

KdanAuto 使用帮助

一、为什么要做这玩意？

QA 做为 ComPDFKit_Conversion(以下统称 SDK) 交付品质把关最后一环，人力投入不足，无法满足SDK 内容 **随测**、以及每个版本的 **通测**，更加没有人力去做每次提测版本的全文档、各种情况覆盖测试，SDK 交付品质难以得到保障。同时，如果品质出现问题，整个团队又需要消耗掉更多的时间和精力去应对各种突发状况。

SDK 测试大致工作量如下：

- 转档SDK支持14种格式，每种格式又分别涉及注释、签名、表单等各情况测试；
- 命令行工具各种参数组合保守估计多达300-400种，每一种参数组合按一个 PDF 文档（一页）算
- 工作量（不算转档实际耗时，）只算对比结果，每轮通测所需要的人力消耗大致为（文档转换 14种格式 x 100份文档 / 10小时工作时间 / 60分钟 = 2.3 人.天）
- 同时纯人工识别对照，还存在一定的失误率（例如一篇文章中，错了几个字，缺了几个字是无法肉眼察觉的）。

Note：随测和功能通测都是最耗费人力的

二、KdanAuto 能干什么？

KdanAuto 里面直接集成转换 SDK，能够对指定目录内的文档，进行一键式自动转档，并同步将转换出来的档案与前一个（SDK）版本进行预览结果对照、生成对照报告。给 QA 后续将要执行的新功能测试做粗过滤、做参考。

对于字符串文档格式，如 RTF、TXT，由于要对比的内容为纯字符，人工无法快速进行字符准确度精准对比，可以使用KdanAuto 工具提供的字符对比工具，进行逐个字符对比，并在测试报告中，给出字符对比效果。以利观察字符对照差异。

同时可以对【命令行工具 注释、高亮、加水印、提取等编辑功能自动化执行生成的结果文档】做对照，粗过滤出存在差异的内容，供 QA 精准测试。

具体功能有

- KdaAuto对照工具 -> 导入文档到不同格式转档目录 -> 执行一键转档 -> 自动对照转档结果（与之前版本结果的差别） ->生成对照报告
- KdanAuto字符对比工具 -> 导入文档到不同格式转档目录 -> 执行一键转档/OCR(RTF、TXT) -> 自动对

比转档结果（与对照文档的字符差别）->生成字符对比报告

- 命令行对照工具->任意平台，命令行批量编辑（注释、加水印、高亮等）->将结果文档导入 KdanAuto -> 一键（与之前版本结果进行）快照对照->生成对照报告
- 跨平台字符对比工具->任意平台，批量转换/OCR RTF/TXT->将转档结果文档导入 KdanAuto -> 自动对比转档结果（与对照文档的字符差别）->生成字符对比报告
- 可考虑作为工程师自测工具，确保开发的功能块，不会对其它的功能模块产生影响

三、具体有哪些功能？

1、新建测试项

以添加“TXT”格式测试项为例，右击左侧列表中的格式行，点击弹出菜单中的“新建转【TXT】测试项”，在弹出来的窗口中，设置以下属性：

- 测试项名称：用于测试项目的名称，仅用于阅读识别分类区分
- 测试项文件夹：对应于“(功能) 3”中，设置测试文件夹目录 中的子目录名称
- 测试类型：测试类型目前支持4种，快照对照测试，PDF 对照测试，字符对比测试、跨平台对比测试

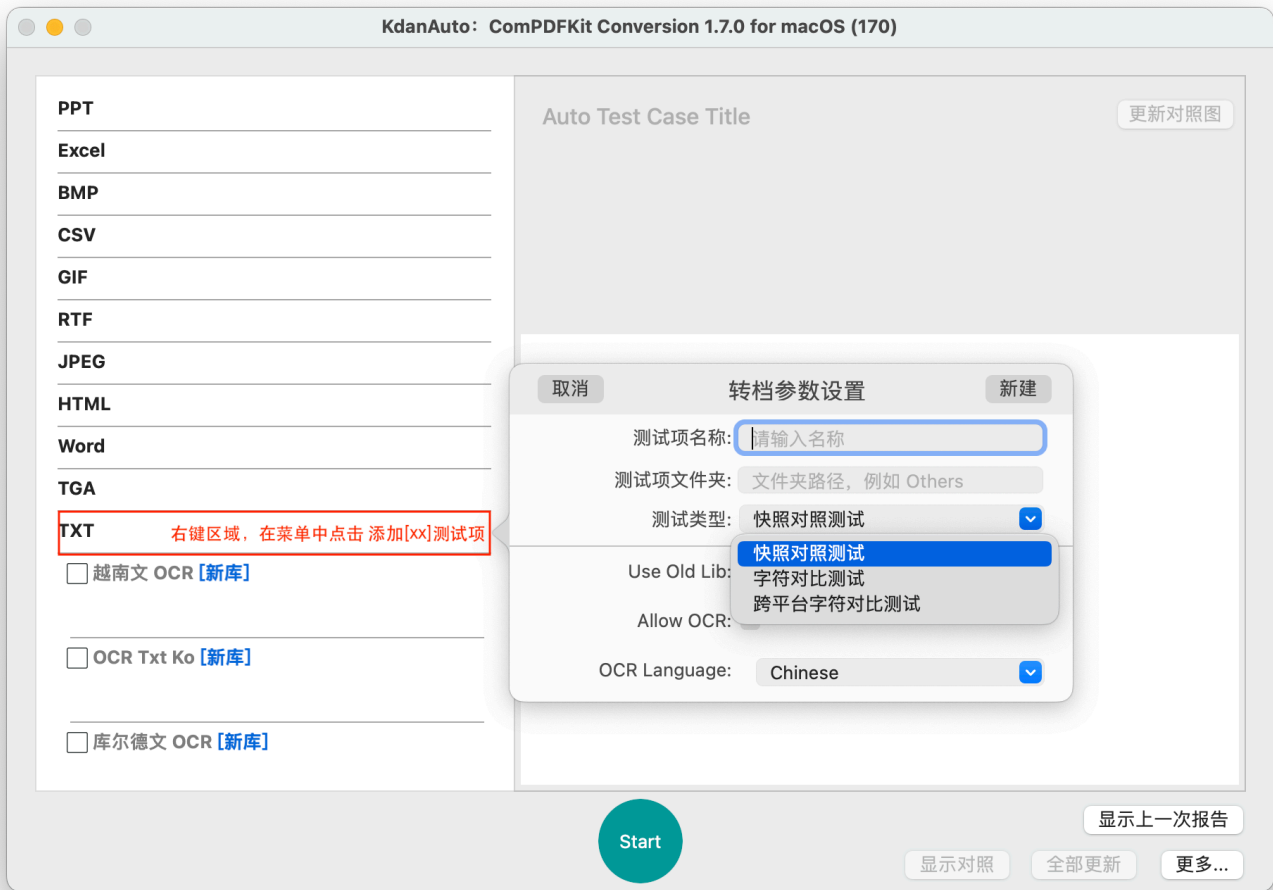
// 其它参数

- Use Old Lib： 是否是使用老库，用于切换 ComPDFKit 新库和之前徐建伟那套库
- Allow OCR： 转 TXT 时，SDK 中提供的参数，是否开启 OCR
- OCR Language： 转 TXT 时，SDK 中对应指定的 OCR 识别语言

Note： 测试类型说明

- 快照对照测试
 - 对新（如1.7.0）/旧（如1.6.0）发布的转档SDK版本，转档/OCR出来的文档，进行快照，生成 JPG 图片（除JPG、PNG 等图片格式外，其他格式均只有一页），进行图片相素级对比，对生成文档差异进行粗过滤
 - 支持左侧列表中，除 PDF 格式外的所有格式类型
- PDF对照测试
 - 用于对其他平台（如 Windows、Android、Saas、命令行工具），以任意形式（如编辑、添加注释等）操作后的文档，导入，跟 前一次（做同样操作）处理后的文档，进行快照，生成 JPG 图片（全部页面），进行图片相素级对比，对生成文档差异进行粗过滤
 - 仅支持左侧列表 PDF 格式类型
- 字符对比测试
 - 用于对PDF 文档，进行 OCR/转档 生成 RTF/TXT 格式，再将生成的文档，与原始（RTF/TXT）文档，进行逐个字符对比，以检测出在转档/OCR 过程中转档/OCR生成字符的准确度
 - 仅支持 TXT、RTF 格式
- 跨平台对比测试
 - 用于其他平台（如 Windows、Saas 等），对PDF 文档进行 OCR/转档生成 RTF/TXT 格式，再将生成的文档，与原始测试范例素材资料（RTF/TXT）文档，进行逐个字符对比，以检测出在转档/OCR 过程中，对字符处理的准确度
 - 仅支持 TXT、RTF 格式

相关属性设置好后，点击右上角【新建】。（如下图）



2、编辑测试项

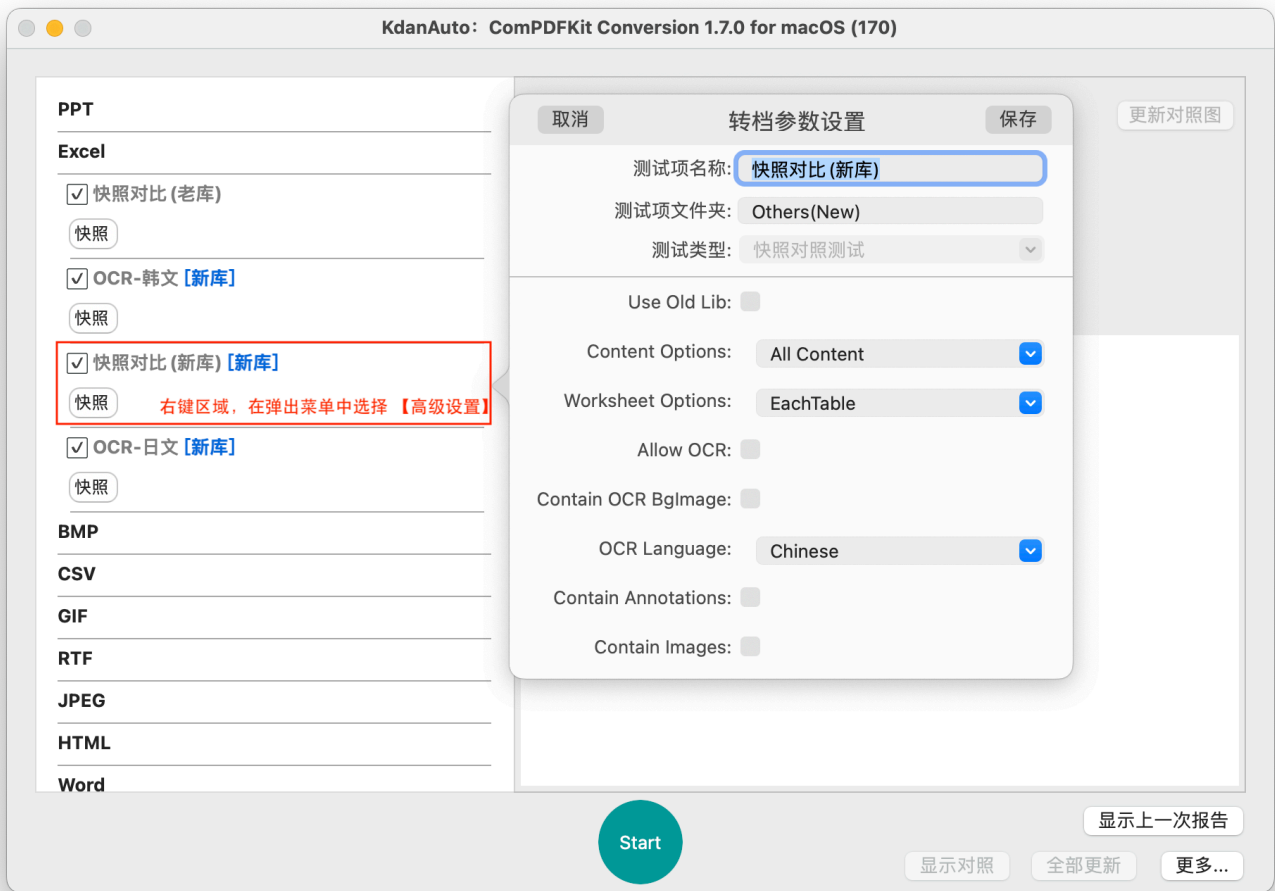
以修改“Excel”格式“快照对比(新库)”测试项为例，右击左侧列表中的 Excel 格式中的 '快照对比(新库)', 弹出菜单中的“高级设置”，在弹出来的窗口中，设置以下属性：

