**西安邮电大学本科毕业设计(论文)** **选题审批表**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 申报人 | 杨武军 | | | 职称 | | 副教授 | | | 学院 | | 通信与信息工程学院 | | | |
| 题目名称 | 基于**Lucene的可定制搜索引擎设计与实现** | | | | | | | | | | | | | |
| 题目来源 | 科研 | |  | | | | | 教学 | | |  | 其它 | | √ |
| 题目类型 | 硬件设计 | |  | | 软件设计 | √ | | 论文 | | |  | 艺术作品 |  | |
| 题目性质 | 实际应用 | | | √ | | | | | 理论研究 | | |  | | |
| 题目  简述 | 现代互联网应用需要处理大型数据集来为用户提供信息服务，管理和搜索大规模数据集成为此类应用最具挑战性的任务之一，通常主流的门户网站和商务网站都会内置高性能的搜索服务。  本题研究互联网搜索引擎的组成和工作原理、重点研究Apache Lucene开源项目的主要功能特征，包括创建索引、搜索索引，结果排序的原理和方法。并基于Lucene设计一个新闻内容的搜索引擎原型系统，实现原始数据的获取、按需创建索引，按用户查询进行结果排序等功能，并允许根据用户偏好定制排序模型。 要求独立完成实验方案的设计和搭建，建立实验原型系统，完成测试分析和毕业论文。 | | | | | | | | | | | | | |
| 对学  生知  识与  能力  要求 | 1.有较好的TCP/IP网络基础知识。  2.了解Java编程语言，以及基本的数据分析和建模能力。  3.对开源软件有一定的了解，熟悉数据结构和算法的基本知识。  4.富有钻研精神和合作精神。 | | | | | | | | | | | | | |
| 预期  目标 | 1）独立搜集并学习Lucene开源库及Java编程技术，了解本课题涉及的相关内容对社会、安全、法律等方面的影响，制定工作计划、完成开题报告。  2）根据工作计划，综合考虑时间、经济、实验室设备限制等约束条件，提出可行的实验方案。  3）独立完成实验数据源的采集，实验方案设计、环境搭建以及索引和排序算法模型的建立。  4）基于Lucene开源库接口，设计搜索引擎原型系统并测试性能，评价其有效性。  5）了解开源社区及通信领域的知识产权保护及工程职业规范，能在毕设和论文撰写过程中自觉遵守、履行责任。按照撰写规范和质量要求，完成毕业论文撰写、验收及答辩。 | | | | | | | | | | | | | |
| 时间  进度 | 2020-3-16至2020-03-23： 研究和学习搜索算法及Lucene开源库，学习搜索引擎实现方法。  2020-03-24至2020-04-19：学习并搭建Java开发环境，完成Glassfish/Tomcat等运行环境搭建（也可使用其它服务器软件）。  2020-04-20至2020-05-19：完成数据源构建，设计搜索引擎的索引构建和结果排序模块，编写代码。基于Lucene编程库接口实现实验原型、并测试功能  2020-05-20至2020-05-31： 优化原型系统、编写文档撰写论文，完成论文初稿  2020-06-01至2020-06-11 ：完善并修改毕业论文，准备答辩。 | | | | | | | | | | | | | |
| 系（教研室）主任  签字 | | 年 月 日 | | | | | 主管院长  签字 | | | 年 月 日 | | | | |

田志晨 通工1614 05162002

Telenum：15691720916

QQ：943775910