

发票去重 · 使用文档

上传一批 PDF 发票，自动识别发票号码并查重：哪个文件重复、与哪张重复，一目了然，一键导出。

一、功能介绍

- 「发票去重」是发票处理工具的第三个功能页签，解决"一个文件夹里混进了重复发票"的盘查问题：
- 自动识别：上传后系统自动解析每个 PDF 的发票号码，无需任何手动输入
 - 智能判重：同一发票号第二次及以后出现即标记「重复」，并列出于哪张文件重复；首次出现的保留为「唯一」
 - 结果清晰：统计栏（总数 / 唯一 / 重复 / 未识别）+ 明细表（文件名 / 发票号码 / 查重结果 / 重复于）
 - 批量高效：单次最多 300 个 PDF（超过 50 个自动分批，跨批结果统一汇总）；支持选择整个文件夹自动收录（含子文件夹）
 - 随时再查：上传区常驻不消失，结果直接显示在下方，换一批文件接着查
 - 一键导出：查重结果导出 Excel，留档或转发同事
 - 隐私安全：识别在服务器独立进程进行，完成后即删；网站不保留任何形式的用户数据，用完即删
- 与扫码核对查重的区别：扫码核对用扫码枪逐张扫、边扫边查，适合报销窗口实时拦截；发票去重适合事后批量盘查一个文件夹里的历史发票，一次全查完。
- 适用人群：月末盘查报销附件、归档前查重、交接清点电子发票存档的财务与审计人员。

二、详细操作步骤

第 1 步：选择文件

- 三种方式任选：
- 点击上传：点击上传区 → 弹出文件选择框 → 多选 PDF 文件
 - 拖拽上传：从资源管理器拖一批 PDF 到上传区
 - 整个文件夹：点「选择整个文件夹」→ 文件夹内全部 PDF 自动加入（含子文件夹）
- 要求：PDF 电子发票，单文件 ≤ 20MB，单次 ≤ 300 个。混入的非 PDF 文件自动跳过。

第 2 步：等待识别

松开鼠标后系统自动开始解析发票号，页面显示"正在识别发票号..."进度提示。超过 50 个时自动分批查重（每批 50 个），每批完成的结果即时追加显示，跨批发票号统一汇总判重。识别在服务器独立进程进行，完成后自动展示结果——上传区保留在原处，不需要等待页面变化。

第 3 步：查看结果

- 识别完成后，下方显示：
- 统计栏：共 N 个文件 / 唯一发票 N 张 / 重复 N 张 / 未识别 N 个（无重复或无未识别时对应项自动隐藏）
 - 明细表：每行一个文件，4 列——

列	说明
---	----

文件名	上传时的原始文件名
发票号码	自动识别出的发票号码（未识别为空）
查重结果	唯一（绿）/ 重复（红）/ 未识别（灰斜体）
重复于	重复项指向首次出现的同名发票文件

发现重复时弹出提示"查重完成：发现 N 张重复发票"，重复行红色标注，一眼定位。

第 4 步：导出结果（可选）

点「导出查重结果」→ 下载 Excel 文件（文件名含时间戳，如`发票去重结果\_20261008-220101.xlsx`），4 列与页面表格一致，可直接留档或转发。

换一批再查

上传区常驻——直接再选/再拖一批新文件即可开始新一轮查重，新结果自动替换旧结果。

三、常见问题解答

Q1：怎么判断重复？

系统自动识别每个 PDF 的发票号码，同一号码第二次及以后出现即标记「重复」，并列出于哪张文件重复，首次出现的保留为「唯一」。

Q2：能处理多少文件？

单次最多 300 个 PDF（选文件夹时自动加入全部；超过 50 个自动分批查重，跨批结果统一汇总）。识别在服务器独立进程进行，完成后即删，不保存任何数据。

Q3：「未识别」是什么意思？

该文件无法解析出发票号：可能是扫描件（图片式 PDF）、损坏文件或不是发票的 PDF。这类文件不计入重复，建议人工确认。

Q4：和扫码核对查重有什么区别？

扫码核对用扫码枪逐张扫、边扫边查，适合报销窗口实时拦截；发票去重适合事后批量盘查一个文件夹里的历史发票，一次全查完。

Q5：重复的发票文件会被自动删除吗？

不会。本功能只识别和标注重复，不删除、不移动任何文件——删除哪些、保留哪张，由您人工决定后再操作。

Q6：同一张发票既有 PDF 又扫过码，两边能互通吗？

不能。两个功能的数据互相独立：发票去重只对本次上传的 PDF 文件判重，扫码核对只对本次扫入的二维码判重。都需要时请分别使用。

Q7：文件名不同但发票号相同，算重复吗？

算。判重只看发票号码，与文件名无关——同一张发票用不同文件名存了两份，照样标出"重复"。

Q8: 导出的 Excel 里有什么?

与页面表格完全一致的 4 列: 文件名、发票号码、查重结果、重复于。查重结果列为"唯一 / 重复 / 未识别"文字标注。

Q9: 会上传保存我的发票吗?

不会。文件仅用于本次识别, 解析在服务器独立进程完成后即删; 网站不保留任何形式的用户数据。

Q10: 下载的文件在哪里?

在浏览器的默认下载文件夹 (通常是「下载」目录)。点击浏览器右上角的下载图标可直接找到, 也可按 Ctrl + J 打开下载列表定位。

如果有其他问题或需求, 可以联系技术处理, 技术微信: lj101450