

档案软件 - 识别结果全文检索功能【云丰档案软件功能介绍】

档案管理软件 · 功能介绍

把 OCR 出的文字真正用起来：在成千上万页里按关键词一秒定位。识别结果全文检索对 OCR 识别后的全文建立索引，输入关键词即定位到具体页与具体位置，扫描件也能像网页一样搜索。

一、为什么需要识别结果全文检索

OCR 出了文本，若不建索引，还是不能搜。全文检索才把识别成果转化为实际的查找能力。

有文本不能搜：识别出了文字却没索引，查找仍要逐页翻；

定位不精确：告诉在哪卷不够，还要知道在第几页；

命中不高亮：搜到了却不知道是哪句话命中，还要再找；

海量里难找：几十万页里靠关键词，没有索引根本来不及。

全文检索让 OCR 文本变成搜索引擎：关键词一输，直接跳到命中页，扫描件查找效率质变。

二、怎么用：输入关键词，定位到页

在检索框输入关键词，系统在已识别的全文中查找，返回命中的案卷并高亮定位到具体页。

表1 识别结果全文检索流程

步骤	操作要点
输入关键词	在全文检索框输入要找的词；
建立索引	系统对 OCR 全文预先建索引；
秒级返回	在海量页面中找出命中位置；
高亮定位	打开页面高亮命中词句；
按需查阅	跳转到命中页直接阅读。

检索质量取决于 OCR 质量。全文能搜多准，前提是识别文本校得准。错字会导致漏搜，因此关键档案的 OCR 校对直接影响检索体验。

检索结果高亮命中词句，直接跳到对应页。结果按本人权限过滤，不越权暴露。

检索同时命中题名、著录与正文文本，一处关键词多源命中。结果可按相关度与年度排序，快速定位目标页。

全文检索让扫描件像网页一样可搜。检索同时命中题名、著录与正文文本，一处关键词多源命中，结果高亮命中词句并直接跳到对应页，按相关度与年度排序。几十年的纸质档案数字化后，按人名、项目名、事项名即可秒级定位到具体页，查档效率发生质变。结果严格按本人权限过滤，不越权暴露。随着 OCR 与校对质量提升，检索准确率持续改善。这一能力把沉睡的历史资料变成可即时调用的知识，档案服务的深度与广度都上了台阶。

全文检索把 OCR 成果的价值彻底释放。它与题名、著录检索互为补充，让用户既能按目录找、也能按内容找。随着校对与批量识别能力提升，检索命中越来越准，档案从静态资料真正变为可即时调用的知识资源。

三、核心能力

表2 核心能力一览

能力	一句话价值	关键细节
全文索引	海量里秒搜	对 OCR 全文建索引，几十万页也能秒级命中。
定位到页	翻到具体处	结果不仅给案卷，还定位到具体页与词句。
高亮命中	一眼看到	命中关键词在页面中高亮显示，不必再找。
模糊容错	近似也能搜	支持近似词与常见错字的容错，提高命中率。
权限过滤	只搜有权的	检索结果按本人权限过滤，不越权暴露。

几十年的纸质档案数字化后，按人名、项目名一秒定位到具体页，查档效率发生质变。

四、典型使用场景

查旧文件：在几十年的老档案里按关键词找到某份批示。

主题调研：搜某人名或项目名，找出所有相关页面。

证据查找：审计时按关键词在海量卷中定位证据页。

五、价值与产品定位

全文检索的价值在于把 OCR 成果转化为查找力：扫描件像网页一样可搜，几十年档案一秒定位。扫描件像网页一样可搜，档案利用深度由此提升。

它是云丰档案利用能力的核心，建立在 OCR 与双层 PDF 之上，使档案从“存得下”走向“找得到、用得上”。