**大学生创新创业训练计划项目**

**结题报告书**

**项 目 名 称：基于文本挖掘的法律裁判文书的数据处理平台**

**项 目 编 号：\_\_\_Q2018107490042\_**

**项目负责人：\_\_\_\_\_苏有福\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**项 目 级 别：\_\_\_\_ 区级\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**所 在 学 院： 信息工程学院\_**

**起 止 年 月：\_\_\_\_2018.4-2018.12\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**指 导 教 师：\_\_\_\_\_张海燕、杜方\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**其 他 成 员:\_\_\_\_\_曹虎、马飞龙\_**

**填 表 日 期：\_\_\_\_\_2018.12.8\_\_\_\_\_\_\_**

**宁 夏 大 学**

1. **基本情况**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 项目名称 | | 基于文本挖掘的法律裁判文书的数据处理平台 | | | 项目类型 | 创新训练 **√**  创业训练 □  创业实践 □ |
| 成果形式 | | 软件系统、软件使用文档 | | 立项时间 | 2018年 4 月 12 日 | |
| 完成时间 | | 2018年 12 月 15 日 | | 验收时间 | 2018年 12 月 29 日 | |
| 项  目  主  要  研  究  人  员 | 序号 | 姓名 | 学号 | 专业班级 | 所在学院 | 项目分工 |
| 1 | 苏有福 | 12016242316 | 计算机科学与技术二班 | 信息工程学院 | 非结构文本处理和分析 |
| 2 | 曹虎 | 12016242326 | 计算机科学与技术二班 | 信息工程学院 | 平台的设计和搭建 |
| 3 | 马飞龙 | |  | | --- | | 12016242327 | | 计算机科学与技术二班 | 信息工程学院 | 法律裁判文书的数据处理 |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |

1. **项目执行情况简介**

|  |
| --- |
| 本项目针对法律裁判文书设计实现文书的数据分析和处理平台，由于法律裁判文书中是关于涉案人员、案情描述、法院判决结果等信息以非结构化的文本形式表示，而文书内容中包含了很多有价值的信息，这些信息对于未判决案件有着参照价值，对于法律专业人士分析不同类型案件的特点具有分析价值。因此，本项目采用文本挖掘的技术对非结构化的法律裁判文书进行文本数据的分割和分析处理，提取出法律裁判文书中对于未来司法实践有价值的关键性信息，比如：法院、地区、案件类型、涉案人员、案情描述的关键词、依据法条、判决结果等信息，这些具体的关键信息的提取必然会成为依据法条推荐、司法裁判预测、司法要素辅助等实践应用的基础性工作。  本项目针对某律师事务所2年受理的法律判决文书数据（1G数据量）分析并设计与实现法律裁判文书分析与处理平台，该平台软件包括四个主干模块：（1）设计并实现“法律裁判文书处理”模块。该模块完成将非结构化文书数据的分割和分析处理成结构化的知识点过程，并将知识点数据存入Mysql数据库以备司法实践应用，以及提供了专业法律人士导入新裁判文书的功能。（2）设计并实现“法律裁判文书要素检索与展示”模块，该模块完成了清楚地展示法律裁判文书中所包含的关键知识点信息的检索与展示。（3）设计并实现“相似案件的文书推荐”模块，该模块完成了用户输入未判决案件的案件类型、地区、基本案情描述信息后，系统会从数据库检索出与该案件匹配度高的已判决的案件判决文书供用户参考。（4）设计并实现“法律裁判文书统计分析”模块，该模块完成了根据用户的不同需求统计分析不同地区、不同法院、不同案件类型、不同涉案人员的案件发生情况的统计直方图展示，以及统计不同案件类型下的案情描述中的关键词（频率出现高）的词云展示。  本项目不仅完成了申请书中的非结构数据的分析处理内容，还进一步将项目推进到法律知识点的展示和分析应用实践。项目所设计与实现的平台软件是智能司法研究工作的有意义的探索和实践，也是法律裁判文书的司法实践成果的有益延伸，同时，是计算机学科与法律学科的交叉实践应用探索了实际有效的合作模式。  在项目的实施阶段，每个人分工明确，主要由苏有福负责非结构文本处理和分析，各组员进行辅助配合任务，曹虎主要负责平台的设计和搭建工作，马飞龙主要负责法律裁判文书的数据处理工作，在整个过程中，各项任务并不是每个人简单的完成，都是相互学习，以及指导老师的指导和帮助。在项目需求确立阶段，各组员积极地提出自己的观点及意见，参与到系统的开发调研工作中，最终确立了需求报告，为系统开发打下了稳定的基础。在系统开发过程中各组员积极参与，对分配的工作及时有效的完成，保证了项目整体进度按时完成，在整个项目组员的不断努力下，完成了需求分析的研讨，数据库的设计，前端界面布局研讨，前端界面的搭建，后端的实现，前后端数据库的传输及显示，系统的功能测试，系统的整体测试等。  本项目按照项目计划有序推进，按照经费预算进行费用支出，主要用于平台设计与实现过程中的调研、资料印刷、工具学习、开发设备的更新和升级等费用开支。本项目共有三位成员，各自任务分工明确，每位成员都能按照项目计划努力完成，同时，遇到问题互相帮助和商量，协同完成了项目的各项任务。本项目按时完成大创项目的要求，并按时保证质量完成所有材料的提交。 |

