



1. 方案概述

本方案建设一套面向单位/部门的文档管理平台，实现资料集中归档、分级授权、全文检索与审计追溯。平台适用于内网或外网服务器部署，支持 Docker 部署，采用系统内置账号密码登录。

很多单位在文档管理上普遍遇到以下问题：

- 资料分散：文件在个人电脑、共享盘、邮件附件等多处保存，交接困难，人员变动后容易“找不到”
- 版本混乱：同一份文件多次修改后，无法快速判断哪个是最终版，容易误用旧版
- 权限不清：共享盘“要么谁都能看，要么谁都不能看”，既不安全也不好
- 扫描件不可检索：扫描 PDF、图片资料只能“看图”，无法通过关键词快速定位
- 追溯困难：误删、误传、越权下载发生后，缺少可核查的记录

方案重点解决：

- 资料统一归档：目录树管理，结构清晰，职责明确

- 权限清晰可控：按目录授权，权限自动继承，支持例外授权
- 查找高效：文件名与内容统一检索，支持模糊匹配与结果高亮
- 扫描件可搜索：对扫描 PDF 与图片进行 OCR 识别并纳入全文检索
- 可追溯：关键操作留痕，满足日常管理与审计需要

建设完成后，用户体验与管理效果体现为：

- 查找资料从“翻文件夹”变成“搜关键词”，资料定位更快、更准
- 权限按目录统一管理，避免“资料共享盘越放越乱、越放越不敢用”
- 扫描件同样可检索，历史资料不再变成“只能看不能搜”的存档
- 管理动作可追溯，发生误删、误传、越权访问时可快速定位责任与影响范围



2. 适用范围

2.1 文档类型

平台覆盖以下资料类型：

- Office 文档：Word、Excel、PPT
- PDF 文档：包含文本型 PDF 与扫描型 PDF
- 图片资料：jpg/png 等
- 扫描件：可包含彩色盖章材料

这些资料既可以按“制度/流程/模板/合同/项目档案”等业务逻辑归档，也可以按“部门-年份-事项”组织。目录结构由单位自行定义，平台负责保障结构稳定、权限清晰、检索可用。

2.2 归档方式

- 按目录逐步上传归档
- 支持上传 zip 压缩包并解压归档为目录与文件

归档的关键原则是“先确定目录结构与归档口径，再上传资料”。这样做的好处是后续资料持续增加时，目录不会越堆越乱，搜索结果也更容易定位到正确位置。

2.3 典型使用场景

- 制度与流程库：按制度类别、年份、版本归档，便于全员检索与查阅
- 合同与项目档案库：按项目、合同编号、供应商归档，支持扫描件检索与校对
- 会议纪要与通知库：按部门、会议类型、日期归档，便于跨人员查找与复盘
- 业务表格与模板库：按业务域归档，保证统一版本与便捷下载

3. 用户与角色

平台采用系统内置账号密码登录。角色可按单位习惯命名与配置，典型角色如下：

- 系统管理员：维护部门、账号、角色权限，查看审计与运行情况
- 部门管理员/资料管理员：维护本部门目录结构与授权范围，整理资料
- 普通用户：在授权范围内浏览、上传、下载资料

平台按权限展示内容，未授权的目录与文件不会显示，也不会出现在搜索结果中。

为了让权限管理更省心，实际落地中通常按“岗位/部门”分配角色：管理员只需要维护好目录授权规则，人员变动时只需调整账号所属部门或角色即可，无需反复给每个人逐条授权。

典型使用流程如下：

1. 管理员创建部门与账号，并配置角色权限
2. 部门管理员建立本部门目录结构，按目录配置授权
3. 用户在授权目录中上传资料或上传 zip 压缩包归档
4. 系统自动进行内容提取与 OCR 识别，完成后即可全文检索
5. 对识别不准确的扫描件，授权人员可人工校对并更新检索结果

4. 权限与目录管理

4.1 目录树管理

平台以“目录树”组织资料，目录结构与电脑文件夹一致，支持新建、重命名、移动等维护操作（由授权人员执行）。

为了避免目录随意生长，目录维护建议由“部门管理员/资料管理员”负责，普通用户只在既定目录中上传资料。这样目录结构更稳定，后续检索与统计也更准确。

4.2 目录级权限

权限以目录节点为边界配置，常见权限包括：

- 查看：可见目录与文件
- 上传：可在目录上传文件
- 下载：可下载文件
- 维护目录：可新增/调整目录结构
- 删除文件：可删除文件（推荐由资料管理员掌握）
- 授权管理：可为他人配置权限（推荐由部门管理员掌握）

