**高校论文查重系统**

**可行性评估**

王显龙 15301051

林泽鑫 15301125

刘华秋 15301126

骆人杰 15301130

王瑞华 15301136

王婉铃 15301158

## 项目背景及简介

## 1.项目背景

近年来各大高等院校愈加重视学术打假。每年有大量的本科应届生、研究生提交自己的学术论文，其作为学术造假的一个主要源头应被严格把关。但现有的公共论文查重平台，如CNKI、万方等，虽然收录了大量期刊、会议论文，但对高校本科生、研究生所发表的论文收录并不全面，故出现了对高校内本科生、研究生论文搭建论文库进行查重的需求。高校论文查重系统提供了这样的平台，既为各高校提供定制化论文库的论文查重服务，更为各高校提供了共享私有论文数据库，形成高校论文联盟、提高查重效果的平台。

## 2.项目功能简介

高校论文查重系统主要分为专为高校内网使用的论文查重系统和架设在公网的论文查重系统两种形式。高校内网使用的论文查重系统主要功能如下：

1、论文库添加论文

2、查看论文库论文列表

3、删除论文库中的论文

4、单篇论文的查重

5、多篇论文的查重，其中既包括论文库查重，也包括多篇论文之间的查重

6、导出论文查重报告

面向公共网络的论文查重系统，在此之上还增加了注册、认证、高校之间共享论文库的功能。高校使用公共网络的论文查重系统时，仍然可以自建论文库，进行论文查重。高校进行高校认证后，可以被其他高校检索。认证高校可以向其他认证高校发起论文共享，各高校同意共享后，即可共享论文库。

## 项目技术可行性

## 前端设计技术可行性

前端主要使用BootStrap框架进行页面设计，主要技术包括HTML、CSS、JavaScript。前端设计主要包括论文库增删查页面、论文查重功能页面、查重结果展示页面。主要难点在于查重结果展示形式的设计，可以通过表格、图片等形式展示结果。在公网环境中的论文查重系统，还需要设计登录注册、高校审核和认证高校搜索页面，技术上均可通过BootStrap框架得以实现。

## 2.服务器端技术可行性

服务器端主要使用Spring + SpringMVC + MyBatis框架进行逻辑处理，包括登录注册、认证高校搜索、添加论文到论文库、删除论文、论文格式转换和论文查重功能。其中论文格式转换功能涉及到pdf、word、txt等格式，需要对这些相关格式的转换工具进行研究，设计转换逻辑，并对论文中不需要比对数据（如承诺书、目录等内容）的删除逻辑进行设计。已知可使用的格式转换工具有Acrobat接口、Office接口、PdfBox、pypdf2等。论文查重算法上，需要考虑查重算法的时间复杂度，同时需要考虑语句重复判定的逻辑，需要对现有查重逻辑、自然语言处理等相关内容进行研究。在自然语言处理上，有word2vec模型，可以将文字转换为向量，进行相关计算。

服务器端功能开发有一定难度，但在技术方面都有相应的解决方案，需要花费一定时间进行学习和设计。

## 项目市场可行性

本项目在开题前，已经询问部分学院领导，确认了高校有内部论文查重的需求，本项目为高校提供自建论文库、自由进行论文查重的功能，能在一定程度上解决高校论文内部查重的痛点，有一定的市场价值。

## 项目时间可行性

项目中主要难点在于论文格式转换、查重算法设计和查重结果展示的设计，这也是项目的核心功能。项目的时间主要花费在前端界面设计、查重算法设计、格式转换工具的学习和使用上，可以通过合理的分工和合理的时间规划，在课时内完成项目的开发。