



配置式档案数据治理系统 V2.0

数据架构员

操作说明书



上海珂绘数码影像科技有限公司

2024 年 5 月

目录

一、 数据架构.....	5
1、 查询浏览设置.....	5
1.1 类目架构设置.....	5
1.2 数据集元数据.....	5
2、 数据 ETL 设置.....	10
2.1 数据 ETL 规则.....	10
2.2 上传数据至表.....	12
2.3 全文数据处理.....	15
2.4 EXCEL 模板设置.....	16
3、 数据收集设置.....	21
3.1 数据著录架构.....	21
3.2 表字段元数据.....	22
3.3 审核显示字段.....	25
4、 四性检测设置.....	27
4.1 四性检测规则设置.....	27
二、 知识库.....	41
三、 工具集合.....	43
1、 格式转换.....	43
2、 PDF 工具.....	43
2.1 压缩 PDF 文件.....	44
2.2 合并和拆分 PDF.....	44
2.3 PDF 分割成 JPG.....	44
2.4 PDF 添加水印.....	45
2.5 PDF 添加条码.....	45
2.6 PDF 添加数字证书.....	46
2.7 PDF 文档加解密.....	46
2.8 查找和替换 PDF 文档中的文本.....	47
2.9 PDF 添加骑缝章.....	47
2.10 PDF 重叠页面.....	48
2.11 PDF 设置过期时间.....	48
2.12 PDF 添加动态图章.....	48
3、 检查工具.....	49
3.1 PDF 图纸检查工具.....	49
3.2 档案数据四性检测.....	50
4、 OCR 工具: Umi-ocrV2.1.1.....	52
4.1 主界面.....	52
4.2 批量文档(可生成双层 PDF).....	53
4.3 截图 OCR.....	54
4.4 批量 OCR.....	55
4.5 二维码.....	55
联系方式.....	57
软件反馈单.....	57

前言

随着数据化时代的到来，企业对于数据的需求越来越高。档案作为一种重要的信息资源，其管理方式的科学性和高效性对于企业的运营和发展具有重要意义。传统的档案管理方式存在着诸多问题，如人工操作繁琐、效率低下、易出错等。因此，开发一款符合档案数据治理体系的配置式档案治理系统显得尤为重要。该系统可以通过自动化的方式快速实现档案的三员分离、著录、导入导出、存储、全文检索、四性检测等功能，提高企业的工作效率和管理水平。

上海珂绘数码科技有限公司公司（下称“公司”）的《配置式档案数据治理系统》（以下简称“系统”）是一款功能强大且易于使用的档案管理软件，具有以下特点：

一、通用的数据架构平台

支持企业或组织内数据整合、模型定义、存储管理以及数据流动的框架，允许用户根据自己的需求定义档案的属性和字段。用户可以添加、修改或删除字段，以适应不同类型的档案管理，便于之后著录、导入和导出流程。

二、通用的档案工具平台



专为档案管理和文档保存而设计，提供了各类 **PDF** 文件的通用工具、数据检查；支持对不同格式的档案源数据文件的转换、及各类 **PDF** 工具，并针对 **PDF** 竣工图开发了图层、签章及属性的检查工具。对于数据表和成品 Access 数据库，可以使用四性检测功能。

三、通用的查询调阅平台

提供高效的模糊查询、字段精准查询、全文检索、数据追溯、四性检测报告查询等功能

帮助利用者快速准确找到目标档案，大幅提高模糊信息下的查档效率和精准度，通过档案信息系统可实现类百度式一体化检索，在线浏览利用，无需到档案库房，不受时间及空间的限制，简化流程。



四、通用的数据收集平台

支持多种数据源的连接和数据抽取、数据转换、数据加载、数据验证等功能，针对众多档案类型的多手段配套加工功能，还有自定义创建模板与数据导出，系统提供了灵活的模板设计功能，数据使用者可以根据自己的需求创建各种档案类型的模板，并将数据导出为常用的文件格式，如 json、Excel 等



