**法律裁判文书说明文档**

目录

[**1.** **平台建设的意义** 1](#_Toc532475700)

[**2.** **平台建设的基本内容和目的** 2](#_Toc532475701)

[**3.** **后台法律裁判文书处理** 4](#_Toc532475702)

[**4.** **法律裁判文书要素的检索** 7](#_Toc532475703)

[**5.** **相似案件文书推荐** 8](#_Toc532475704)

[**6.** **法律裁判文书统计分析** 9](#_Toc532475705)

1. **平台建设的意义**

党的十八大以来，党中央高度重视司法的公开工作，2013年11月最高人民法院正式运行中国裁判文书网，推动裁判文书全面在互联网公开。这意味着裁判文书以标准化电子文本形式向世人公开，裁判文书是标准的官方记录法治要素的重要文件，它的公开对于法律大数据的研究和应用无疑提供了一个重要的数据宝库。

第二个方面，随着国家法治中国的不断推进，针对以往案多人少、司法不公开等法律现状，引入人工智能和大数据分析技术来辅助司法工作，这是推动和保证法治中国的重要举措。裁判文书是法律大数据的基础数据来源，因此，分析和处理基础数据是法律各种应用研究和实践的重要且根本的工作。

当前，裁判文书中的信息都是以半结构化和非结构化的文本形式存在，孤立的看，这些信息毫无意义，但是，如果能找到这些信息之间的联系，这些信息必将产生巨大的价值。当前，因为裁判文书的工作才刚刚开始公开化，针对这方面的研究工作也属于起步阶段。

目前，本项目已获得了超过1G的裁判文书的数据量，这对项目的研究和实施工作有着重要的意义。

1. **平台建设的基本内容和目的**

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。

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. **后台法律裁判文书处理**

文书的处理所依赖的一个基础条件是，法律文书是一类有着较为严格的书写规范的，各部分有着较为明确界限的文档类文件。根据这个条件，我们对文书进行的初步的分割，就转换为了“边界标签”类文字或语句的识别与检索，也即初等的自然语言识别。通过我们的努力，文书可以被初步的分割为标题、文书编号、地区和法院等独立且包含不同信息的部分。在接下来的工作中，我们运用所掌握的数据库知识，确定了以关系数据库存储的方案，为了方便与这些工作对接，在老师的指导下，我们将初步处理的文书以json文书格式，用一个个key-value对将文书存储，之后再将一个个json文件导入数据库。实现了较为准确，合理的文书处理。下面是后台的主要界面：

登录界面（需要权限）



后台管理系统界面



文书导入界面



案件类型界面  


1. **法律裁判文书要素的检索**

根据用户在搜索框中数输入的关键字，从大量裁判文书中进行关键字匹配，得出符合用户需求的裁判文书，展示每一个案件的标题、案件类型、案件年份、相关法院、涉案人员，用户可根据进一步点击标题来访问相关裁判文书的详细内容。并且我们已经提取出了一系列法院，地域，和案件类型供用户去选择，用户也可选择自己所需的要素进行裁判文书的查找。下面是主要的界面：



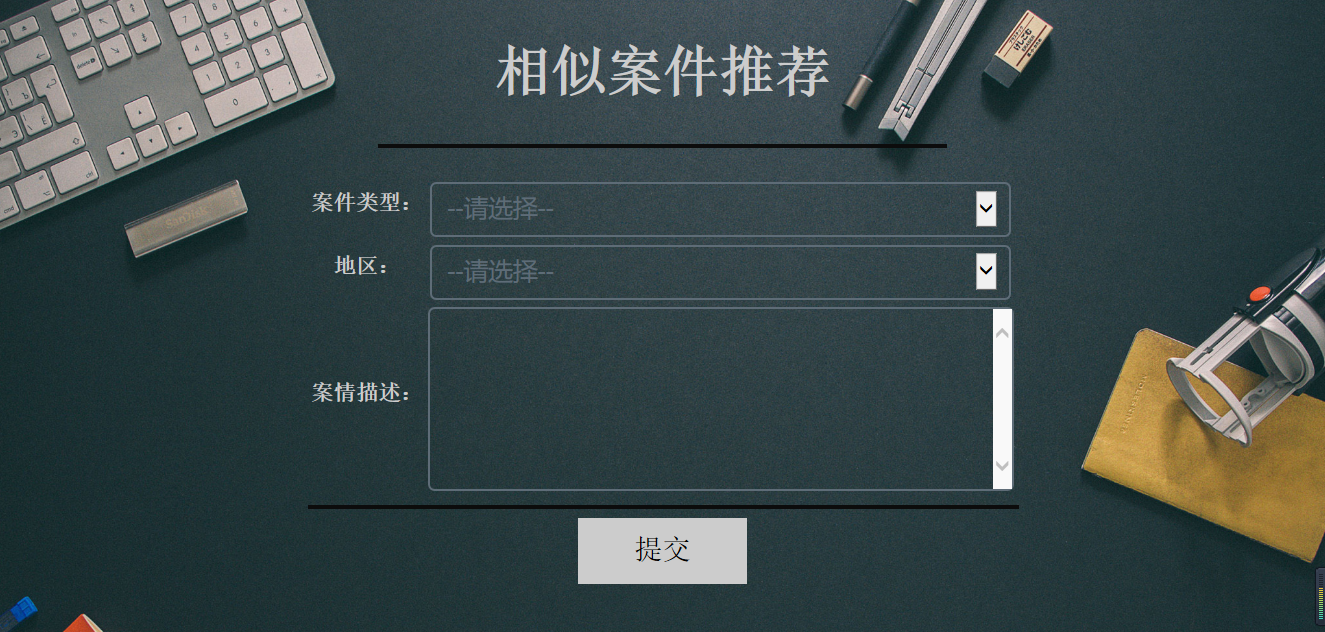
主界面



检索界面

1. **相似案件文书推荐**

用户可选择案件类型和地区，输入案情描述来得到与所输入的内容相似并且根据相似度进行排序过的裁判文书，展示每一个案件的标题、案件类型、案件年份、相关法院、涉案人员，用户可根据进一步点击标题来访问相关裁判文书的详细内容，便于用户进行比较和总结。下面是主要的界面：



相似案件推荐界面 

显示结果

1. **法律裁判文书统计分析**

法律裁判文书的统计主要是显示出关键字的统计，案件类型的统计，法院进行统计和地区进行统计。这些统计界面主要是为了让用户更好的分析案件，以及提供有效的帮助和建议，大大节省法律人士的时间，提高的效率。在海量的裁判文书中找到需求的案件，然后进行分析和统计，不管在时间上，还是效率，都是一个很大的提升，让法律人士按照自己的需求可以进行查询，这体现出来了统计的重要性。下面是统计的界面：