- 测试项名称：用于测试项目的名称，仅用于阅读识别分类区分，可修改
- 测试项文件夹：对应于“(功能) 3”中，设置测试文件夹目录 中的子目录名称，不可修改
- 测试类型：测试类型目前支持4种，快照对照测试，PDF 对照测试，字符对比测试、跨平台对比测试
 - 但 Excel 格式只支持 快照对照测试，所以当前示例不可修改

// 其它参数

- Use Old Lib: 是否是使用老库，用于切换 ComPDFKit 新库和之前徐建伟那套库
- Content Options: 转 Excel 时，SDK 中提供的参数
- Worksheet Options: 转 Excel 时，SDK 中提供的参数
- Allow OCR: 转 Excel 时，SDK 中提供的参数，是否开启 OCR
- Contain OCR BGIImage: 转 Excel 时，SDK 中提供的参数，OCR 是否包含背景图片
- OCR Language: 转 Excel 时，SDK 中对应指定的 OCR 识别语言
- Contain Annotations: 转 Excel 时，SDK 中提供的参数，是否包含 Annotations
- Contain Images: 转 Excel 时，SDK 中提供的参数，是否 包含图片

修改相关属性设置好后，点击右上角【保存】。（如下图）



3、设置测试文件夹目录、建立文件夹、导入文件

点击 App 右下角“更多...”按钮，在弹出的设置菜单中，设置自动化测试文件根目录（如下图所示）。



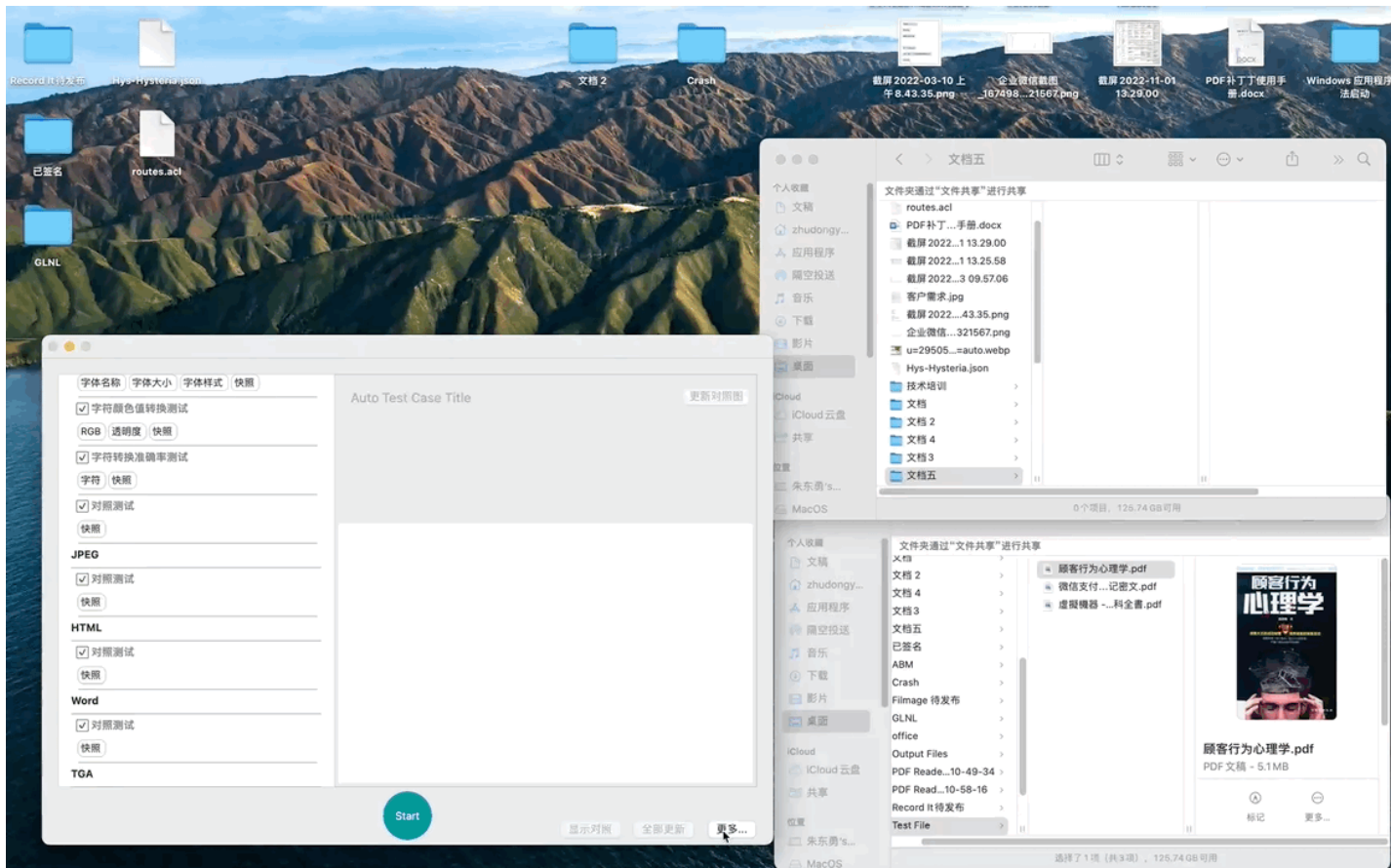
其中根文件目录中的子文件，可以手动创建，App 也能根据实际情况自动创建（从 App 导入文件时）

【自动创建】

通过 App 给不同转档类型导入文件（支持文件复制或 超链接导入），自动生成“Origin Files”目录；在转档过程中，再自动生成“Result Files”和“Check Files”目录结构，如下演示操作

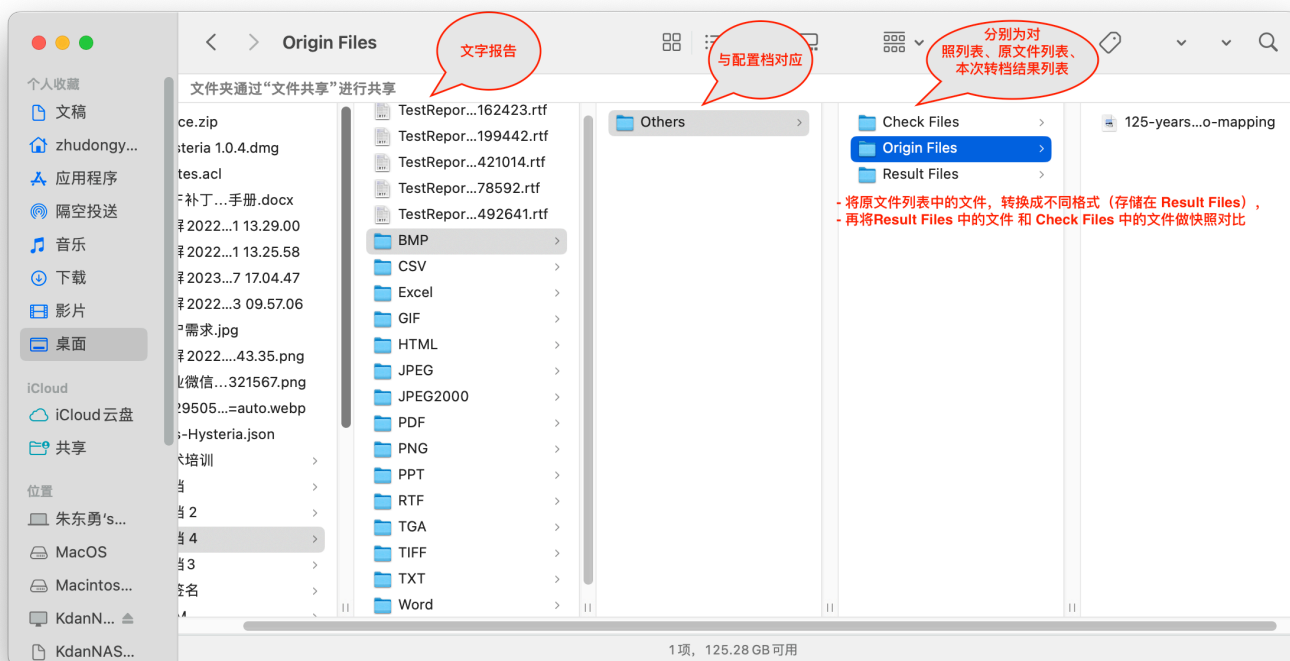
Note:

- PDF文档格式对照测试，还需要导入 Result Files
- RTF/TXT 本地转档对比测试，还需要导入 Check Files
- RTF/TXT 跨平台对比测试，还需要导入 Result Files 和 Check Files