数据架构员操作功能介绍

一、数据架构

1、查询浏览设置

该模块为数据浏览模块中数据显示的前置设置

1.1 类目架构设置



在左侧调阅树架构可对架构右击执行节点的增删改操作，并选择【第一类数据表链接】、【第二类数据表链接】两者为数据显示的两种模式。

设置完成后即为数据浏览模块中左侧的树状结构。同时【初始化类目(慎用)】，此功能是将树架构内容重置为类目设置。

1.2 数据集元数据

为不同档案类型在数据浏览模块中字段显示的设置，同调阅软件的 config 设置

首先输入本设置的名称及档案类型名称、水印

数据元数据设置

← 显示及搜索字段详细参数

设置名称

设置名称*: 城建

档案类型*: 城建档案

附件水印设置:

上一步 > 取消

选择在数据库中对应的数据表、需要显示浏览的字段、where 子句为字段限制条件设置(比如归档年度在输入时要求大于和/或小于某值, 即 a.归档年度>2010 and/or a.归档年度<2015)、orderby 字段为数据的顺序逻辑字段, 如案卷号及顺序号, 在数据表中可选择多个数据库, 如图第一个数据库为 a, 第二个为 b, 即在字段名设置上, 设置不同数据库时, 要在字段名前分别写明 a.XXX 和 b.XXX, “a.”、“b.” 称之为别名

数据元数据设置

← 显示及搜索字段详细参数

显示数据SQL语句设置

数据SQL语句(多张表的, 可以按 选定数据表->选择字段->选定数据表->选择字段 进行)

选定数据表: 城建database

数据表*: 城建database a,城建database b

选择字段: 案卷顺序号, varchar, 案卷号, va...

字段名*: a.案卷顺序号,a.案卷号,a.顺序号,a.责任者,a.文件图纸名称,a.文件图纸编号,a.页数,a.图幅,a.起始页号

where子句*: 1=1

orderby字段*: a.案卷号,a.顺序号

上一步 > 取消

设置要在浏览界面显示的字段中文名称、隐藏字段、显示时的宽度百分比、查询的字段范围

在 pgsql 中，只能对 String 类型使用模糊查询，因此这里的可搜索只代表了字段检索字段，模糊查询时会自动过滤掉其中不是 String 类型的字段。

下一步是设置档号字段、图像名字段、图像路径字段，其中档号字段是必选，另两个字段有就选，没有就不选；设置数据显示的样式，有字段 gridview 和卡片 cardview 两种形式



【子表单设置(假如有)】针对同时有案卷和卷内的情况，案卷目录的子表单即为对应档案的卷内目录，设置父表(案卷表)与子表(卷内表)之间的关联数据字段，即可根据案卷目录的档号直接找到卷内目录内对应该档号的数据



子表单字段	父表单字段	
a.室编档号	a.室编档号	

设置数据导出时需要关联的模板，有目录、封面、备考表三种，这些模板在 EXCEL 模板设置中设置添加

数据集元数据设置

显示及搜索字段详细参数

EXCEL表格输出设置

目录表格输出列表	封面表格输出列表	备考表格输出列表
<input checked="" type="checkbox"/> 文书卷内 <input type="checkbox"/> 业务案卷 <input type="checkbox"/> 业务卷内 <input type="checkbox"/> 城建卷内 <input type="checkbox"/> 城建案卷 <input type="checkbox"/> 文书案卷	<input checked="" type="checkbox"/> 文书封面	

下一步 >

取消

最后一步点击完成，即完成了设置

数据集元数据设置

显示及搜索字段详细参数

确认各设置参数

搜索字段类型: string,string,string,string,string,string,string,string,string

文件名显示字段: a.aid

图像名字段: a.relpah

图像路径字段: a.cdpath

显示表单样式: gridView

输出Excel表单设定:

子表单显示设置ID: 0

子父表单关联字段:

