

**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 马艳娇**

**学 号 20162106628**

**指导教师 松云**

**日 期 2019.11.19**

**计算机科学技术学院制**

**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。比较有名的搜索引擎有Baidu, Yahoo， Google 等。无论哪方面的内容，这些搜索引擎都能帮助人们快速地找到相关的网页。用户只需输入一些关键字，它们马上就会搜索出相关的网页。但是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。其中主要有三个方面:

1. 相关性信息太多。传统的搜索引擎返回的页面太多，用户很难快速准确地定位到所需的信息。例如，用户在Google上输入几个关键字，它有可能返回成千上万个页面，用户将浪费大量时间在这些网页中寻找自己所需要的信息。

2． 用户难以通过几个关键词的简单组合来表达检索需求。因为人们的检索需求往往是非常复杂而又各具特殊性，通过简单的几个关键词组合用户不可能将自己的检索意图表达清楚，搜索引擎自然也就没有办法找到令用户满意的答案了 。

3． 检索效果差强人意。以关键词为基础的索引、 匹配算法尽管简单易行， 但是对语义信息的理解不够充分，因此检索效果不是很好。

长久以来，对于解决此类问题的方法层出不穷，各个国家研究人员通过研究分析设计出了一种自动问答系统。对于问答系统，用户不需要把自己的问题分解成关键字，用户可以把整个问题直接交给问答系统。问答系统结合自然语言处理技术，通过对问题理解，能够直接提交给用户想要的答案。而我们要研究的是面向于专业网站的FAQ系统,它就像一个知识渊博的专家，可以快速准确地回答任何问题。在我国，研制开发专业网站的FAQ系统的意义尤为重大，这是因为：

1.我国的专业网站不计其数， 自然语言问答系统所面向的恰好是专业网站， 它使人们能够迅速地在专业网站这一有限的领域内， 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. 对于我国绝大多数 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。

3.由于采用友好自然的自然语言接口，可以消除计算机用户特别是不擅长英语的用户对Internet 的神秘感和与计算机进行交互的障碍，使计算机真正成为最广大的计算机用户的良师益友，对于Internet 技术在我国的进一步普及， 具有不可估量的重大意义。

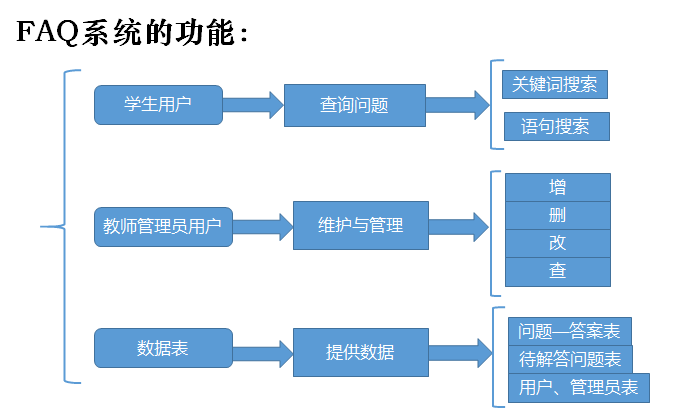
2．国内外在该方向的研究现状及分析

自动问答系统涉及到多种自然语言处理技术，包括词语、组块、句子、文本相似度计算等。这些技术的研究和发展对于自动问答系统问题库的构建和维护都起到了提高和改良的作用。

自动问答系统作为自然语言处理的一部分，近些年取得了飞速的发展。与传统检索系统相比，自动问答系统允许用户用自然语言而非关键词进行提问，并返回一个简洁、准确的答案，而不是一些网页。因此能够更好地符合用户的检索目的，具有准确、快捷和高效等特点，是目前自然语言处理领域的一个研究热点。

麻省理工大学已经开发出了问答系统 START，并且从1993年开始就发布到网上。通过该系统可以回答一些有关地理、文化、科技、娱乐等方面的简单问题。另一个比较成熟的系统是 Answer Bus，它不仅可以回答英语问题，还可以回答法语、 西班牙语、德语、意大利语和葡萄牙语的问题。国内在自动问答系统的研究上也取得了重大进展。陆汝玲院士主持开发的盘古知识库并以此为后台的对话系统—Auto talk System。Auto Talk System 是具有常识知识的对话系统。该系统的任务是模仿一个学龄前儿童与人交谈，输入输出都是自然语言。

3．主要研究内容

主要围绕基于操作系统课程FAQ系统的开发展开设计与研究。

FAQ系统的优点：

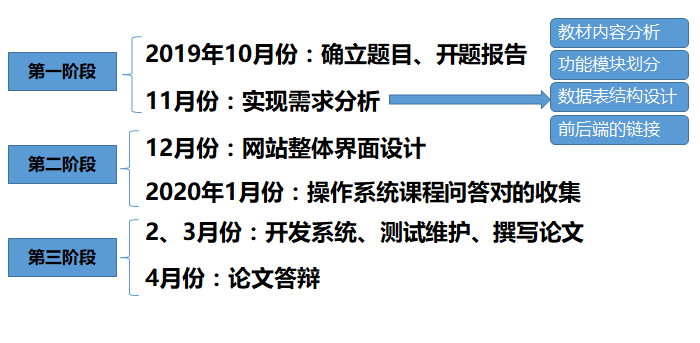
1、 针对某一特定领域而设立，所以能够对这一领域的问题做出相对准确地回答

2、 对有关问题和答案进行收录并分类，根据不同的问题，给出用户不同的答案，更加容易精确的获取到相关的知识，并且可以提高获取知识的效率

4. 研究方案

以小组讨论的形式展开研究，首先拟定论文题目，与指导老师沟通过后确定最终题目，组长与组内成员分工合作，设计系统界面，分析课程内容，然后根据教材的知识结构，划分主题，把学生经常遇到的问题划分成 3 类， 一类是概念型问题，第二类是说明型问题，第三类是方法型问题。再开始问答对的收集，建立数据库，定期与指导老师沟通，针对操作系统课程FAQ系统要做到增、删、改、查，四个方面，确定用户与管理员两大模块，最后对平台进行管理与维护

5．进度安排，预期达到的目标



6．课题已具备和所需的条件、经费

课题已具备和所需软件已全部安装完毕，不需要经费。

7．研究过程中可能遇到的困难和问题，解决的措施

1、词条收集过少——在搜索引擎上进行查找作为参考

2、页面设计过于简陋——学习界面设计相关软件进行美化

3、相关制作技术不成熟——向相关技术人员寻求帮助

8．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．