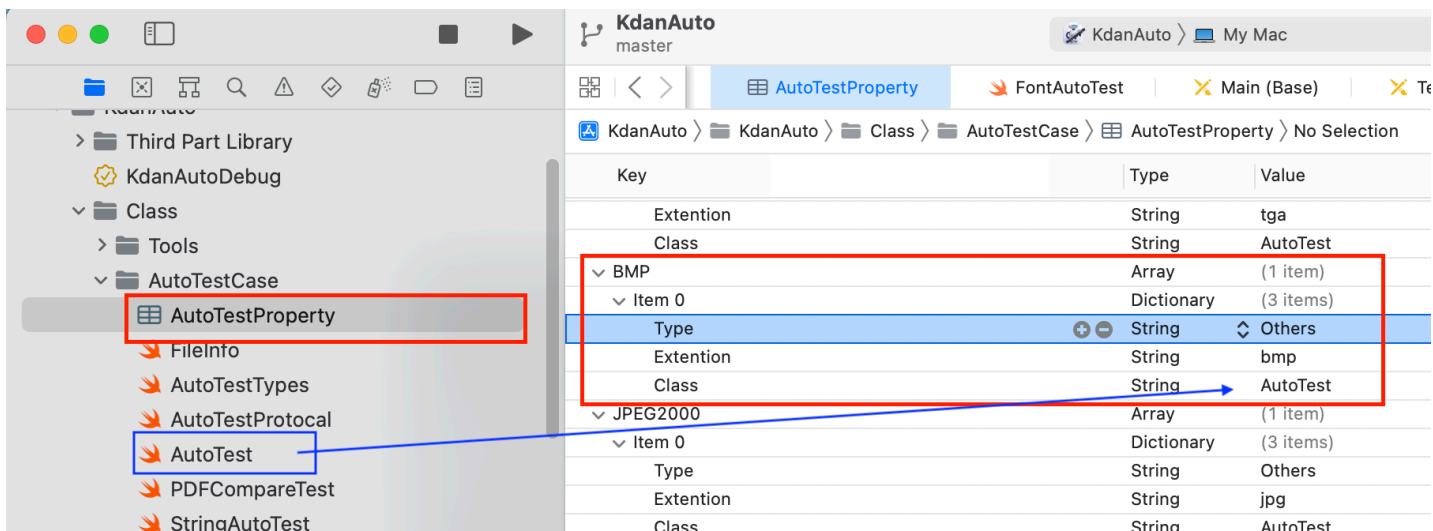


【手动创建】（已采用配置自动化，已弃用，采用自动生成目录方案）

在根目录内，手动创建如下图所示目录结构（每种文件，对照测试）目录结构类似。



上述列表中，相关目录中的转档类型，留意与 App 配置目录保持一致

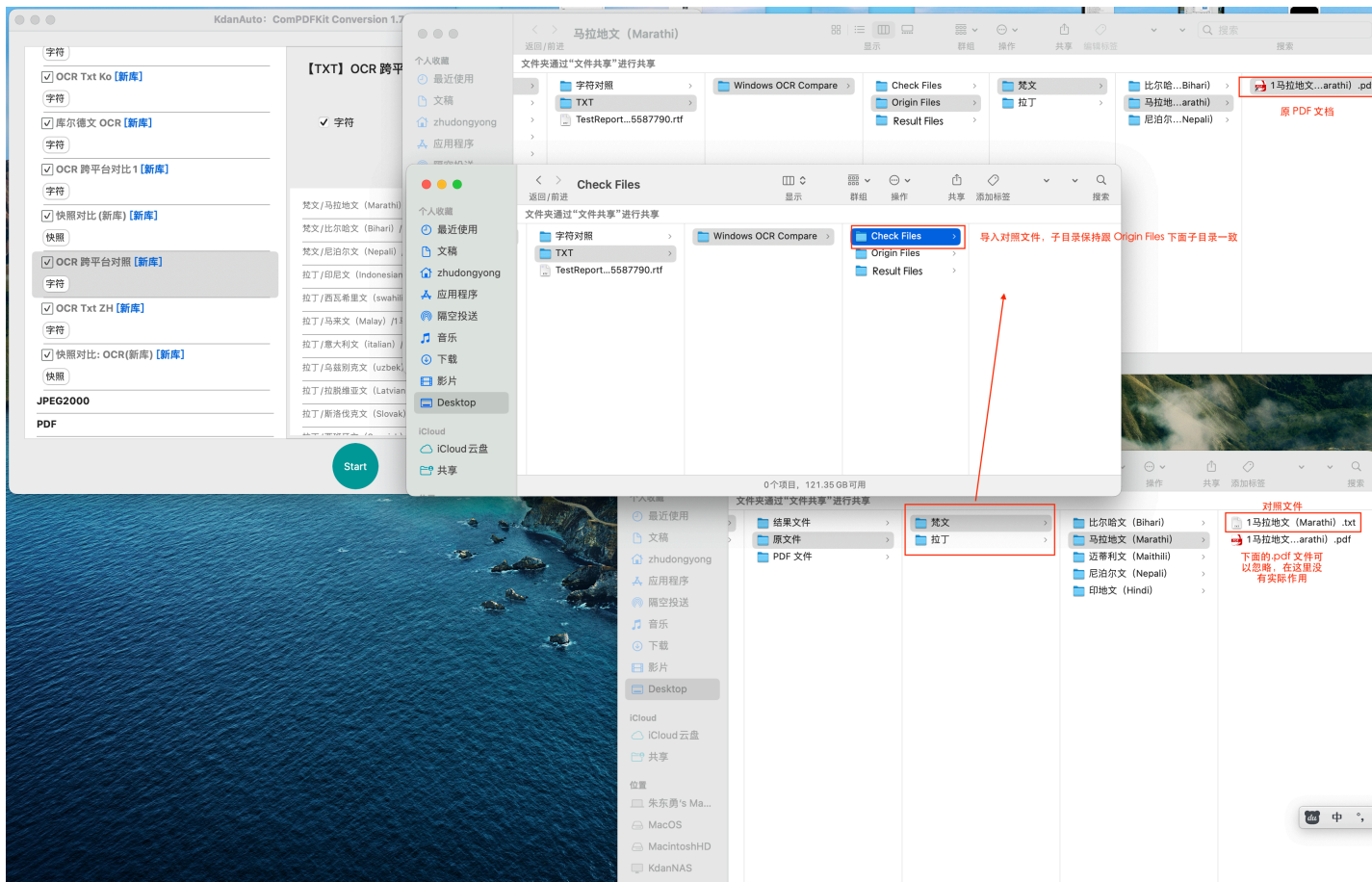


接着再可以考虑手动或自动往各类型的“Origin Files”测试(原)文档目录中拷贝 (/导入) 测试文档

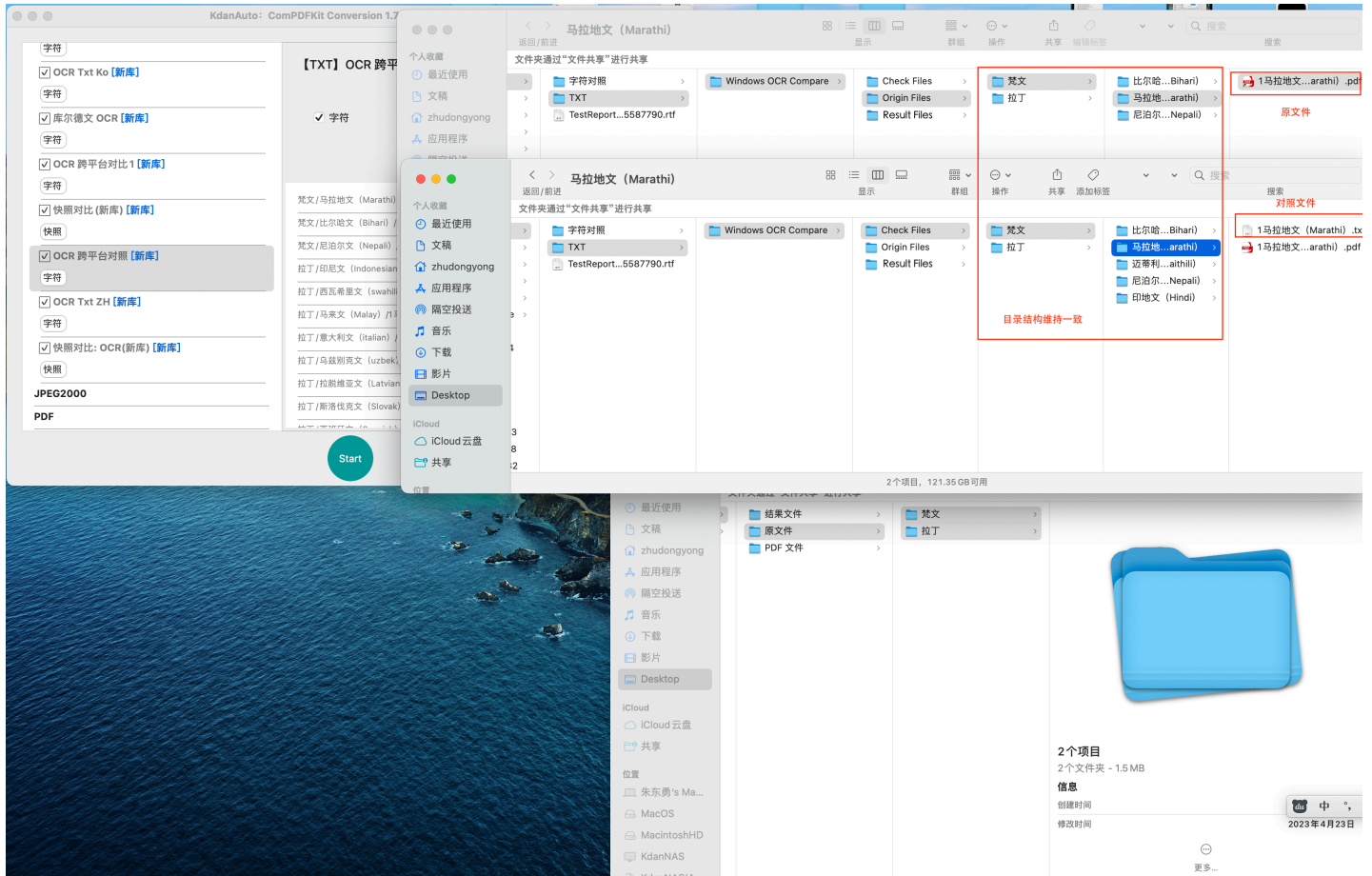
【导入对比文件】

对于PDF、RTF（字符对比）、TXT（字符对比）功能，需要手动从外部导入对比（参照）文档，以 TXT 格式跨平台 OCR 字符对比为例：导入“Result Files”和“Check Files”

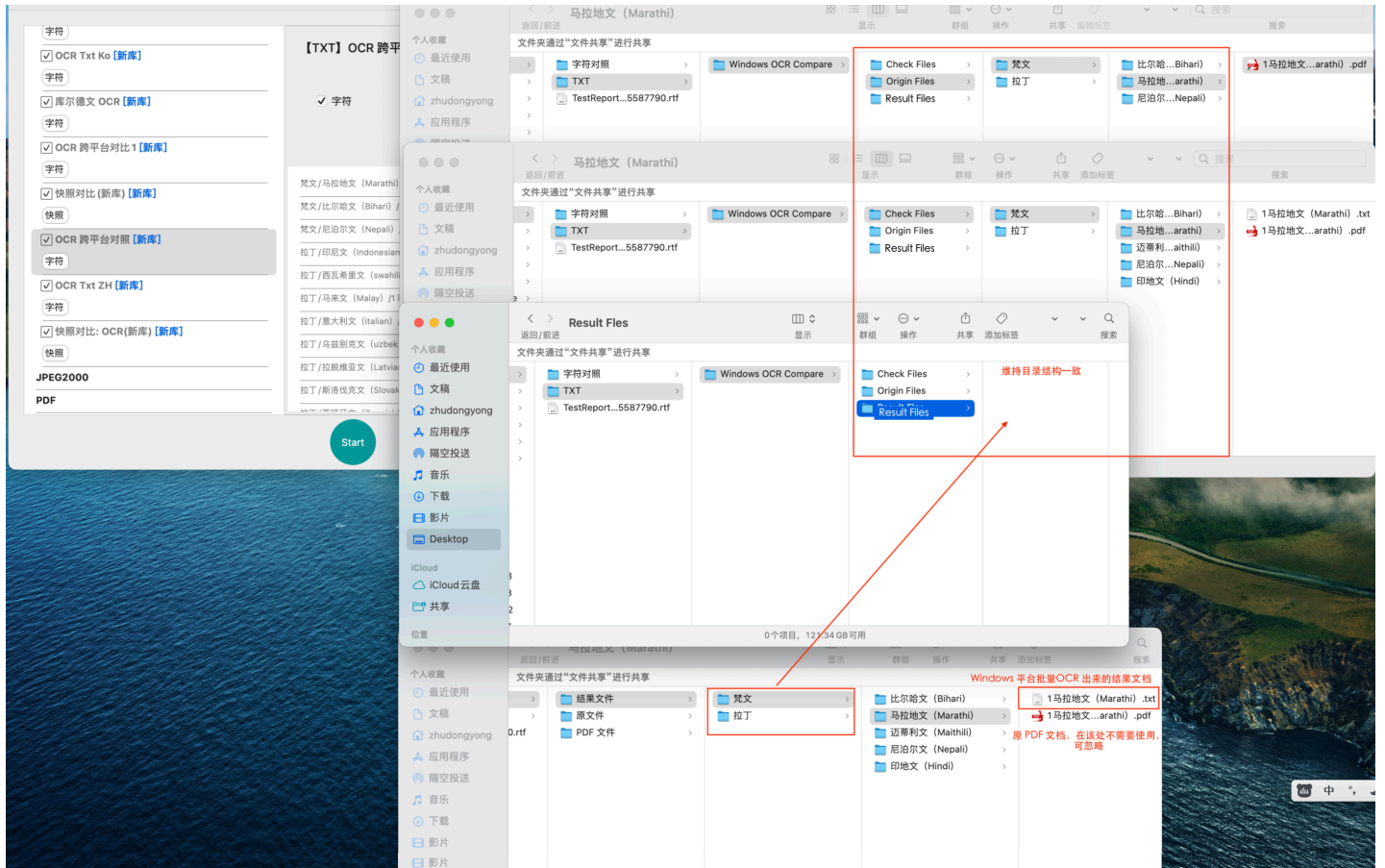
导入命令行、其他平台转换/OCR 结果文件，不是快照对照测试（快照是 KdanAuto 自动生成，并保存备用），就需要导入