权限设置遵循“按目录统一授权”的思路：同一类资料尽量集中在同一目录下，授权只做一次，目录下的所有文件自然遵循同一规则，管理成本更低。

4.3 权限继承与例外授权

- 权限默认向下继承：父目录授权后，子目录默认同步生效

- 支持例外授权：可在某个子目录额外开放或收回指定主体的权限

举例说明：

- “部门资料/2026”目录授权给部门全员查看后，“部门资料/2026/合同”与“部门资料/2026/会议纪要”会自动生效，无需重复授权
- 若“部门资料/2026/合同”仅允许少数岗位查看，则可以在该子目录单独配置更严格的权限

目录授权主体支持：

- 账号授权：针对具体人员设置权限
- 角色授权：对一类岗位统一授权，减少反复配置
- 部门授权：部门整体可见/可用，方便批量管理

4.4 审计追溯

授权、收权、删除等关键操作记录如下信息：

- 操作人、时间、操作类型
- 生效对象（目录/文件）
- 来源信息

除授权与删除外，平台也可对上传、下载、预览、OCR 校对等关键动作记录日志，方便在出现争议或问题时快速核查发生过什么。

5. 核心功能

5.1 文档管理

- 目录树管理：按权限维护目录层级
- 文件管理：上传、预览、下载，支持软删除恢复
- 标签与分类：支持按标签与业务分类筛选

文件管理能力包含：

- 上传校验：文件名、类型与目录权限校验，确保上传行为可控
- 预览体验：常用格式支持在线预览，减少反复下载
- 归档路径展示：搜索结果可回溯到所在目录，便于整理与复查

对客户来说，“文档管理”最重要的是两点：一是资料放进来后不会丢、不会乱；二是权

限按目录统一生效，能看什么、能做什么一目了然。平台通过目录结构 + 权限规则 + 审计留痕，把这两点落到日常操作中。

5.2 全文搜索

- 统一检索：同时检索文件名与文件内容
- 模糊匹配：支持模糊搜索与结果高亮
- 权限过滤：仅返回用户有权限访问的结果

全文搜索不仅支持“精确匹配”，也支持“只记得大概关键词”的模糊查找。用户可以像使用搜索引擎一样输入关键词，平台返回命中的文件，并高亮显示命中片段，避免打开多个文件反复确认。

搜索结果呈现包含：

- 命中关键词高亮片段
- 文件所在目录路径，便于定位与归档核对
- 常用筛选项：按部门、时间范围、上传人、标签等快速缩小范围

为了确保“安全 and 好用兼顾”，搜索结果会自动做权限过滤：用户看不到的目录与文件，也不会出现在搜索中。

5.3 扫描件 OCR 与校对

- OCR 识别：对扫描 PDF 与图片进行文字识别并纳入检索
- 人工校对：支持人工修正 OCR 识别文本
- 自动更新：校对完成后自动更新索引，搜索结果以校对后的内容为准

扫描件的痛点在于“能看但不好找”。平台会对扫描 PDF/图片自动做 OCR 识别，把图片中的文字转成可检索文本；当识别不够准确时，支持人工校对，保证关键资料的检索准确率。

