

档案软件 - 双层 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

原图层加文字层叠在一起，既看得到原件又搜得到文字，这就是双层 PDF。双层 PDF 生成把扫描图像与 OCR 识别文本合成双层 PDF，下层是清晰原件、上层是可检索文字，归档标准格式由此形成。

一、为什么需要双层 PDF 生成

纯图片 PDF 看着是原件却不能搜，纯文字 PDF 又丢了原貌。双层 PDF 把两者优点合到一个文件里。

图片不可搜：扫描件 PDF 只能翻页，关键词检索无门；

文字失原貌：转成纯文字后，手写签章与版式全没了；

格式不合规：档案归档要求保留原貌又要可检索，普通 PDF 难两全；

查阅不便：不能复制文字、不能定位，利用效率低。

双层 PDF 是扫描档案的标准归宿：原貌一层、文字一层，看着是原件，搜得到、也复制得出。

二、怎么用：图像加文字层合成

系统把扫描图像作为底层、OCR 文本作为上层叠加，导出为标准双层 PDF 归档。

表1 双层 PDF 生成流程

步骤	操作要点
图像入库	扫描图像完成去污纠偏；
OCR 识别	识别出页面文字；
叠合成层	文字层与图像层对齐叠加；
生成 PDF	导出为双层 PDF；
归档利用	可检索可复制，保留原貌。

双层不是两个文件。图像与文字封装在同一个 PDF 里，不额外占用文件数。视觉上看到的是原件，选择时才能选中文字，二者互不干扰。

图像与文字封装在同一文件，视觉上看原件，选中时才取文字。整卷批量合成。

双层 PDF 的文本层与图像层严格对位，复制文字不串行。文件体积经优化，便于长期在线调阅。

双层 PDF 是电子档案长期保存的标准形态。系统把原始扫描图像与 OCR 文本层封装在同一文件中：视觉上看是原貌，选中文本时则可复制、可检索，两层严格对位、复制不串行。文件体积经优化，便于长期在线调阅与离线归档。这样的 PDF 既忠实保留档案原貌，又具备文本的可利用性，两全其美。整卷批量合成后自动纳入全文索引。对存量纸质档案，生成双层 PDF 是数字化成果入库的关键一步，使其同时满足真实归档与便捷利用的双重要求。

双层 PDF 是连接扫描原貌与电子文本的桥梁，也是电子档案长期保存的推荐形态。它一份文件同时满足真实归档与便捷利用，批量生成后纳入索引，是存量数字化从加工走向正式入库的关键转化。

三、核心能力

表2 核心能力一览

能力	一句话价值	关键细节
图文双层	原貌可检索	下层保留原件图像，上层叠加可搜文字。
定位复制	像网页一样用	可复制文字、按关键词定位到页。
标准格式	归档合规	符合电子档案格式要求，长期可读。
层级对齐	字图对应	文字层与图像位置精确对齐，不串位。
批量生成	整卷产出	一卷扫描件批量合成双层 PDF。

这样的 PDF 既保留原貌可阅，又能检索复制，正是电子档案长期保存的标准形态。

四、典型使用场景

存量数字化：老档案扫描件统一转双层 PDF 归档。

合规移交：按移交要求提交可检索的双层 PDF。

日常查阅：调阅时既看原件又能复制引用。

五、价值与产品定位

双层 PDF 的价值在于让扫描档案“两全其美”：既是原貌电子件，又是可检索文本，这正是数字化归档的标准形态。扫描档案两全其美，既是原貌又是可检索文本。

它是云丰档案 OCR 与图像处理能力的最终产物，与全文检索、著录体系衔接，是电子档案长期保存的标准格式。