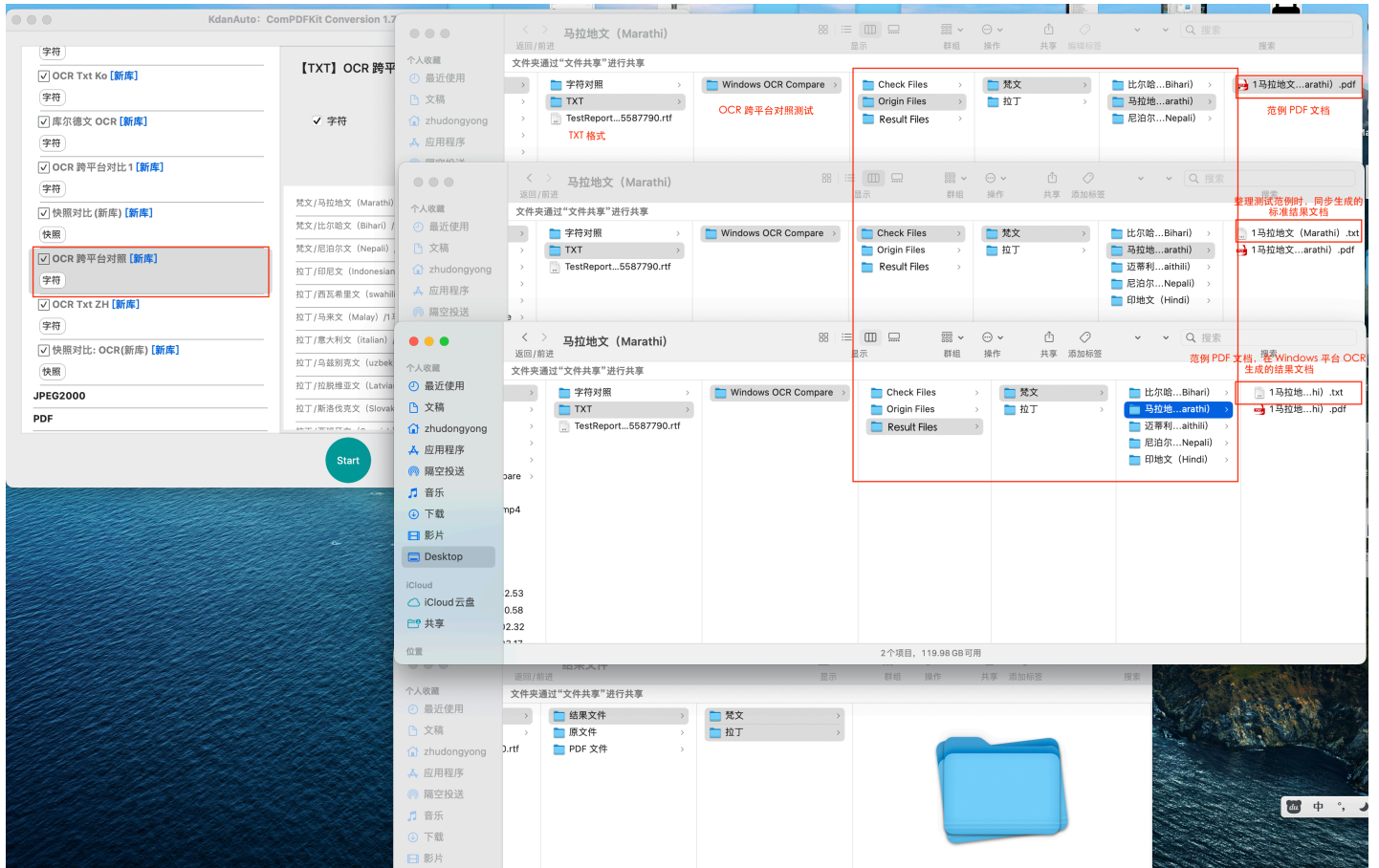
导入后效果，留意文件夹结构跟 Origin Files 目录文件夹结构保持一致



导入对照文件，不是使用 KdanAuto 进行转换/OCR得到的文件 进行 对照/对比，就需要导入

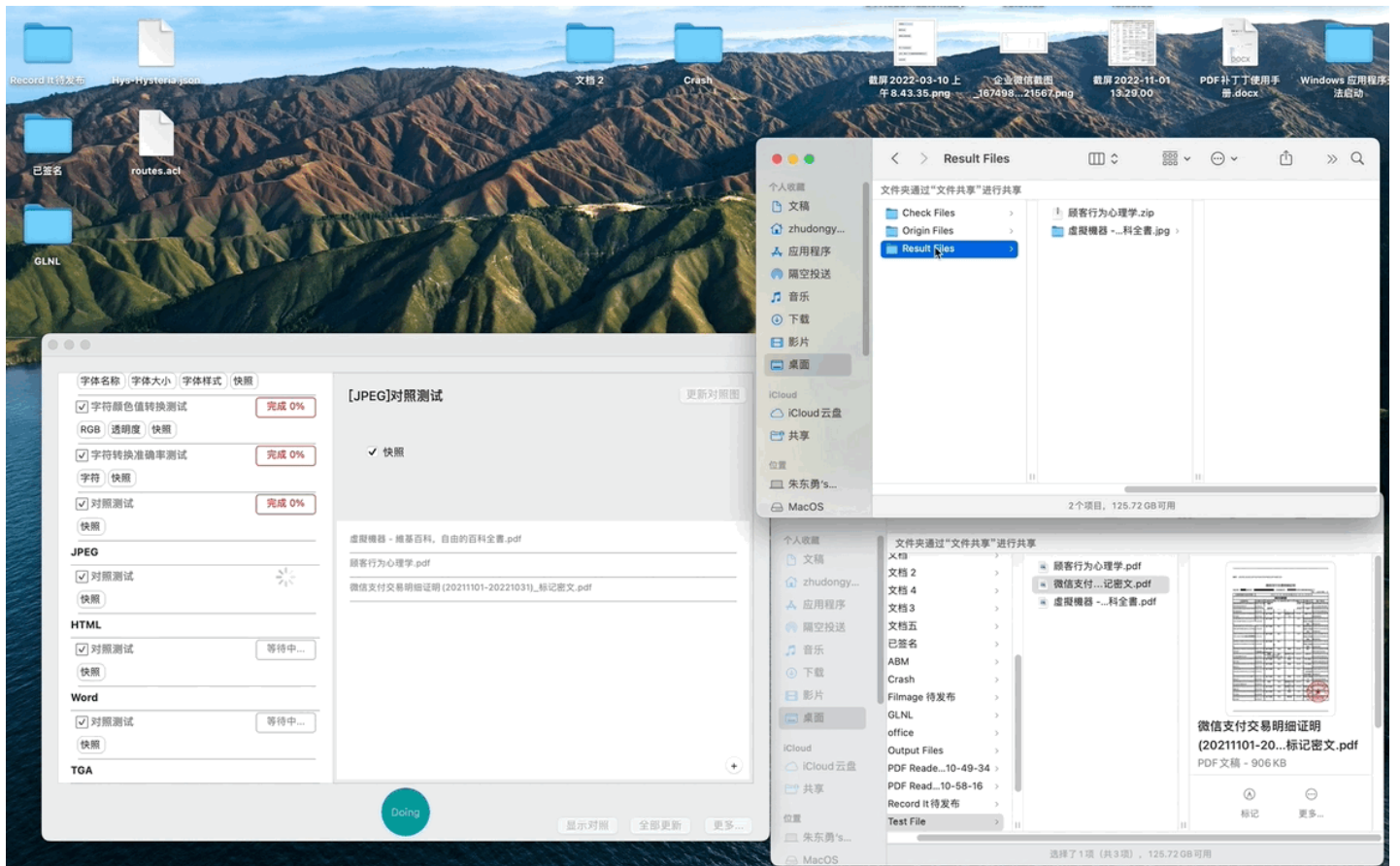


导入后效果，留意文件夹结构跟 Origin Files 目录文件夹结构保持一致



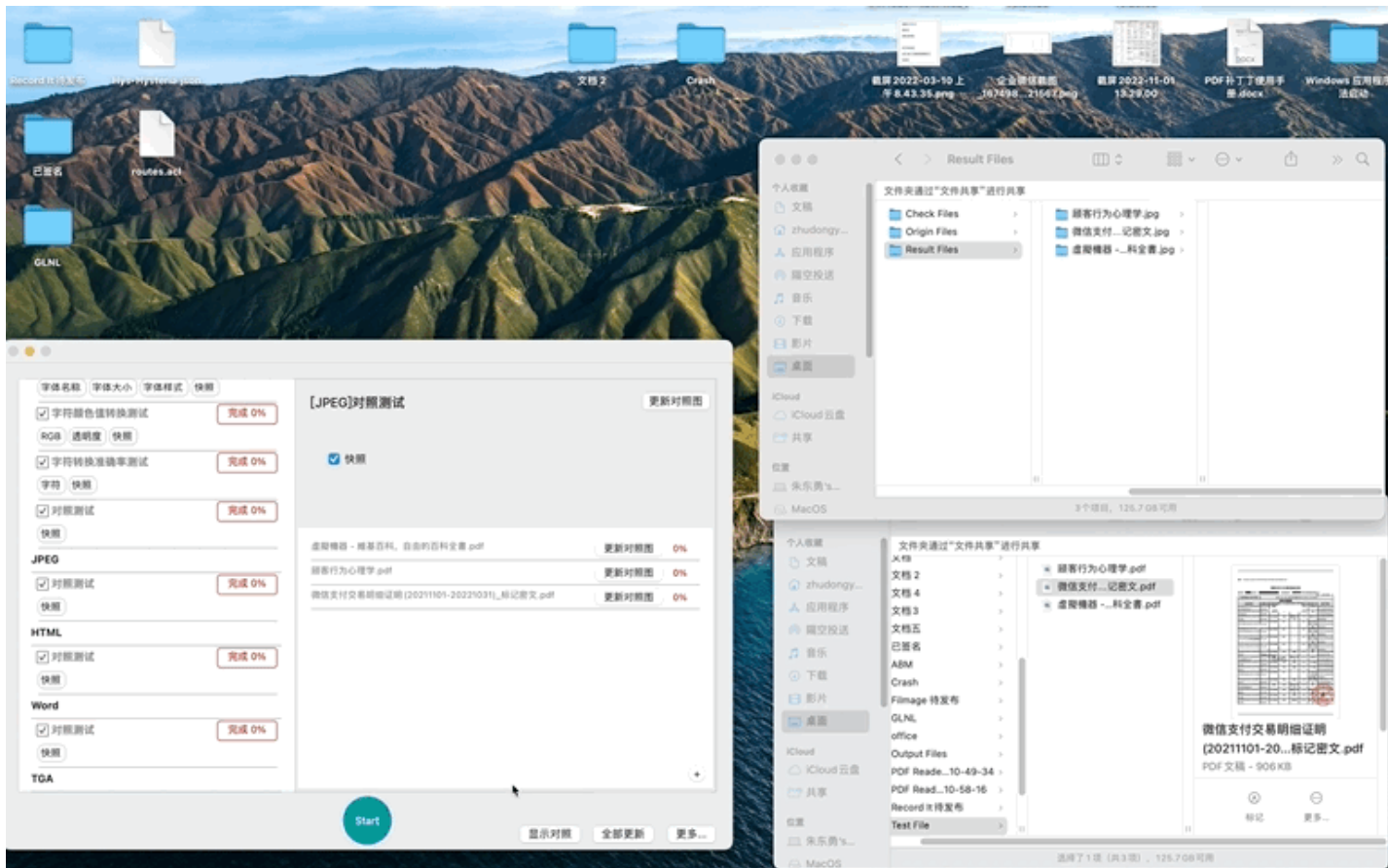
4、执行对照测试 & 更新对照文档

第一次测试，由于没有对照文件，所以测试对照结果均会为0%（模拟SDK第一轮发版），这个时候是需要QA人力对所有文档做确认，看有哪些功能是有完善的，哪些是有欠缺的。

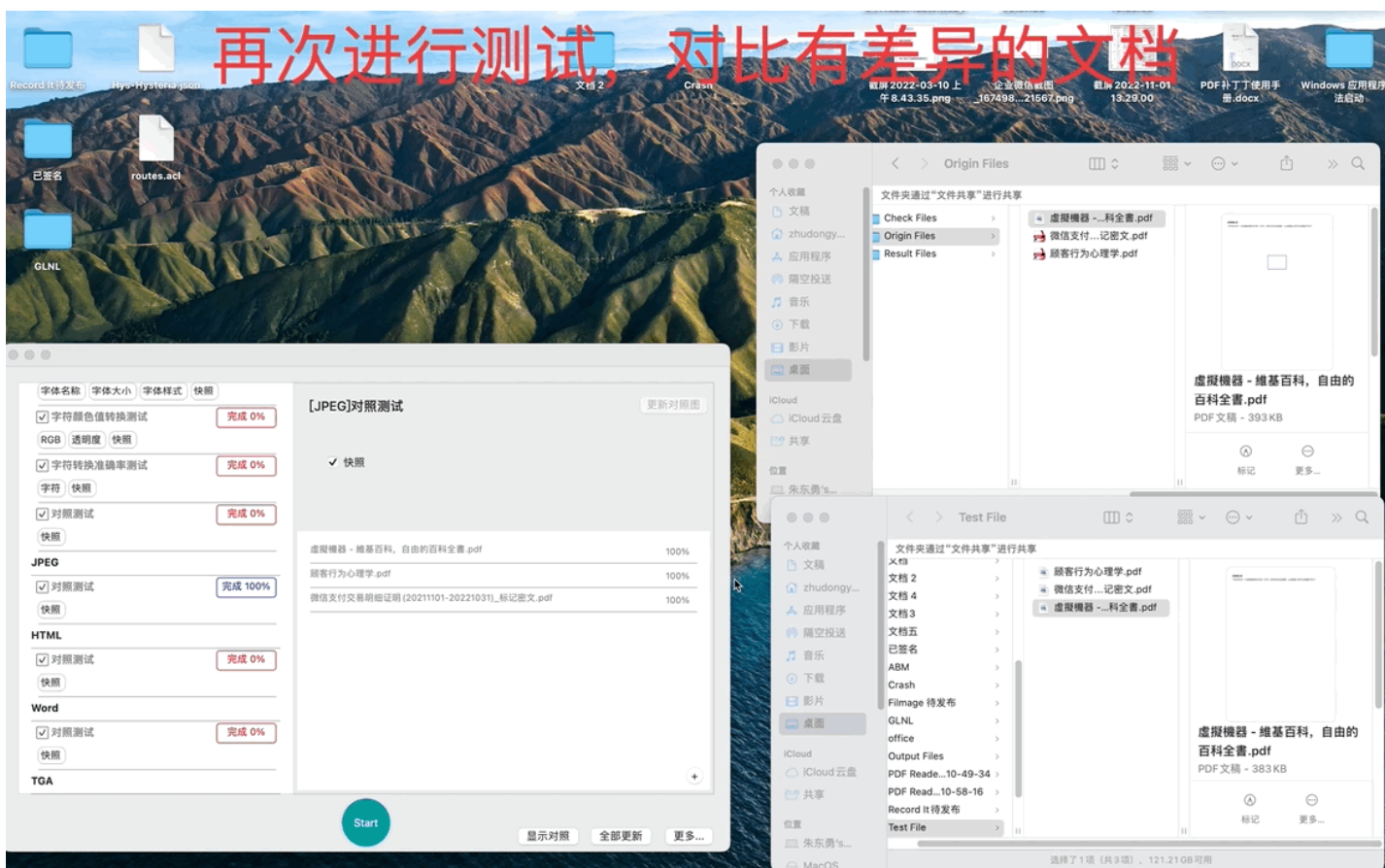