OCR 校对机制满足：

- 校对入口明确：可在文件详情页直接查看与编辑识别文本
- 校对留痕：记录校对人、校对时间，便于追溯与复核
- 校对后即生效：无需重新上传文件即可更新检索内容

5.4 压缩包归档

- zip 上传：支持上传 zip 压缩包

- 自动解压：按压缩包内目录结构解压并归档到目标目录

压缩包归档用于快速整理历史资料，例如“按年份/项目/供应商”一次性归档，降低人工逐个上传的工作量。

5.5 统计报表

提供基础统计，便于管理与分析：

- 按部门/人员统计上传量、下载量
- 热门资料与高频检索关键词统计

统计的目标是“辅助管理”，例如：哪个部门资料沉淀更充分、哪些资料被频繁使用、大家常搜哪些关键词。对后续目录优化、资料补齐、制度宣贯都能提供参考。

6. 内容提取与检索机制

平台在文件上传后执行内容提取与建索引，使资料可被全文检索：

- 文本类文档：直接提取文字内容并建立索引
- 扫描类资料：先进行 OCR 识别，再将识别文字建立索引

OCR 效果与扫描清晰度、对比度等因素相关。平台提供人工校对机制，以满足对准确率要求更高的业务场景。

为保障可用性，平台对处理过程提供状态可见性：

- 文件状态：已上传、处理中、可检索、处理失败
- 失败处理：明确失败原因，支持重新触发处理

这意味着用户上传文件后，即使 OCR/内容提取需要一定时间，平台也能让用户清楚知道当前进度；如果某个文件处理失败，也能看到原因并重新处理，避免“上传后搜不到但不知道为什么”的体验。

7. 系统组成

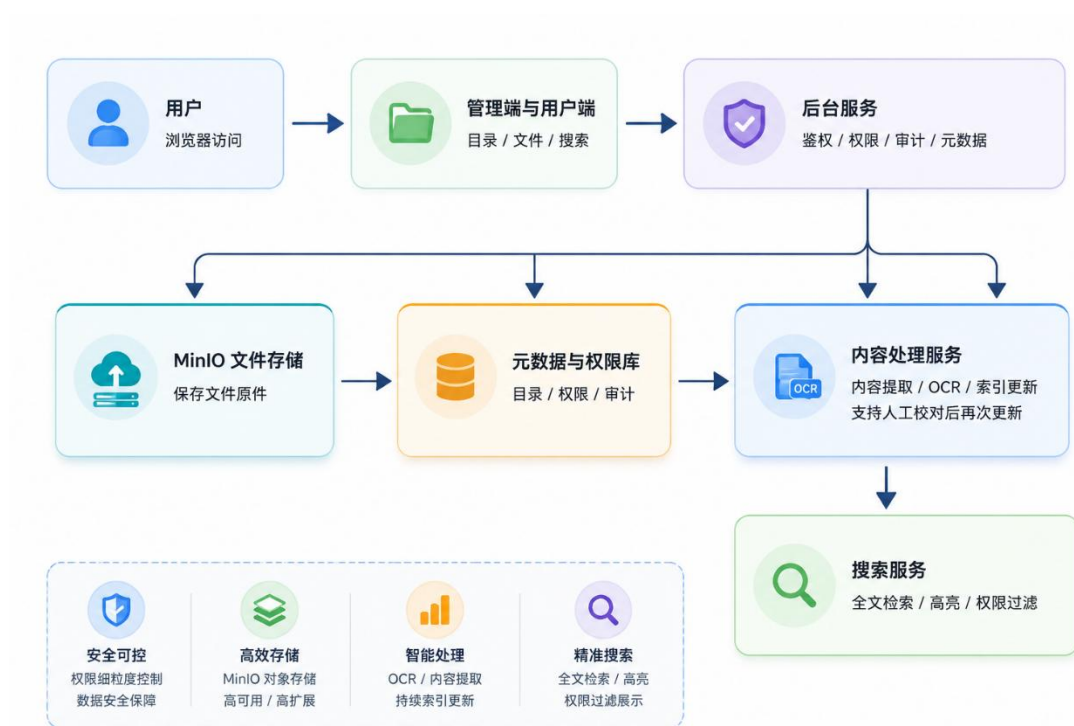
平台由以下模块组成：

- 管理端与用户端：目录、文件、搜索、授权、统计报表
- 后台服务：登录鉴权、权限校验、目录与文件元数据管理、审计
- 文件存储：使用 MinIO 在服务器侧保存文件原件
- 内容处理服务：内容提取、OCR、索引更新

- 搜索服务：全文检索与结果高亮

从客户视角理解，这套系统可以简单理解为：文件原件放在 MinIO 里，文件信息与权限在“后台服务”里，文件内容的索引在“搜索服务”里。这样既能保证检索速度，也便于后续扩容与运维。