1. **研究总结报告**

|  |
| --- |
| 本项目不仅完成了申请书中的非结构数据的分析处理的任务，还进一步将项目推进到法律知识点的展示和分析应用实践，以此更清楚和灵活地展示和应用法律裁判文书中的重要信息。  主要成绩：  本项目针对某律师事务所2年受理的法律判决文书数据（1G数据量）分析并设计与实现法律裁判文书分析与处理平台，该平台软件包括四个主干模块：（1）设计并实现“法律裁判文书处理”模块。该模块完成将非结构化文书数据的分割和分析处理成结构化的知识点过程，并将知识点数据存入Mysql数据库以备司法实践应用，以及提供了专业法律人士导入新裁判文书的功能。（2）设计并实现“法律裁判文书要素检索与展示”模块，该模块完成了清楚地展示法律裁判文书中所包含的关键知识点信息的检索与展示。（3）设计并实现“相似案件的文书推荐”模块，该模块完成了用户输入未判决案件的案件类型、地区、基本案情描述信息后，系统会从数据库检索出与该案件匹配度高的已判决的案件判决文书供用户参考。（4）设计并实现“法律裁判文书统计分析”模块，该模块完成了根据用户的不同需求统计分析不同地区、不同法院、不同案件类型、不同涉案人员的案件发生情况的统计直方图展示，以及统计不同案件类型下的案情描述中的关键词（频率出现高）的词云展示。  不足：  （1）由于项目组采用最新的Python数据处理语言作为平台的开发工具，此开发工具对于项目成员是全新的设计语言，由于熟练度不是很强，所以在平台开发的一些功能模块上界面和后台处理得有些部分不太完善和精致。  未来的工作：  （1）设计并实现从国家文书网爬取更多的文书数据处理，不断更新和扩充法律文书知识库中的信息。  （2）平台软件中的三个应用示范模块的功能进一步地完善。  （3）随着法律文书知识库中的信息不断扩张，如何高效地进行知识展示和应用，这是未来要考虑的重要问题。 |

**四、经费使用情况**

|  |
| --- |
| 经费合计1141.9元，其中，学校配套资助 元，学院（所）配套资助 元，其他经费 元。 |
| 经费支出情况：  项目硬件环境搭建：购买硬盘，优盘，键盘，内存条等共计676.9元  打印费：200元  车票及耗材：151元  服务器：114元 |

1. **学院评审意见**

|  |
| --- |
| 内容提示：院系专家组对结题的意见，包括对项目研究工作和研究成果的评价等。  专家组组长（签章）：  年 月 日 |

1. **学校创新创业训练计划领导小组审核意见**

|  |
| --- |
| 负责人（签章）：  年 月 日 |