第一次测试的结果，可以存为对照结果（例如，我们可以拿线上已经发布上线的版本转档快照用作对照结果）。

在后续每轮次测试的时候，（之前已经有生成对照结果文档的测试档，）就会对应地对比出结果，如下演示



接着用将 PDF 库替换成新版本（例如从1.5.0 -> 1.6.0）（演示视频中的效果，是用 Pro Mac 将原文档进行适当的编辑，来模拟得出不同的结果进行演示）



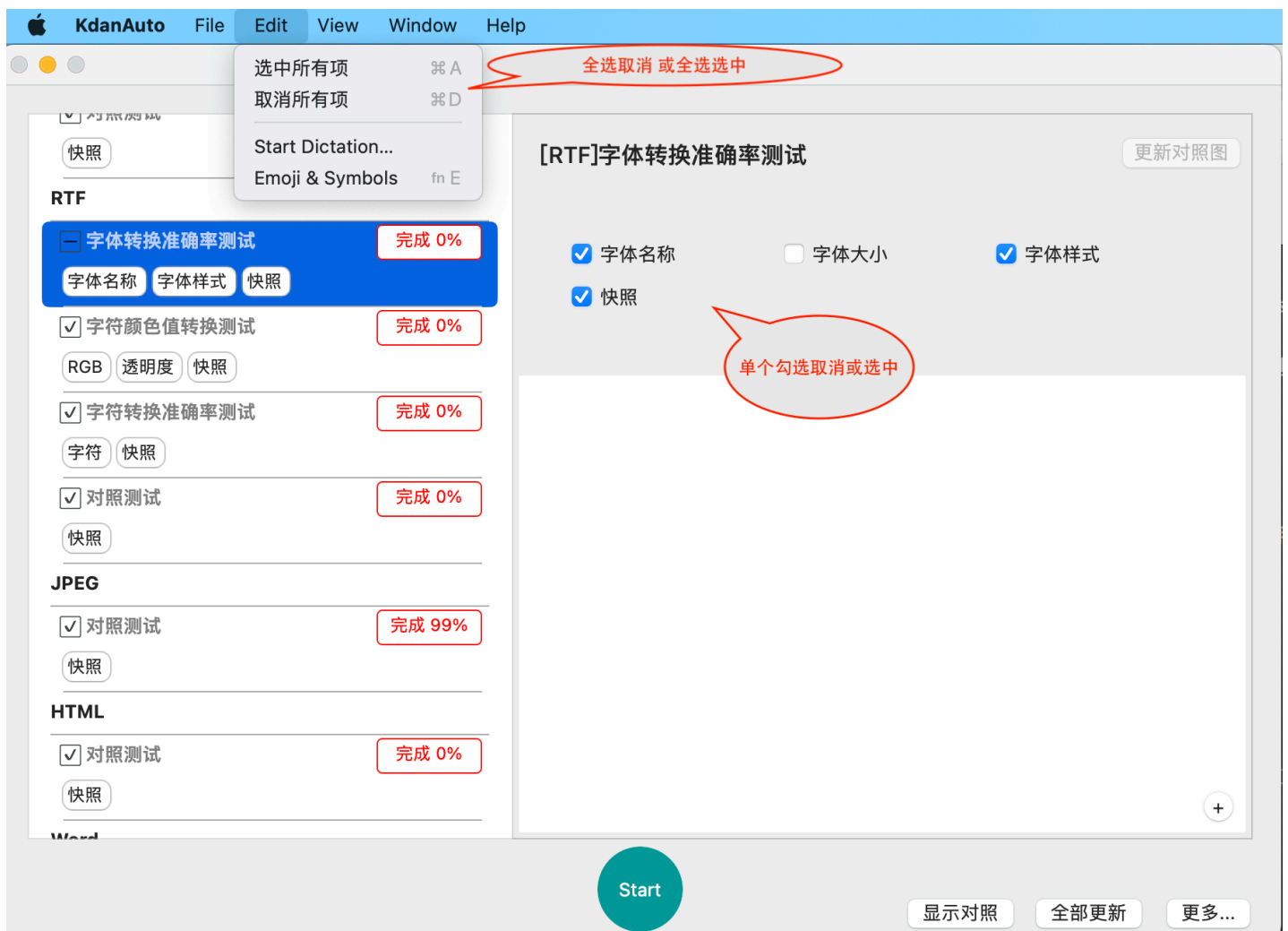
Note:

- 第一次测试演示，没有对照快照，所以对照结果相似度为0%
- 第二次测试演示，相同的3份文档，相同的库，对照会是完全一致的；
- 第三次测试演示，对比出来有差异的项，是用 PDF Reader Pro，将原文档进行一定的修改，模拟不同 SDK 版本，转档出来的内容有差异时，会是怎样的对照效果

