

**内蒙古师范大学计算机科学技术学院**

**毕业设计（论文）开题报告**

**题 目： 操作系统课程FAQ系统设计与开发**

**专 业 计算机科学与技术**

**学 生 白杨**

**学 号 20162106640**

**指导教师 松云**

**日 期 2019.11.19**

**一、开题报告主要内容**

1．课题来源及研究的目的和意义

随着互联网的迅速发展和广泛普及， 导致网上信息爆炸性增长。 如何能够在庞大的互联网上获得有价值的信息日益成为网民关注的问题。 随着搜索引擎技术的发展和普及， 人们能够通过搜索引擎方便的得到想要的各种信息。尽管搜索服务提供者在研发搜索技术方面已经花费了大量的时间和精力， 但是目前的搜索引擎仍然存在很多不足的地方。所以操作系统课程FAQ系统设计与开发的意义尤为重大， 这是因为：

1. 我国的专业网站不计其数， 操作系统课程FAQ系统所面向的恰好是专业网站， 它使学生能够迅速地在这有限的领域内， 查到他所关心的问题的答案。 因此，这类系统具有非常广阔的市场前景，它是各个专业网站吸引更多用户所必需的信息搜索技术。

2. 对于绝大多数 Internet 用户而言， 制约他们上网查询信息的瓶颈就在于网络速度的缓慢，这种迅速达到查询目标的技术自然会受到苦苦等待搜索结果的广大计算机用户的欢迎。

1. 国内外在该方向的研究现状及分析

问答系统涉及到句子相似度计算。 国外对于该层面的相似度研究主要集中在字符串的相似度计算方面。在国内， 对于汉语句子的相似度计算研究方面， 穗志方、 俞士汶提出了基于骨架依存树的语句相似度计算模型并用于基于实例的机器翻译；李伟等抽取文本中的关键句以及关键句中的关键词用于文本相似性计算；李素建基于知网和同义词词林，提出了语句相关度的定量计算模型；吕学强等考虑词形相似度和词序相似度两个因素， 提出了句子相似模型和最相似句子的查找算法；车万翔等利用改进编辑距离进行中文相似句子的检索；秦兵等采用 TFIDF 法和基于语义的方法， 面向常问问题集计算问句间的相似度；崔桓等在基于网络的问答系统中综合考虑关键词的顺序、关键词之间的距离、 以及问句和答案的长度等信息，用于计算其相似度。现有的这些方法主要是以词语为基本处理单元， 通过计算相同词语所占的比重确定句子之间的相似度。 部分在句法分析的基础上计算相似度的方法， 由于现有句法分析的精度限制， 难以取得较好的效果。

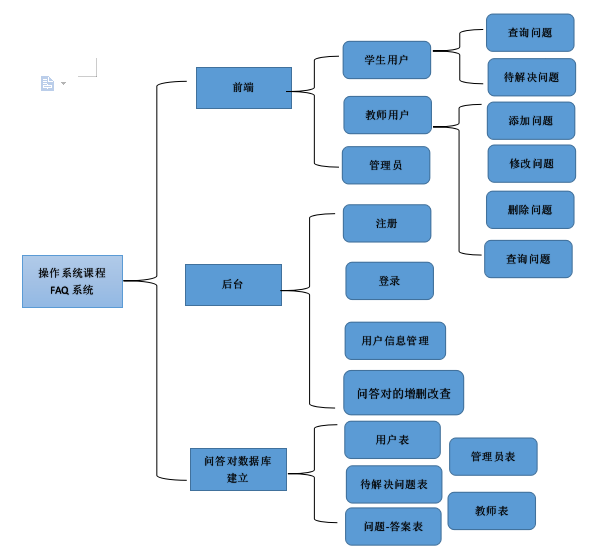
文本相似度计算是自然语言处理, Web 智能检索、 问题分类和文本分类研究中一

个比较基础的课题。 它所研究的是用什么样的方法来计算或比较两个文本的相似性，该方法也可应用于计算一个文本与某个类别的相似性和用于个性化推荐服务计算用户目 标和文本的匹配程度。在国内， 余刚提出了一种基于《知网》 的词汇语义相似度计算的文本相似度算法。 通过知网的词汇描述方式， 引入了词汇的语义计算， 即引入计算任何两个中文词汇的语义相似度方法;接着计算两个文本向量空间词条之间的相似度再采用基于二部图的最大权匹配算法获得两个文木的相似度值。

1. 主要研究内容

主要针对操作系统课程这门课程的基于FAQ的问答系统问题库的构建和维护展开研究。

1. 学生用户：查询、提交未解决问题功能
2. 教师管理员：管理用户、对问答对进行维护（增、删、改、查）功能
3. 数据表：问答-答案表、用户-管理员表、待解决问题表



1. 研究方案

首先将其分为三部分：前端、后台、问答对数据库建立。

1. 前端：用户（登录、注册、查询、提交未解决问题界面）、教师用户（增删改查界面）、管理员（用户信息、教师信息界面）、主界面（常见问题、搜索、网站导航）
2. 后台：实现用户登录注册、教师管理员对问答对的增删改查
3. 问答对数据库建立：问答对收集、建问题-答案表、管理员表、待解答问题表
4. 进度安排，预期达到的目标

第一阶段：2019年10月份——确定题目、开题报告

2019年11月份——需求分析：功能模块、数据表结构设计

第二阶段：2019年12月份—2020年1月份——教师用户及管理员界面设计

1. 实现教师用户登录注册功能
2. 实现教师用户增删改查功能
3. 实现管理员维护用户信息功能

第三阶段：2020年2、3月份——开发系统、维护测试、论文撰写

2020年4月份——答辩

1. 课题已具备和所需的条件、经费

课题所需软件已经安装完毕，机房也有分配。

1. 研究过程中可能遇到的困难和问题，解决的措施

词条收集太少——从网上查找

页面设计欠缺——参考相关软件页面设计

1. 主要参考文献

[1]  郑实福,刘挺,秦兵.中文自动问答系统综述[J].中文信息学报, 2002, 6(16): 46‐52.

[2]  刘涛． 受限领域问答系统 FAQ 的研究与实现[D]． 北京： 北京理工大学硕士学位论文， 2006．

[3]  陈康。 受限领域问答系统的文件分析研究[D]。  北京： 北京理工大学博士学位论文， 2008．

[4]  裴婧， 包宏。 汉语句子相似度计算在 FAQ 中的应用[C]。  计算机工程， 2009,17

[5]  朱毅华， 侯汉清， 沙印亭。 计算机识别同义词的两种算法比较与测评[J]。 中国图书馆学报，2002,28（140:82‐85）

[6]   朱毅华。  智能搜索引擎中的同义词识别算法研究[D]. 南京： 南京农业大学硕士学位论文，2001‐06

[7]  刘群， 李素建。 基于《知网》 的词汇语义相似度计算[C].第三届中文词汇语义学研讨会， 中国台北， 2002.

[8]Mong Li Lee，Hongjun Lu，Tok Wang Ling， Yee Teng Ko． Cleansing data for mining and warehousing[C]． Database and Expert Systems Applications， Florence Italy， 1999．

[9] Agirre E，  Rigau G． A proposal for word sense disambiguation using conceptual distance[C]． Recent

[10]Advances in Natural Language Processing RANLP'95， Tzigov Chark， Bulgaria， 1995．