系统架构示意图如下：



8. 部署方案

平台支持在内网或外网服务器部署，支持 Docker 部署方式，也可部署在物理机或虚拟机环境。服务器资源配置按预计用户数、并发在线人数、文件总量与年增长进行规划。

外网部署场景通常会配合单位的网络边界策略（例如只开放必要端口、配置访问控制、使用 HTTPS 证书等），以满足公开访问与安全管控的平衡。

8.1 单机部署

适用于小规模单位或初期上线：

- 单台服务器部署平台与基础组件
- 扫描件较多时可配置 GPU 以提升 OCR 处理效率

8.2 分布式部署

适用于多人同时使用、扫描件较多、对稳定性要求较高：

- 后台服务支持多实例部署，便于扩容与容灾
- OCR 与内容处理服务独立部署，便于按处理压力扩展
- 文件存储与搜索服务独立部署，便于容量扩展与运维

当文档数量持续增长或 OCR 处理量变大时，分布式部署可以通过增加内容处理服务的资源来提升处理能力，而不影响用户端的日常访问与检索体验。

9. 实施交付流程

- 需求梳理：组织结构、目录规则、权限策略、审计范围
- 环境准备：服务器、网络与存储资源，证书与访问策略（外网部署时）
- 系统部署：Docker 或传统方式部署并完成初始化配置
- 组织与权限初始化：部门、角色、账号、目录结构与授权规则配置
- 资料归档：按目录上传资料，支持 zip 解压归档；执行内容提取与 OCR 建索引
- 校对流程启用：为扫描件提供 OCR 文本校对入口与管理流程
- 培训与验收：管理员与用户培训，按验收用例验收并形成验收记录
- 上线运行：持续监控、备份、容量管理与问题处理

以上流程的交付成果是：账号体系与目录结构跑通、权限配置落地、资料开始可检索、扫描件可 OCR 校对、审计日志可查询。客户可以在日常工作中直接使用，并在后续持续把资料归档到平台中。

10. 交付与验收

10.1 交付物

- 系统软件：管理端、后台服务、内容处理服务、搜索服务部署包
- 文档资料：
 - 管理员手册：部门/账号/角色/权限/审计/运维要点
 - 用户手册：目录、上传、zip 归档、搜索、预览、下载

- 运维手册：部署、备份、监控、告警、升级与故障处理
- 培训与验收：
 - 管理员培训
 - 上线支持
 - 验收用例与验收记录模板

10.2 验收要点

- 权限控制：不同账号在同一目录下的可见、可上传、可下载行为与授权一致
- 目录管理：授权人员可维护目录树结构，未授权人员无法操作
- 文件上传与预览：常用格式上传成功且可预览，下载行为可审计追溯
- zip 归档：压缩包上传后按原目录结构解压归档，文件可正常预览与检索
- 全文检索：文件名与内容可检索，命中片段高亮展示
- OCR 检索：扫描件可检索，OCR 文本可人工校对，校对后检索内容更新
- 审计日志：授权、删除、下载等关键操作留痕可查询

验收时建议准备一套样例资料（包含 Word/PDF/扫描件/图片/zip），以及两到三类典型账号（管理员、资料管理员、普通用户），用真实操作来验证权限、检索、OCR、审计等关键能力。

11. 运维与审计

- 权限最小化：高风险权限按岗位授予并可审计追溯
- 审计与操作日志：长期保留，管理员可按管理要求手动清理
- 备份策略：数据库与文件存储按周期备份，明确恢复演练机制
- 访问控制：按网络边界与访问范围配置访问控制策略

日常运维包含：

- 运行监控：接口可用性、任务队列积压、OCR 处理耗时、搜索响应时间
- 容量管理：文件存储容量增长、索引容量增长、备份空间增长
- 权限巡检：定期复核关键目录的授权范围与高风险权限分配

在实际管理中，建议把“关键目录的授权变更”纳入固定流程，例如由部门管理员执行、管理员定期复核，既保证日常调整灵活，又避免权限长期漂移。

12. 风险与应对

- OCR 识别率波动：通过扫描规范与人工校对机制保证可用性与准确率
- 权限规则过细：以目录继承为主、例外授权为辅，保持规则可维护
- 资料归档工作量：先规划目录结构与归档口径，再按目录逐步上传，优先处理高频资料
- 搜索效果优化：持续沉淀业务词库与常用关键词，提升检索命中率与可用性

为了让客户更省心，建议把风险应对固化为“可执行的管理动作”：

- 扫描规范：统一分辨率、对比度与方向，显著提升 OCR 命中率
- 校对流程：对关键扫描件明确“谁校对、校对到什么程度、校对后谁复核”
- 授权治理：目录继承为主，例外授权尽量少且可追溯，避免权限越配越复杂
- 词库沉淀：把常用简称、单位内部术语逐步沉淀为检索关键词，提高搜索体验