5、勾选测试选项

常规格式，目前暂时只支持预览对照，后续再可以逐步添加对文档内容做更精准的对比测试，例如可以直接读取 Word 文档，看 PDF 转 word 文档的内容准确度。

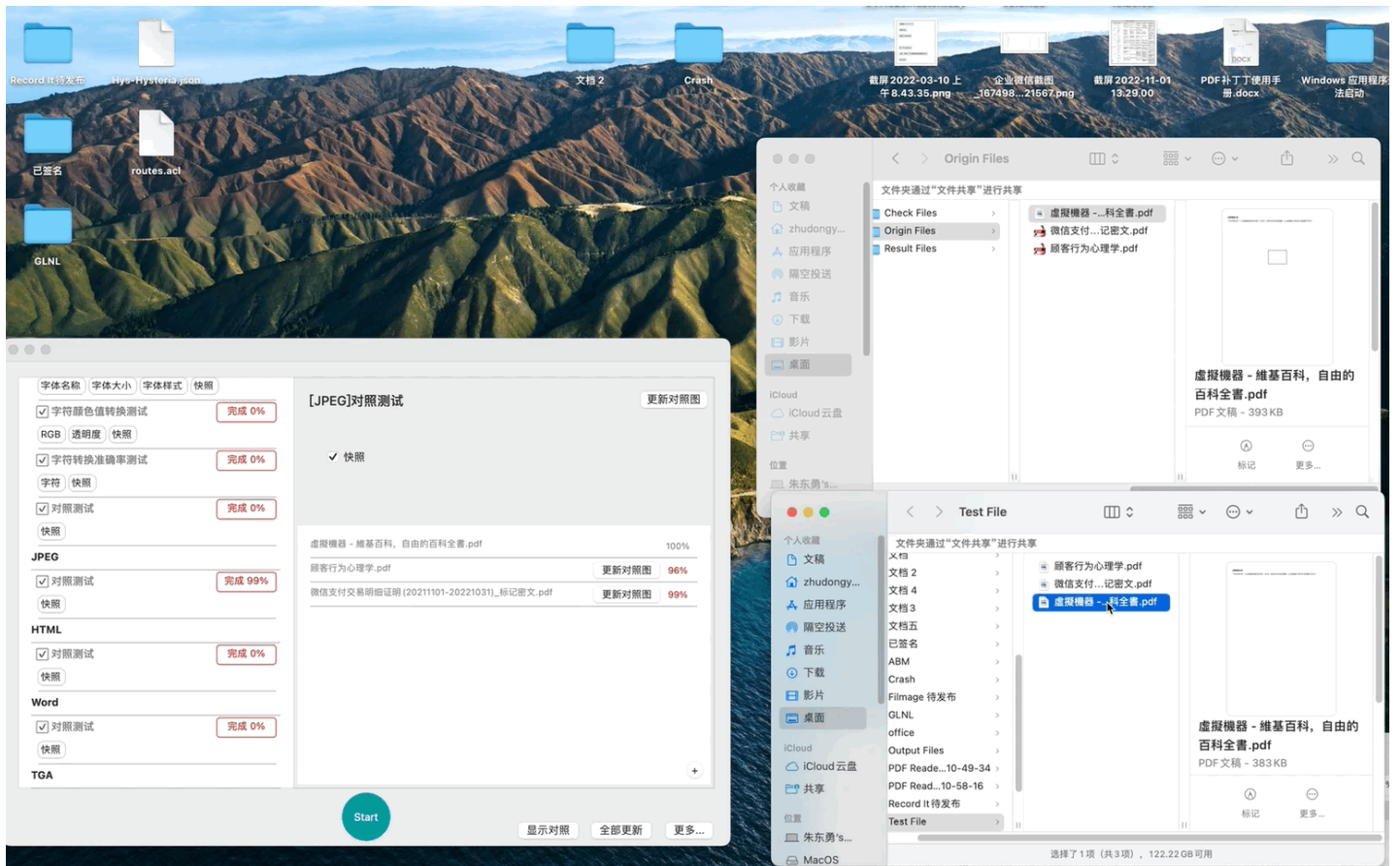
目前可以支持从 每种类型选项中，按需逐个【勾选】或【取消】，也可以从菜单中，点击【全选】和【全部取消】



6、导出测试报告

文字报告，在原本选择的测试目录（文档五）下，会生成文字报告；

对照图报告导出如下

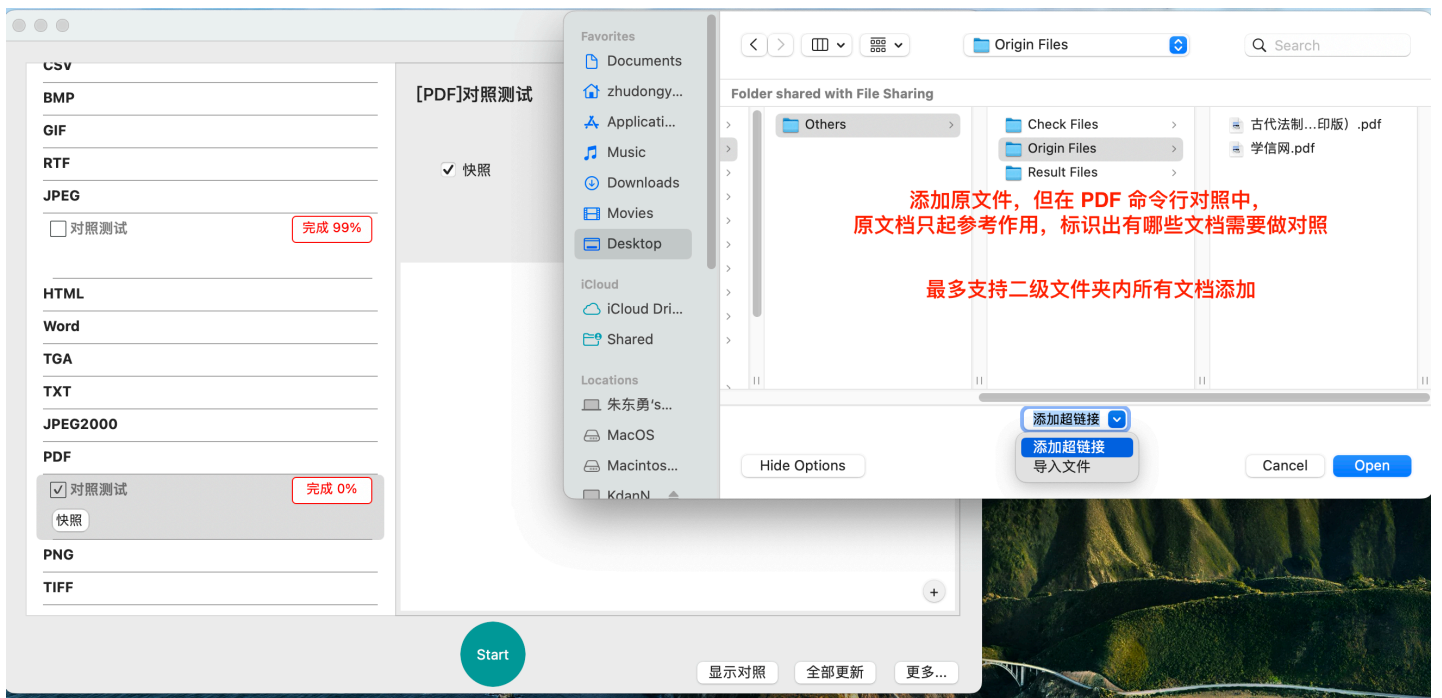


Note:

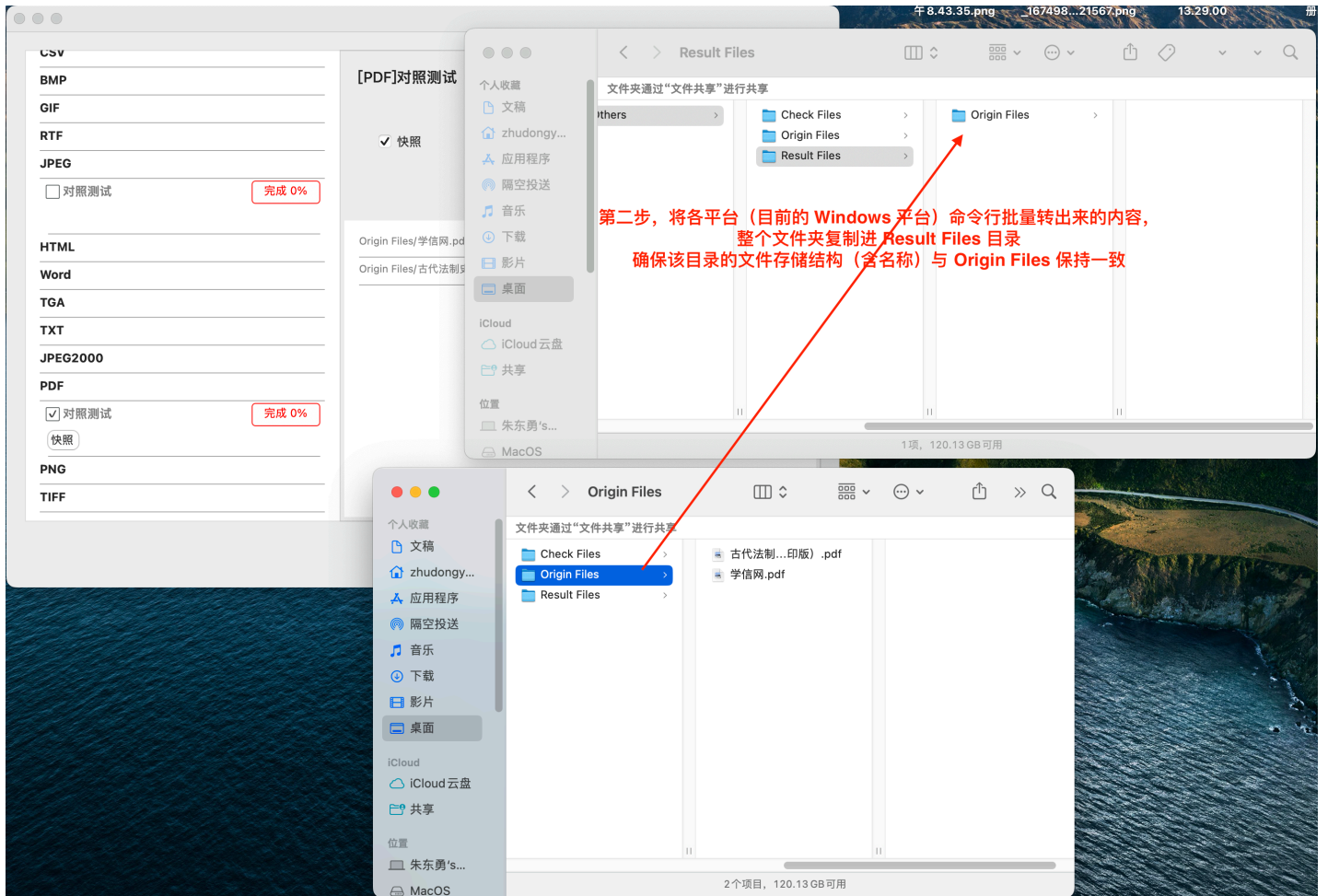
- 对照报告导出, 可以只指定导出有差异的项;
- 对照报告导出, 可以只导出某个文件, 可以只导出某种格式, 也可以全部导出

