**电 子 科 技 大 学**

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

**学士学位论文**

**BACHELOR THESIS**



论文题目  **基于机器学习的问答推荐算法设计**

专 业 **计算机科学与技术**

学 号 **2014060108020**

作者姓名 **伍峰**

指导教师 **俸志刚**

摘 要

当今互联网发展如此迅速,日新月异的各种技术层出不穷,什么大数据,云计算,人工智能等正潜移默化的进入了人们的生活,这里面我们都能看到一个技术的身影,那就是机器学习,面对互联网信息爆炸的今天,信息量爆炸,用传统的信息处理方式已经不足以满足现在的数据量,也就有了机器学习的用武之地.本文以问题答案匹配为机器学习的对象,以爬虫爬取数据,索引构建,模型构建,机器学习等为主要内容.

**关键词：**机器学习,问题答案推荐,模型构建,文本处理,关键词提取,爬虫,搜索引擎,索引

ABSTRACT

with the development of the Internet so quickly, so many new technology are born just like an endless stream,like big data,cloud computing,artificial intelligence and so on.

……

**Keywords:** time-domain electromagnetic scattering, time-domain integral equation (TDIE), marching-on in-time (MOT) scheme, late·-time instability, plane wave time-domain (PWTD) algorithm

第一章 绪 论

1.1 研究工作背景和意义

1.2 机器学习国内外研究历史和现状

1.3 LTR研究现状

1.4 本论文结构安排

第二章 机器学习基础

2.1 机器学习分类

2.2 机器学习基本原理

2.3 深度神经网络

2.4 LTR分类

第三章 爬虫基础

3.1 爬虫基本原理

3.2 爬虫常见问题

第四章 搜索引擎基础

4.1 搜索引擎基本原理

4.2 文本特征值处理

4.3 索引

4.4 LTR

第五章 问答推荐系统研究与设计

5.1 数据的爬取

5.1.1 爬虫框架选取

5.1.2 数据结构设计

5.1.3 爬虫代码设计

5.1.4 爬虫中间件设计

5.2 数据预处理

5.2.1 文本特征值提取

5.2.2 建立正向索引

5.2.3 建立逆向索引

5.2.4 相关度标记

5.3 机器学习模型

5.3.1 模型的选取

5.3.2 模型的训练

5.4 优化

5.4.1 爬虫优化

5.4.2 数据库优化

第六章 全文总结和展望

6.1 总结

6.2 后续工作展望

致谢

参考文献