水印选择: 水印第一种

完成

取消

2、数据 ETL 设置

针对利用成品数据直接导入系统利用的情况，分别设置导入规则和全文数据处理

2.1 数据 ETL 规则

选择需要上传的数据库或 excel 文件所在的文件夹



选择刚才上传的数据库或 excel 文件的数据库名称/工作表名称、
orderby 字段为数据的顺序逻辑字段、where 字段则根据需要能将数据的顺序逻辑字段部分抽取指定数据范围的逻辑设置，如案卷顺序号 <50，则只导入案卷顺序前 50 的数据(假设案卷顺序号为 int 类型)，如果全部导入则不修改、图像文件夹名称字段、图像文件名排序字段



上传数据规则设置

目录与图像
设置相关的参数

数据表设置：

数据表表名：

orderby字段：

where子句：

目录与图像对应关系(以MDB字段,比如 档号 表示文件夹名,以 顺序号 定位图像文件排列)

图像文件夹名称：

图像文件名排序：

< 上一步 下一步 > 取消

输入图像存放文件夹名，默认为 FTP 中的 backup、选择导入的数据库名称、图像文件名称字段、图像文件夹字段、及档号所在字段

上传数据规则设置

设置服务端
数据表中的图像文件字段、图像路径字段、图像存放服务器文件夹

服务器端FTP上传图像存放文件夹名

图像存放文件夹名：

服务器端数据表中的图像文件字段、图像路径字段

目标数据表*：

图像文件名字段：

图像文件夹字段：

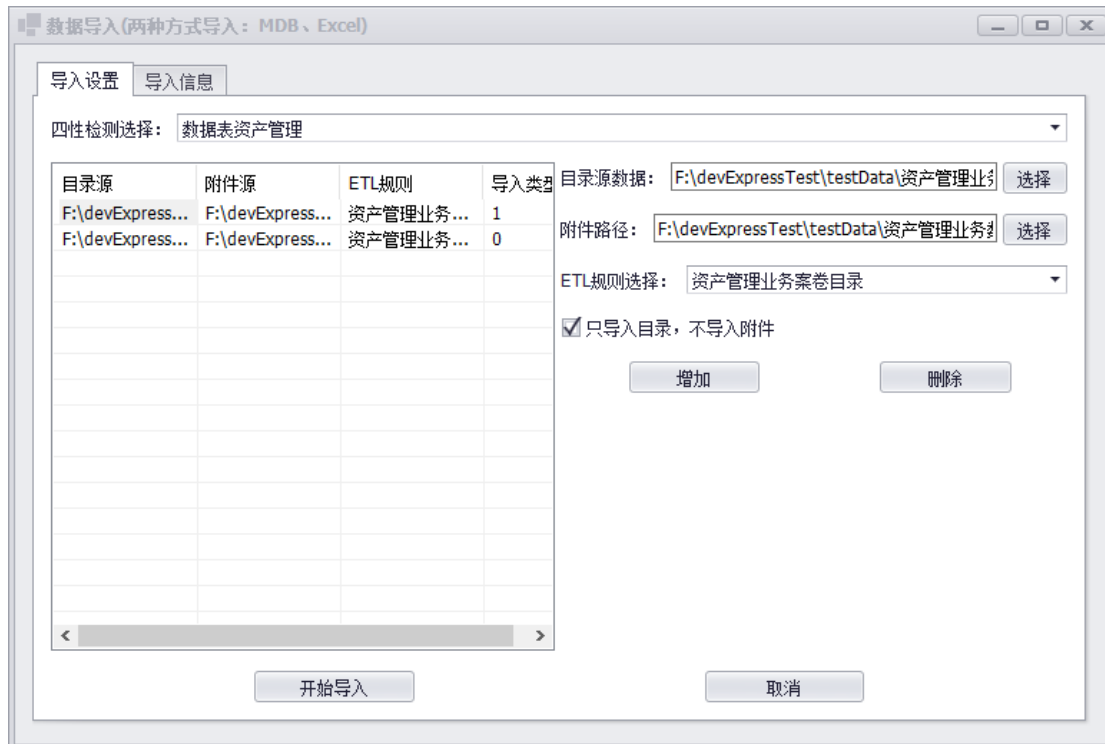
自增列：

档号字段*：

图像文件字段和图像路径字段不需要在下一步 源数据与字段对应关系 中存在。假如没有，空自增列是指目标数据表的自增列，一般为ID，假如没有，空

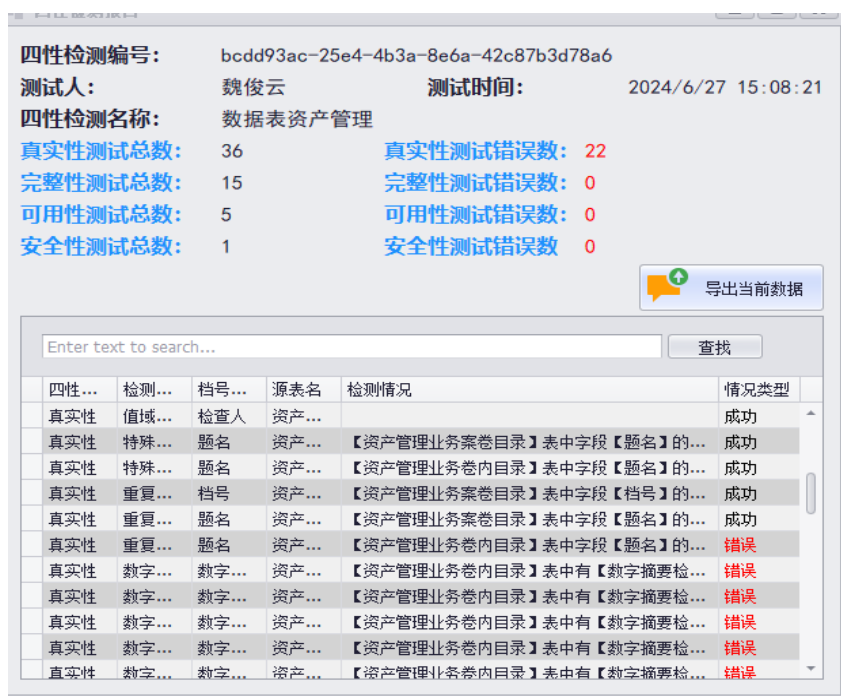
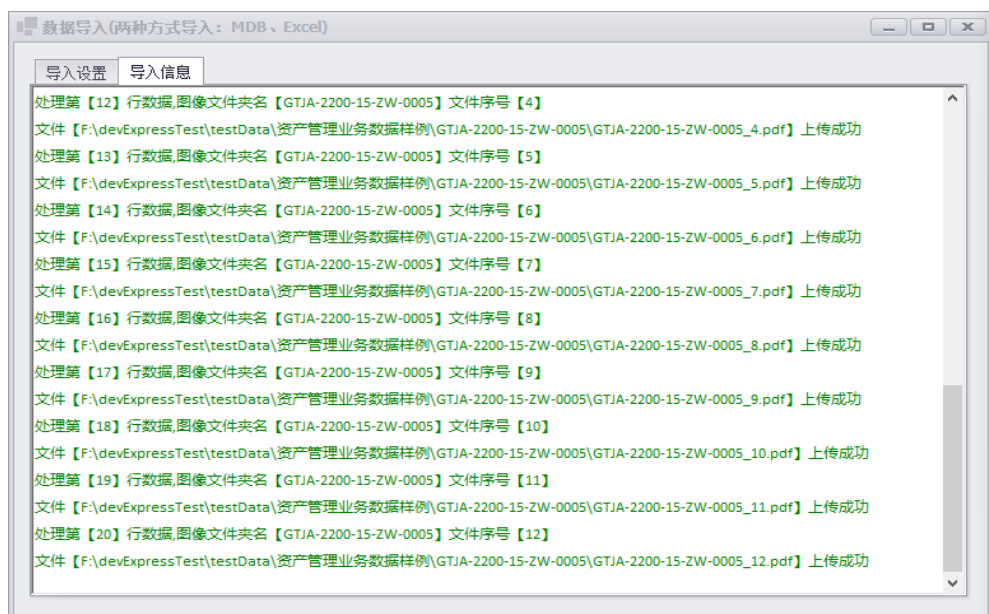
< 上一步 下一步 > 取消

