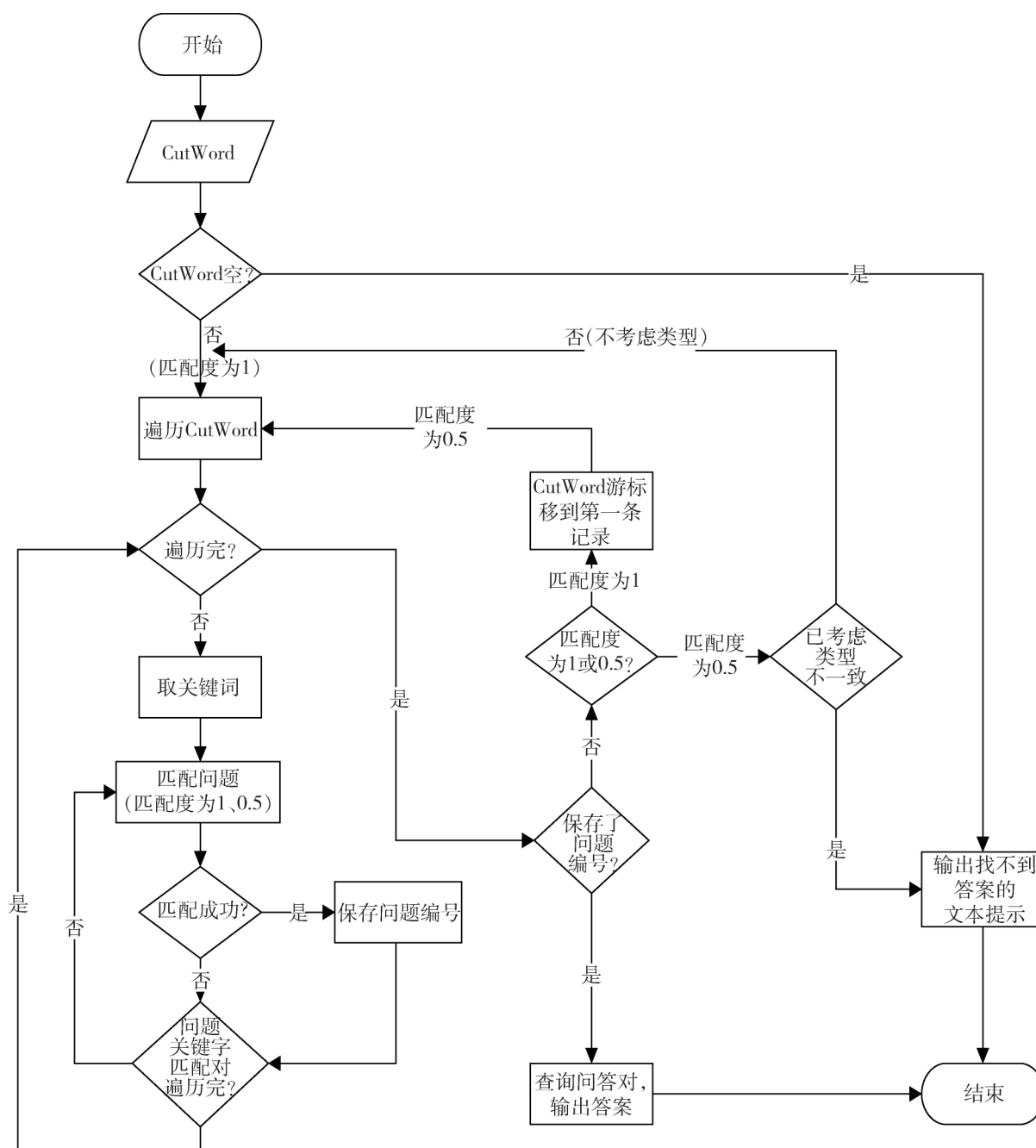


图4 问题匹配算法流程图

3 系统实现

3.1 系统运行模式

系统采用基于 B/S 模式的三层分层架构实现 如图 5 所示。

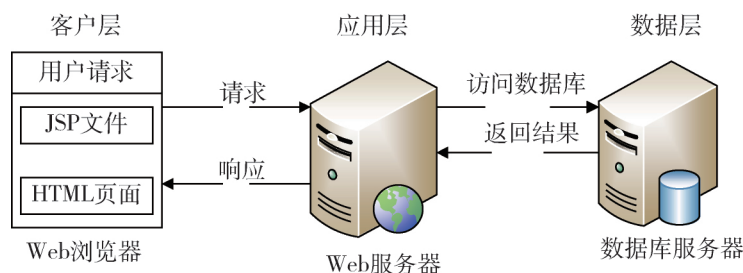


图5 三层分层体系结构

客户层即表示层,系统显示界面,用户与系统交互接口,具体由 Web 浏览器完成,采用 JSP 和 HTML 技术实现。应用层负责业务逻辑处理,包括处理用户层请求、建立数据库连接、数据库查询或更新、处理结果返回给用户层。本系统中问题分析与处理、答案检索由应用层完成。数据层负责系统对数据库的操作,提供对数据库中各类数据的访问和管理。三层体系结构具有分布灵活、逻辑处理集中和管理能力强的特点,具体表现在:高效率、易于维护、安全性增强、可伸缩性和移植性好^[12-13]。

