

档案软件 - 可检索 PDF 生成功能【云丰档案软件功能介绍】

档案管理软件 · 功能介绍

换个角度说，双层 PDF 也就是能搜的 PDF，是数字化归档的目标格式。可检索 PDF 生成将扫描件经 OCR 后导出为可按关键词搜索、可复制文字的 PDF，让每份电子档案都能被全文检索命中。

一、为什么需要可检索 PDF 生成

不少单位的扫描档案仍是“死图片”，要找一段内容只能逐页翻。可检索 PDF 是让档案真正可用的关键一步。

查找靠翻：纯图片 PDF 搜不了，查内容逐页翻找；

引用靠抄：要引用文字只能照着敲，慢且易错；

归档价值低：不能检索的电子档案，和纸质差别不大；

跨库难搜：不可检索就无法纳入全文检索范围。

生成可检索 PDF 让每份扫描档案都能被搜索命中：找内容一秒定位，要引用直接复制。

二、怎么用：OCR 后导出可检索

扫描件经 OCR 识别后，系统导出带文字层的可检索 PDF，纳入全文索引。

表1 可检索 PDF 生成流程

步骤	操作要点
扫描加工	完成去污纠偏等预处理；
OCR 识别	识别页面文字并校对；
导出 PDF	生成带文字层的可检索 PDF；
建索引	文字纳入全文检索；
归档利用	关键词即可定位。

可检索建立在识别准确上。PDF 能不能被准确搜到，取决于 OCR 与校对质量。重要档案建议先校对再生成，避免错字导致漏检。

生成后自动纳入全文索引，按关键词即定位。重要档案建议先校对再生成，避免错字漏检。

生成后自动入库全文索引，检索零延迟。可检索程度随校对质量提升，错字越少检索越准。

可检索 PDF 让扫描档案从看得到升级为搜得到。系统在 PDF 中嵌入文本层，生成后自动纳入全文索引，按关键词即定位、零延迟。重要档案建议先校对再生成，避免错字导致漏检；随着校对质量提升，可检索程度不断提高。把历史死图片档案批量转为可检索，全库按主题查找成为可能，数字化才算真正落到实处。对依赖快速查档的审计、稽核、业务部门，可检索 PDF 极大缩短了资料定位时间，档案的利用价值被充分释放。

可检索 PDF 把死图片变成活文本，是数字化成果价值兑现的标志。生成即入索引、按词即定位，让历史档案服务于日常业务。它与校对质量相辅相成，持续提升检索体验，是档案利用深度的集中体现。

生成即入索引、按词即定位，把历史图片档案转为可即时调用的资源，是数字化成果价值兑现的标志性一步。

三、核心能力

表2 核心能力一览

能力	一句话价值	关键细节
关键词可搜	一页不漏	PDF 内文字可被全文检索命中。
文字可复制	引用不手敲	正文文字可选中复制，免去照抄。
保留原貌	看图是原件	视觉上仍是原始扫描件，不失真。
批量产出	整卷转换	一卷文件批量生成可检索 PDF。
纳入索引	全库可搜	生成后自动进入全文检索范围。

把历史死图片档案批量转为可检索，全库按主题查找，数字化才算落到实处。

四、典型使用场景

存量转可检索：把历史死图片档案批量转为可检索 PDF。

新卷归档：新扫描件直接生成可检索格式入库。

跨库检索：纳入全文索引后，全库按主题查找。

五、价值与产品定位

可检索 PDF 的价值在于把扫描档案从“看得到”升级为“搜得到”：每份文件都能被关键词命中，数字化才算落到实处。扫描档案从看得到升级为搜得到。

它是云丰档案 OCR 能力的对外交付形态，与全文检索、双层 PDF 一体，支撑档案的深度利用。