### 浙大控股集团档案管理系统详细参数

### 1.总体要求

* 系统需采用B/S架构设计，使用JAVA语言开发，以便支持跨平台、跨数据库部署，节省后续维护成本。系统支持分布式部署，支持利用docker进行部署，支持maven管理与构建。
* 技术上基于REST风格的前后端分离的Web微服务架构，数据库采用Mysql，前端使用基于React或Vue的框架，后端使用SpringBoot的框架。
* 支持主流浏览器，包括IE9.0以上、FireFox、Chrome、Safari，以及多种基于IOS、 Android的手机端浏览器。
* 至少支持1000的并发访问和检索，支持亿级文档的存储和管理，检索响应速度不超过2秒。
* 符合“信息技术应用创新产业”国家战略要求，适配或兼容国产化操作系统（Linux内核）。

### 3.后台功能要求

* 平台具有库群管理功能，可根据档案文献的不同建设不同的数据库，每个数据库又可设置不同的子数据库且子数据库层级不限，每个叶子节点子库都可分别设置不同的数据模型。
* 平台提供遵循Dublin Core标准和规范的常用数据模型，包含却不限于档案、音视频、图片、论文等。
* 平台默认提供常用数据字段，同时提供字段的自由配置，字段符合多媒体规则,可满足视频、音频、图片、文本、富文本等数据的存储、预览、修改，可根据需要自由配置。支持对字段名称、字段描述、字段类型、分词方式等属性进行设置，同时支持字段类型的设置如：string、long、double、boolean、date、ip、object、nested、geo\_point、geo\_shape、completion等。
* 具有模型管理功能，模型通常由多个字段组成，平台默认提供常用模型，同时提供模型的字段配置、字段排序、字段匿名化、检索参数配置等定制化内容。
* 至少支持单字分词、索引分词、词典优先、精准分词、细粒度分词、自然语言分词、空白分词、英文分词、中文等分词方式。
* 具有索引管理功能，用户可以对元数据库索引进行管理，包括删除索引、重建索引、停止重建以及查看存储速度、索引速度等。
* 平台支持索引的分片管理，支持灵活的分布式搜索，平台根据需要索引的文档数量大小来设定分片数，以保证每个分片有最适合当前硬件配置的索引数据量，从而实现最佳的搜索效率。当集群扩容或缩小，系统会自动在节点间迁移分片，以使集群保持平衡。
* 支持对各个元数据表内的数据分片查看，支持对单条数据的修改、删除及录入功能，支持批量元数据的导入。
* 平台采用ES集群提供多种检索方式，支持集群的自由扩展，支持古籍书目简繁的自动转换及识别。支持百万数据毫秒级响应。
* 用户在检索前，可以选择文献类型进行检索。每种文献类型还可以选择相应的数据字段，以进一步缩小检索范围。
* 支持复杂的组合检索条件来提高检索精度，提供与或非的组合查询模式，各查询块可通过选择检索字段以及字段匹配度进一步缩小范围，提供精准以及模糊两种匹配度选择。
* 系统提供二次检索功能，用户可以在上一次的检索结果再次输入关键词进行检索。
* 用户检索时系统可根据用户的输入，提供检索词提示，检索词提示内容来源于平台对于用户的热词统计，也可支持用户指定提示词库。
* 平台针对pdf的全文内容以及目录信息提供数据检索，支持对检索结果进行配置，提供多种模式的全文检索策略如句读、固长等模式，并可通过检索结果直接跳转正文内容。
* 平台可根据标准格式文本文件自动抽取目录结构，编译标准json，提供目录信息的修改和手动添加，用户可在前台通过输入目录内容对数据进行检索。自动解析双层pdf、txt的文本内容，根据抽取结果构建全文检索数据索引。
* 平台内置通用词表，支持用户自定义词表的导入和维护。用户自定义词表包括行业词表、停用词表、过滤词表、同义词表、歧义词表以及规则词表。系统支持对单个词表进行查看和编辑，支持通过词表文件进行词表的创建与批量更新。
* 分词器支持加载多种类型的词表，包括用户词表、停用词表、同义词表和歧义词表。
* 提供多层级的数据管理功能，支持对数据的增删改、控制数据的状态、数据的权限、数据的版权，以及单条数据的多媒体附件绑定等功能。
* 对系统中存在的多媒体附件进行统一管理，提供系统内部的多媒体预览，能够支持jpg、pdf、mp4、mp3、jpeg、jp2、txt等多种格式的多媒体。
* 平台针对数据以及附件的删除提供回收站功能，用户需要在回收站中清除数据才能彻底删除内容，保证了数据存储的安全性，减少误操作带来的安全隐患。
* 平台提供可配置的水印信息，支持对水印的不透明度、密度、正文内容等信息进行配置，在线阅读书目可根据自定义的水印规则进行水印的调度及展示，能够在不影响阅读的情况很好的保护数据的权限。
* 平台内所有的静态资源均经过数据加密，实现访问路径可消耗，单次访问路径使用后销毁，使访问路径不可逆，同时每次访问路径不同，保证资源库的安全。
* 系统提供细致的用户管理，可直观的看到用户的基本信息以及用户启用状态，提供用户过期时间配置。
* 平台提供用户组概念，可对不同用户进行分组管理，不同组的配置信息会直接应用于组内成员，例如 权限、登录控制、日志行为等。
* 平台默认提供用户的注册以及审核功能，同时提供快速三方集成，可通过简单配置完成第三方登录接口的整合，提供IP段免登录、黑名单过滤、注册审核等功能。
* 平台可记录用户在使用过程中产生的流量，可通过配置最大流量来控制用户的流量行为，支持IP段限制、用户组限制、用户单独指定等配置方式。
* 平台可对于用户的基本操作做出权限控制，包括用户对于某一分库中数据的 检索、查看、下载、附件预览、试读时间，试读页数等，平台提供多维度的权限控制，支持从组的维度、个人维度、数据维度进行权限分配，整体权限控制符合DAC安全模型。
* 平台提供快链管理、页面公告 可根据用户需求进行配置。
* 平台针对前台页面提供静态资源配置，可完成背景图、页面标题、文字介绍等前台内容的替换。
* 平台提供详细的日志统计，提供清晰明了的图表展示，包括但不限于用户的访问、检索、下载、阅读等行为，资源统计、数据统计、操作异常、登录监控、检索统计、时间发布、终端统计、类型偏好等，检索统计提供检索词排行、检索热词发布等内容。
* 支持采用nginx负载均衡，扩展网络设备和服务器的带宽、增加吞吐量、加强网络数据处理能力、提高网络的灵活性和可用性。
* 支持ES分布式检索集群，实现近实时搜索，分片自动调度，保证系统健壮，自动故障检测。
* 具有管理员管理、日志管理、模块设置、系统参数设置、系统状态报告、服务器动态监控等功能。