3.2 系统运行效果

实验结果及界面如图 6、图 7 所示:



图6 提问界面

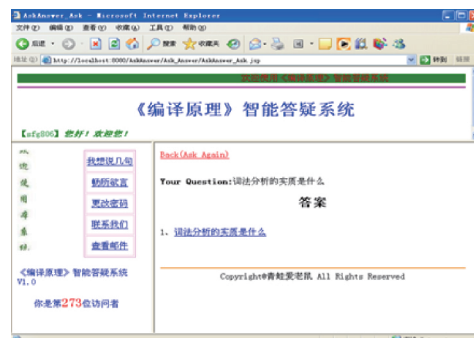


图7 返回答案

答案的具体内容用一个独立页面显示(*.html)。这样返回给用户的答案就可以不是简单文字描述,可以使用多种多媒体方式呈现,形象直观,用户容易理解、吸收,增强了学习者的学习兴趣,从而达到教学的最优化。

4 结语

随着互联网和人工智能技术的快速发展,对日常课程教学及教学课程网站构建提出了新问题,将人工智能应用于课程教学中。问答系统可以有效辅助教学,减轻教师的教学任务。文中对自动问答系统及其相关技术进行了介绍,给出了详细的设计过程。同时,基于《编译原理》课程设计并实现了《编译原理》课程自动问答系统。

参考文献:

- [1] 周蕾. 基于互联网的自动问答系统关键技术研究[D]. 西安: 西北大学, 2018.

(下转第 68 页)

健全课程评价的目的是为了更好的反馈教学,促进教学的改革,高职体育课程教学应根据学生未来职业发展的需求,打破原有传统化的教学模式,创新教学方法,构建多元化、科学性的教学模式,实现学生的主导地位。不仅可以丰富校园体育活动,弘扬体育文化、体育竞赛,同时对培养学生的终身体育意识具有重要的作用,能逐步实现学生职业素养的内化教育,提高学生的综合能力。

参考文献:

- [1] 聂强. 跨界与融合: 基于职业素养教育的高职课程建构研究[D]. 重庆: 西南大学, 2017.
- [2] 翁娟钗. 高职学生职业素养教育研究[D]. 福建: 福建师范大学, 2014.
- [3] 江欣怡. 高职学生职业素养的培养[D]. 广西: 广西师范大学, 2015.
- [4] 全国普通高等职业(专科)院校公共体育课程教学指导纲要(试行) [EB/OL]. <http://fanwen.jianlimoban.net/277492/>
- [5] 张文鹏. 中国学校体育政策的发展与改革研究[D]. 武汉: 华中师范大学, 2015.
- [6] 教育部. 普通高等学校健康教育指导纲要(2017-5). [EB/OL]. http://www.moe.gov.cn/srcsite/A17/moe_943/moe_946/201707/t20170710_308998.html
- [7] 谭远发, 邱成绪. 大学毕业生保留工资落差、工作转换和首职持续时间[J]. 高等教育研究, 2017, 38(7): 50-60.

(上接第 49 页)

- [2] Weizenbaum J. ELIZA-A Computer Program for the Study of Natural Language Communication Between Man and Machine [J]. Communications of the Acm, 1966, 9(1): 36-45.
- [3] Wilensky R, Chin D N, Luria M, et al. The Berkeley UNIX Consultant Project [M]//Intelligent Help Systems for UNIX. Berlin: Springer Netherlands, 2000: 49-94.
- [4] Collins-Thompson K, Callan J, Terra E, et al. The Effect of Document Retrieval Quality on Factoid Question Answering Performance [C]//International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2004: 574-575.
- [5] Zheng Z. Answer Bus Question Answering System [C]//International Conference on Human Language Technology Research. Morgan Kaufmann Publishers Inc. 2002: 399-404.
- [6] Tang X, Minton J R. Computer-implemented Method and System for Enabling the Automated Selection of Keywords for Rapid Keyword Portfolio Expansion: US, US 8036937 B2 [P]. 2011
- [7] Omer L, Rann S, Moshe T, et al. Undivide and Conquer: On Selling a Divisible and Homogeneous Good [J]. B. e. Journal of Theoretical Economics, 2015, 15(1): 1-23.
- [8] 余林峰. 高校教学问答服务机器人的应用研究[D]. 重庆: 重庆理工大学, 2018
- [9] 樊孝忠, 李宏乔, 李良富, 等. 银行领域汉语自动问答系统 BAQS 的研究与实现 [J]. 北京理工大学学报, 2004, 24(6): 528-532.
- [10] Eric J. Glover, Gary W. Flake, Steve Lawrence, William P. Birmingham, Andries Kruger, C. Lee Giles, David M. Pennock. Improving Category Specific Web Search by Learning Query Modifications, Symposium on Applications and the Internet, SAINT 2001, San Diego, California, January 8-12, IEEE Computer Society, Los Alamitos, CA, pp, 2001: 23-31.
- [11] 石凤贵. 智能问答系统中用户问题回答情况的研究 [J]. 电脑知识与技术, 2012, 8(30): 7323-7325, 7328.
- [12] 赖均. 软件工程 [M]. 北京: 清华大学出版社, 2016.
- [13] 耿祥义, 张跃平. JSP 实用教程(第 3 版) [M]. 北京: 清华大学出版社, 2015.