7、命令行转档对照测试

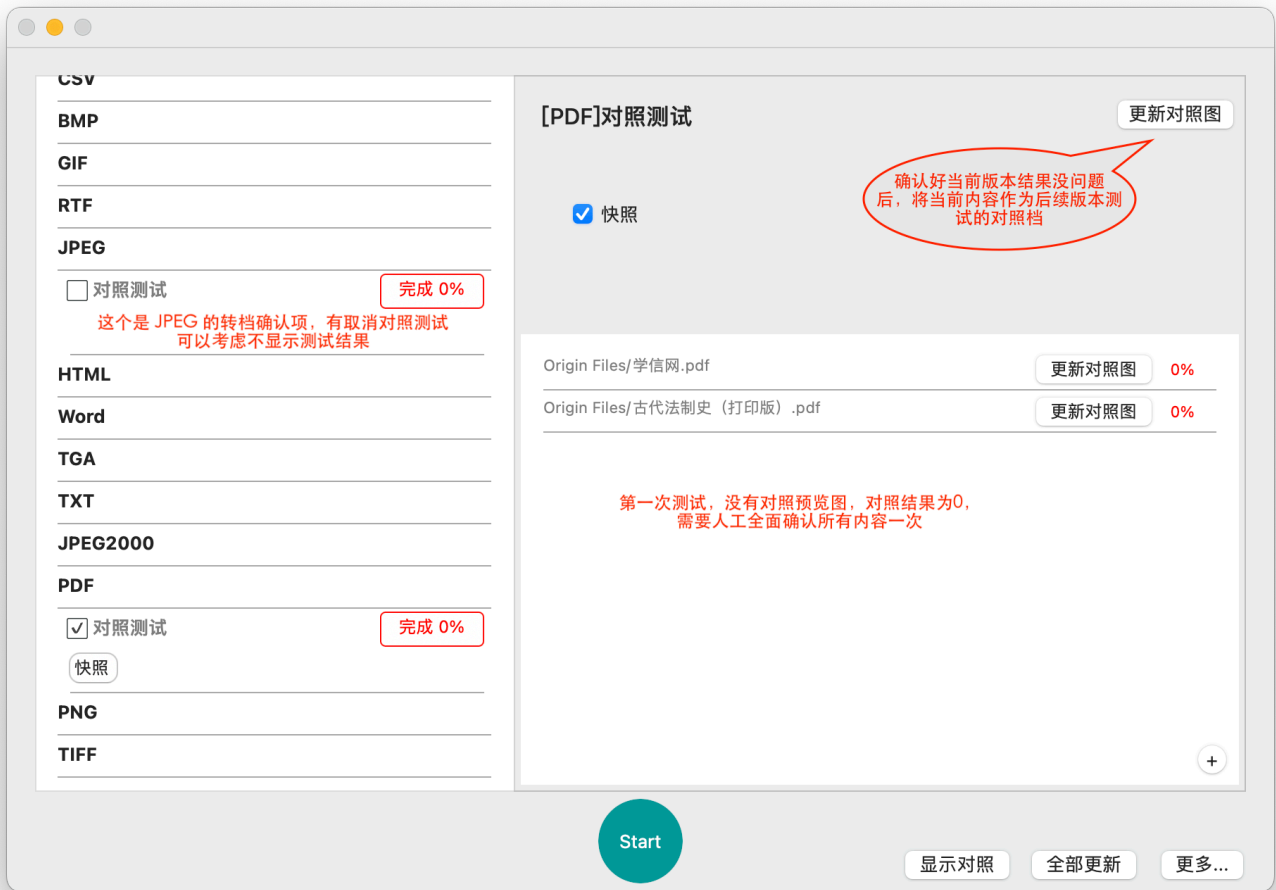
- S1、左侧列表中, 选择 PDF 子选项, 在右侧内容面板中添加文件, (留意最好添加软链接, 原文档在命令行工具中, 只是做为识别文件索引)



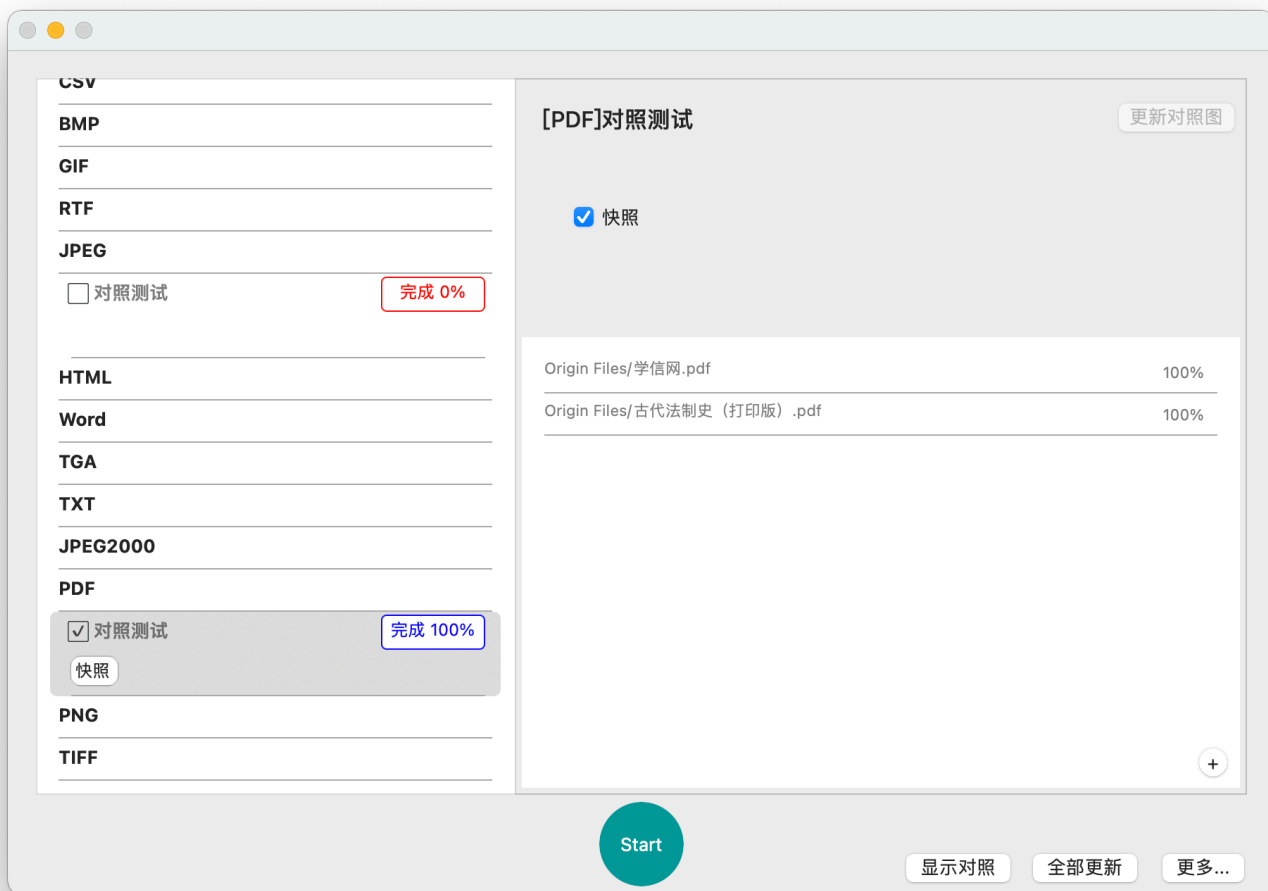
- S2、将命令行工具编辑生成的文件，复制到“Result Files”目录中，留意保持与 Origin Files 中文件结构一致（包括名称和后缀）



- S3、执行对照测试，第一次测试，对照结果会全为0%；



- S4、查看对照结果，及导出测试报告（注：将内容更新作为对照后，第二次测试，会得到实际的对照结果，有版本更新后，将能得到相同文档，不同命令行库版本转换出来的内容差异；差异效果同转档对照）



横向对照效果，如果只是改变了文字，比较难发现有差异

CSV

BMP

GIF

RTF

JPEG

对照测试 完成 0%

HTML

Word

TGA

TXT

JPEG2000

PDF

对照测试 完成 98%

快照

PNG

TIFF

X [PDF:对照测试]Origin Files/学信网.pdf/Page1.jpg 过滤无差异项 Export 相似度:95.01% 横向 覆盖

教育部学籍在线验证报告

更新日期: 2020年8月30日

姓名	蒋和光			
性别	男	证件号码	310116200004305414	
民族	汉族	出生日期	2000年04月30日	
院校	上海政法学院		层次	本科
院系			班级	经济法18051班
专业	法学		学号	1802051038
形式	普通全日制	入学时间	2018年09月01日	学制 4年
类型	普通高等教育	学籍状态	注册学籍(预计毕业日期: 2022年07月01日)	

在线验证

ATHKXZULTLJKUV

在线验证码

1. 扫描“教育部学籍在线验证”小程序

2. 使用小程序扫码验证

注意事项:

- 《学籍在线验证报告》是教育部学籍电子注册备案的查询结果。
- 报告内容验证办法: ①点击报告(电子版)中的在线验证码, 可在在线验证; ②登录中国高等教育学生信息网“在线验证系统”, 输入在线验证码进行验证; ③使用“学信网报告在线验证”的微信小程序, 进行扫码验证。为防止出现假冒报告, 请使用该小程序扫描验证, 不要使用其他第三方扫描程序。
- 报告内容如有篡改, 将以最新在线验证的内容为准。
- 本报告信息权属人同意, 不得将报告用于违背权属人意愿之用途。
- 报告在线验证有效期由报告权属人设置(16个月), 其在报告验证到期前可再次延长验证有效期。

教育部学籍在线验证报告

更新日期: 2020年8月30日

姓名	蒋和光			
性别	男	证件号码	310116200004305414	
民族	汉族	出生日期	2000年04月30日	
院校	上海政法学院		层次	本科
院系	计算科学		班级	经济法2006班
专业	计算机科学与技术		学号	1802051038
形式	普通全日制	入学时间	2018年09月01日	学制 4年
类型	普通高等教育	学籍状态	注册学籍(预计毕业日期: 2022年07月01日)	

在线验证

ATHKXZULTLJKUV

在线验证码

1. 扫描“教育部学籍在线验证”小程序

2. 使用小程序扫码验证

注意事项:

- 《学籍在线验证报告》是教育部学籍电子注册备案的查询结果。
- 报告内容验证办法: ①点击报告(电子版)中的在线验证码, 可在在线验证; ②登录中国高等教育学生信息网“在线验证系统”, 输入在线验证码进行验证; ③使用“学信网报告在线验证”的微信小程序, 进行扫码验证。为防止出现假冒报告, 请使用该小程序扫描验证, 不要使用其他第三方扫描程序。
- 报告内容如有篡改, 将以最新在线验证的内容为准。
- 本报告信息权属人同意, 不得将报告用于违背权属人意愿之用途。
- 报告在线验证有效期由报告权属人设置(16个月), 其在报告验证到期前可再次延长验证有效期。

