超级管理员使用

配置检索方式与处理参数

检索方式（快速/精确）、单文档内容额度、批次与并发。

场景：调整企业资料的检索方式、单文档内容额度和处理批次参数。

角色：超级管理员。

路径：管理后台 → 【检索】。

1 查看当前生效配置

页面顶部的「当前生效配置」显示站点已保存的显式配置：索引模式、检索方式、并发、文档额度、向量模型批次和最近更新时间。新发布任务会把这些值固定为任务快照。



2 选择检索方式

在「配置新发布任务」中选择检索方式：快速优先速度，适合先看有没有相关内容；精确在多路召回和图谱扩展后使用大模型排序，结果更完整但耗时更长。索引范围固定为

完整模式，两种方式基于同一份完整索引，用户搜索时仍可临时切换。

3 调整文档额度

「单文档内容额度」限制新上传解析、Markdown 编辑和新发布的单份规范化内容，默认 8 MiB、最低 1 MiB；页面同时显示管理员配置、运维硬上限和实际生效值。超过运维上限的配置会被明确拒绝。调低额度不影响历史内容，但之后重新编辑或发布必须满足新额度。

4 调整批次与并发

三个参数各有用途：文档 LLM 分块并发（单份文档同时进行的抽取数）、向量模型批次（一次向量请求包含的文本数）、向量索引写入批次（一次写入的记录数）。修改后点击【保存配置】。

> **注意**
>
> 检索方式和执行参数只影响之后创建的发布任务，已发布内容不受影响。调整部署硬上限需由运维修改环境变量并重启服务。