

档案软件 - 识别结果校对功能【云丰档案软件功能介绍】

档案管理软件 · 功能介绍

OCR 再准也会有错，关键错误必须人工校出来，否则检索就会漏。识别结果校对把识别文本与原扫描页并排呈现，供人工逐字核对修改，纠错后的文本才用于检索，保证数字化质量。

一、为什么需要识别结果校对

OCR 对手写、模糊、印章、表格容易认错，错字若直接入库，将来按关键词就搜不到这份档案。

错字漏检索：关键名词认错，将来按正确词反而搜不到；

对照不便：要在图片和文本间来回切换才能核对；

校对无重点：通篇逐字校太费时，不知哪里最可能错；

改了不存：校对结果若不回写，下次又要重来。

校对是 OCR 质量的最后一道关：图文对照、错处即改、改后入库，让识别文本真正可信可用。

二、怎么用：图文对照，逐处修改

打开校对界面，左图右文对照阅读，对识别错误处直接修改，系统标出低置信度字供重点核对。

表1 识别结果校对流程

步骤	操作要点
打开校对	进入待校对的识别任务；
图文对照	原图与识别文本并排显示；
重点核对	低置信度字词优先检查；
直接修改	错处就地编辑改正；
保存回写	校对结果回写，供检索使用。

校对只改文本，不动原图。校对修改的是识别文本层，原始扫描图像保持不动，档案原貌不被改动。这样既保证检索准确，又维护了档案真实性。

校对只改识别文本层，原始扫描图像保持不动，维护档案原貌。低置信度字优先核对。

校对提供热词库，单位名、地名、专业术语可预置，减少同类错字。校对留痕，谁改了哪一处可查，责任清晰。

校对是守住 OCR 质量底线的关键一环。系统只改正识别文本层，原始扫描图像保持不动，完整维护档案原貌；低置信度字优先提示核对，重要档案逐页图文对照。提供热词库，单位名、地名、专业术语可预置，减少同类错字。校对全程留痕，谁改了哪一处可查，责任清晰。因为错字一旦入库，将来检索就会漏，认真校对才能保证按关键词真正搜得到。这道工序看似琐碎，却直接决定了全文检索的可用程度，是数字化成果质量的重要保障。

校对环节看似增加一道工序，实则是 OCR 成果质量的最终守门人。错字少一字，将来检索就多一分准头。系统通过热词、留痕与低置信提示，让校对既有针对性又高效，为全文检索与著录质量打下坚实基础。

三、核心能力

表2 核心能力一览

能力	一句话价值	关键细节
图文对照	校对不费力	原图与文本并排，逐字核对一目了然。
置信度提示	易错处先看	系统标低置信度字词，优先核对高风险点。
就地修改	改得快	错处直接编辑，所见即所得。
结果回写	改了就生效	校对后文本立即用于检索与著录。
原图不动	保真	只改识别层，原始扫描图像保持原貌。

重要档案逐页图文对照校对，确保按正确关键词能搜到；错字若入库，将来检索就会漏。

四、典型使用场景

关键卷校对：重要档案逐页图文校对，确保检索准确。

低质页复核：模糊手写页重点核对识别结果。

批量抽检：加工队对批量 OCR 结果抽样校对。

五、价值与产品定位

识别结果校对的价值在于守住 OCR 质量底线：错字当场纠正，检索才不会“该搜的搜不到”。守住 OCR 质量底线，检索才不会该搜搜不到。

它与批量 OCR、全文检索衔接，构成云丰档案数字化质量闭环，使自动化效率与人的把关各得其所。