[PDF:对照测试]Origin Files/古代法制史 (打印版).pdf/Page1.jpg 相似度:100.00% 横向 覆盖

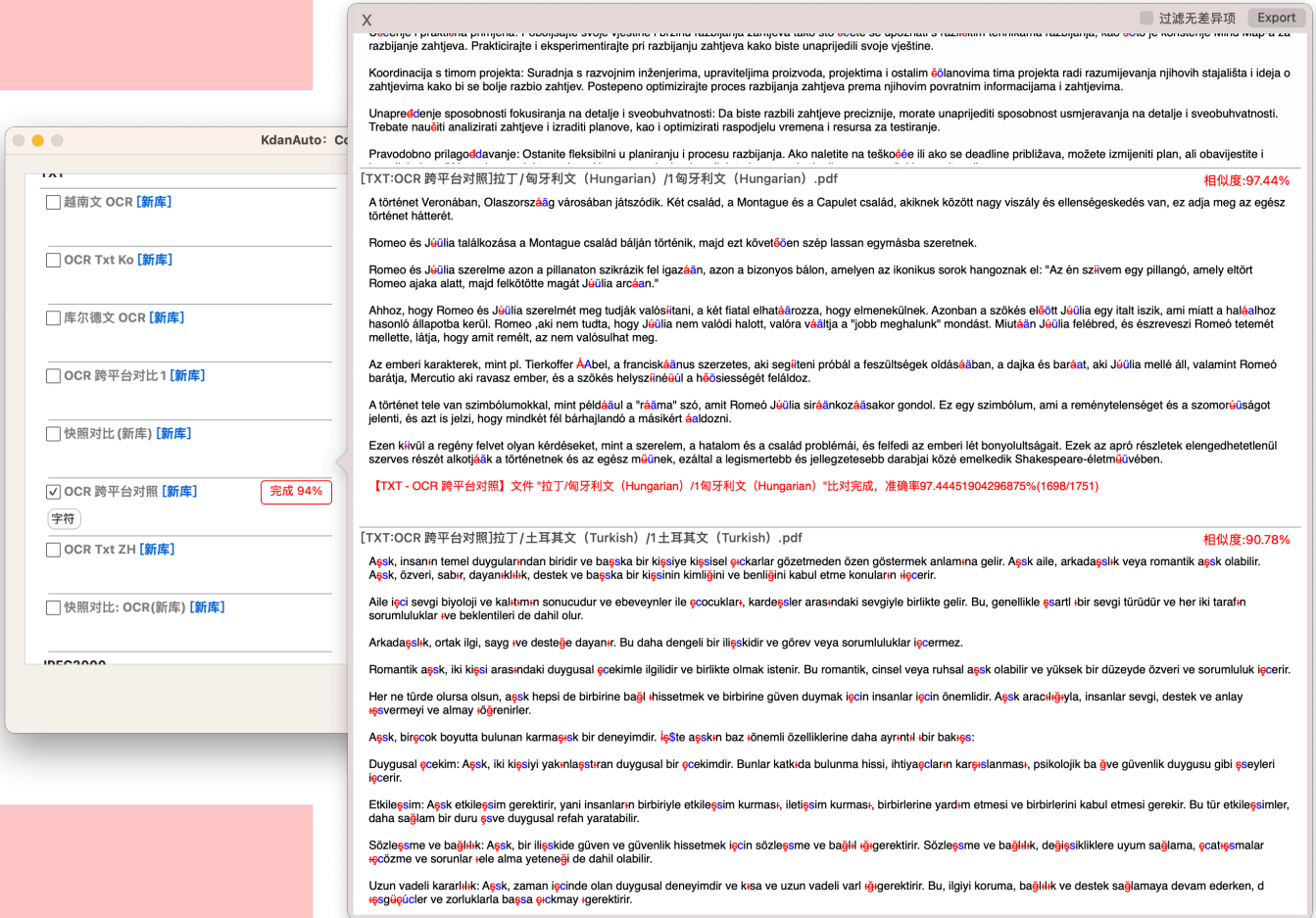
覆盖对比能比较好发现, 中间有多了什么, 少了什么, 以及文字有位移的情况



8、字符串对比测试

通过第4点导入的原(PDF)文件，以及导入 准备测试资料时准备的（标准）对照文件（，以及导入Windows、Saas、命令行工具等 转换/OCR 得到的结果文档），接着就可以进行字符串对比测试

TXT、RTF 格式字符对比测试报告效果如下



9、类型及参数配置（以下参数不再需要手动配置，通过高级设置界面进行设置）

配置表中，目前有预置 **PDF、RTF、Word、PPT、Excel、CSV、HTML、TXT、JPEG、PNG、GIF、TIFF、TGA、BMP、JPEG2000**格式类型，其中**PDF** 暂时考虑用作命令行注释编辑工具，批量执行文档结果对照测试项；而**RTF、Word、PPT、Excel、CSV、HTML、TXT、JPEG、PNG、GIF、TIFF、TGA、BMP、JPEG2000**为自动化转档测试对照测试项；

- PDF 用于 命令行工具，注释、编辑等操作，批量操作文档自动化校对选项
- Word 用于 PDF to Word，转档自动化测试及结果自动化对照测试项
- Excel 用于 PDF to Excel，转档自动化测试及结果自动化对照测试项

Note:

- 只能对照每份文档的左上角部分；
- 交互式表单无法识别（不确定是不是所有）

- PPT 用于 PDF to PPT，转档自动化测试及结果自动化对照测试项，

Note: 只能对照每份文档的第一页

- CSV 用于 *PDF to CSV*, 转档自动化测试及结果自动化对照测试项

Note:

- 只能对照每份文档的左上角部分;

- HTML 用于 *PDF to HTML*, 转档自动化测试及结果自动化对照测试项

Note:

- 新库转档出来是 zip

- RTF 用于 *PDF to RTF*, 转档自动化测试及结果自动化对照测试项 还可用于 *PDF to RTF*, 本地 (或其他平台) 转档/OCR 结果字符对比测试

Note:

- 新库 RTF 格式, 不支持预览, 需要用 WPS 等软件才能打开看效果;

- TXT 用于 *PDF to TXT*, 转档自动化测试及结果自动化对照测试项 还可用于 *PDF to TXT*, 本地 (或其他平台) 转档/OCR 结果字符对比测试

- PNG 用于 *PDF 转 PNG *图片格式, 转档自动化测试及结果自动化对照测试项

Note:

- 新库转档出来是 zip

- JPEG 用于 *PDF 转 JPEG *图片格式, 转档自动化测试及结果自动化对照测试项

Note:

- 新库转档出来是 zip

- GIF 用于 *PDF 转 GIF *图片格式, 转档自动化测试及结果自动化对照测试项

- TIFF 用于 *PDF 转 TIFF *图片格式, 转档自动化测试及结果自动化对照测试项

- TGA 用于 *PDF 转 TGA *图片格式, 转档自动化测试及结果自动化对照测试项

- BMP 用于 *PDF 转 BMP *图片格式, 转档自动化测试及结果自动化对照测试项

- JPEG2000 用于 *PDF 转 JPEG2000 *图片格式, 转档自动化测试及结果自动化对照测试项

支持的参数有 **Type (必选)**、**Name (必选)**、**Extention (必选)**、**Class (必选)**、**Params (可选)**, 对应参数介绍如下:

【Type】 Type为对应格式选项下, 识别类型, 同时对应于当前格式下, 测试类型目录, 例如 【JPEG】 格式

下面的【DPI-300】测试类型

【Name】 *Name*为当前 *Type*, 在 App 中显示的别名, 例如“对照识别”“DPI-300对照识别”

【Extention】 *Extention*为当前 *Type*文件处理后, 所得到的文件的后缀

【Class】 *Class*为当前格式、当前类型 (*Type*) , 用来做自动化处理的 *TestCase* 类, 与测试代码进行关联, 其中通用对照测试使用 *AutoTest*, PDF 命令行工具使用 *PDFCompareTest*

【Params】 *Params*为各自动化测试文件格式中, 可以再从外部 (不需要重新编译App) 来控制的参数配置, 其中各种格式对应的控制参数如下表:

```
- password: PDF 文档密码, String 类型
- useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新库, NSNumber类型 (0/1) 或 String 类型
- paneOptions: HTML 格式, 导出 HTML 页面及导航样式
  - CPDFConvertHtmlSinglePage = 0, // Single page with no na
    - CPDFConvertHtmlSinglePageNavigationByBookmarks = 1, // Navigate single pa
    - CPDFConvertHtmlMultiplePages = 2, // Multiple pages wit
    - CPDFConvertHtmlMultiplePagesSplitByBookmarks = 3 // Split multiple pag
- KMPDFConvertOptionsKeyImageDPI: 老转档库 DPI 参数, 常规设置有50、72、96、144、300、600,
- CPDFConvertOptionsKeyAllInOneSheet: 老库 PDF 转CSV, 多页合并为一个文档, Bool
```

不同格式对应控制参数

- PDF: PDF 文档快照对照
 - password, 所导入PDF 文档的密码, String 类型
- RTF: 转档及结果对照
 - password, 所导入文档的密码
 - useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新转档库, NSNumber类型 (0/1) 或 String 类型 (TRUE, FALSE), 默认缺失时使用新转档库
- Word: 转档及结果对照
 - password, 所导入文档的密码
 - useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新转档库, NSNumber类型 (0/1) 或 String 类型 (TRUE, FALSE), 默认缺失时使用新转档库
- PPT: 转档及结果对照
 - password, 所导入文档的密码
 - useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新转档库, NSNumber类型 (0/1) 或 String 类型 (TRUE, FALSE), 默认缺失时使用新转档库
- Excel: 转档及结果对照

- password, 所导入文档的密码
 - useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新转档库, NSNumber类型 (0/1) 或 String 类型 (TRUE, FALSE) , 默认缺失时使用新转档库
- CSV: 转档及结果对照
 - password, 所导入文档的密码
 - useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新转档库, NSNumber类型 (0/1) 或 String 类型 (TRUE, FALSE) , 默认缺失时使用新转档库
 - CPDFConvertOptionsKeyAllInOneSheet: 老库 PDF 转CSV, 多页合并为一个文档, Bool
- HTML: 转档及结果对照
 - password, 所导入文档的密码
 - useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新转档库, NSNumber类型 (0/1) 或 String 类型 (TRUE, FALSE) , 默认缺失时使用新转档库
 - paneOptions: HTML 格式, 导出 HTML 页面及导航样式
 - CPDFConvertHtmlSinglePage = 0, // Single page with no navigation.
 - CPDFConvertHtmlSinglePageNavigationByBookmarks = 1, // Navigate single page using bookmarks from PDF.
 - CPDFConvertHtmlMultiplePages = 2, // Multiple pages with no split.
 - CPDFConvertHtmlMultiplePagesSplitByBookmarks = 3 // Split multiple pages using bookmarks from PDF.
- TXT: 转档及结果对照
 - password, 所导入文档的密码
 - useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新转档库, NSNumber类型 (0/1) 或 String 类型 (TRUE, FALSE) , 默认缺失时使用新转档库
- JPEG: 转档及结果对照
 - password, 所导入文档的密码
 - useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新转档库, NSNumber类型 (0/1) 或 String 类型 (TRUE, FALSE) , 默认缺失时使用新转档库
 - KMPDFConvertOptionsKeyImageDPI: 老转档库 DPI 参数, 常规设置有50、72、96、144、300、600, String类型, 默认72
- PNG: 转档及结果对照
 - password, 所导入文档的密码
 - useOldLib: 转档类型中, 确认是用旧版本转档库 或是 用新转档库, NSNumber类型 (0/1) 或 String 类型 (TRUE, FALSE) , 默认缺失时使用新转档库
 - KMPDFConvertOptionsKeyImageDPI: 老转档库 DPI 参数, 常规设置有50、72、96、144、300、600, String类型, 默认72

- GIF: 转档及结果对照
 - password, 所导入文档的密码
 - KMPDFConvertOptionsKeyImageDPI: 老转档库 DPI 参数, 常规设置有50、72、96、144、300、600, String类型, 默认72
- TIFF: 转档及结果对照
 - password, 所导入文档的密码
 - KMPDFConvertOptionsKeyImageDPI: 老转档库 DPI 参数, 常规设置有50、72、96、144、300、600, String类型, 默认72
- TGA: 转档及结果对照
 - password, 所导入文档的密码
 - KMPDFConvertOptionsKeyImageDPI: 老转档库 DPI 参数, 常规设置有50、72、96、144、300、600, String类型, 默认72
- BMP: 转档及结果对照
 - password, 所导入文档的密码
 - KMPDFConvertOptionsKeyImageDPI: 老转档库 DPI 参数, 常规设置有50、72、96、144、300、600, String类型, 默认72
- JPEG2000: 转档及结果对照
 - password, 所导入文档的密码
 - KMPDFConvertOptionsKeyImageDPI: 老转档库 DPI 参数, 常规设置有50、72、96、144、300、600, String类型, 默认72

9、问题备注

- PPT、Word、HTML、TXT、RTF 只有第一页能展示
- 交互式表单, 新库解析看起来有问题
- 老库注释会丢失
- 古代法制史 (打印版) .pdf 等比较多文档乱码