指源数据与数据库字段对应关系设置，如两张表字段完全一致则点击字段一键匹配即可，也可通过选择数据源字段和目标数据库字段进行手动的字段一一匹配

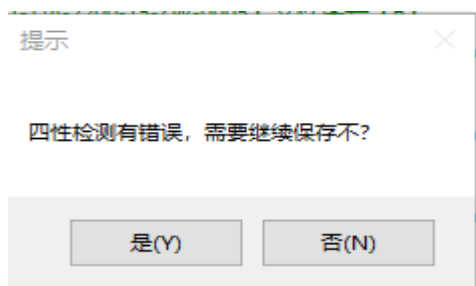


如上图，“增加”按钮为列表增加要导入的 ETL 规则、数据所在位置等。可以混合导入 EXCEL 和 Access 数据库，只要他们可以是一个数据集。点击“开始导入”，会到“导入信息”界面，如下图所示，根据选定的【四性检测规则】进行四性检测，弹出检测报告，然后根据选择是否保存进行数据保存。





假如四性检测报告中错误太多,建议先不要保存,整改数据后再次导入。



2.3 全文数据处理

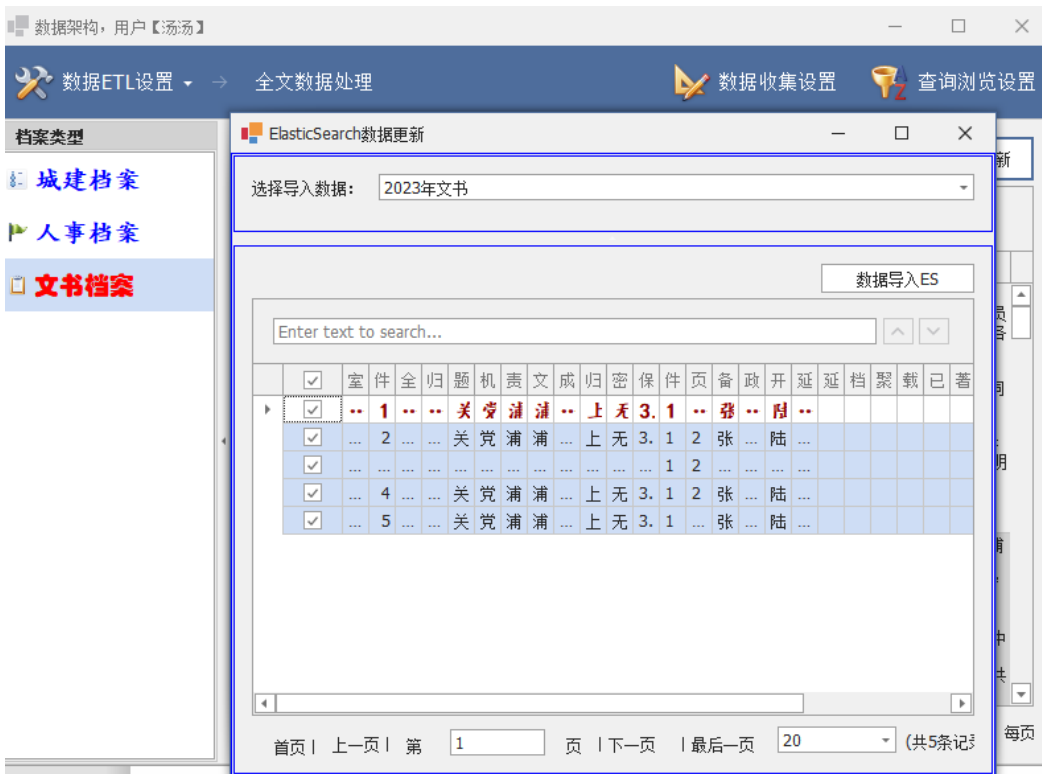
先进入服务器端，双击打开 elasticsearch.bat，如下图



进入本模块后会提示链接成功

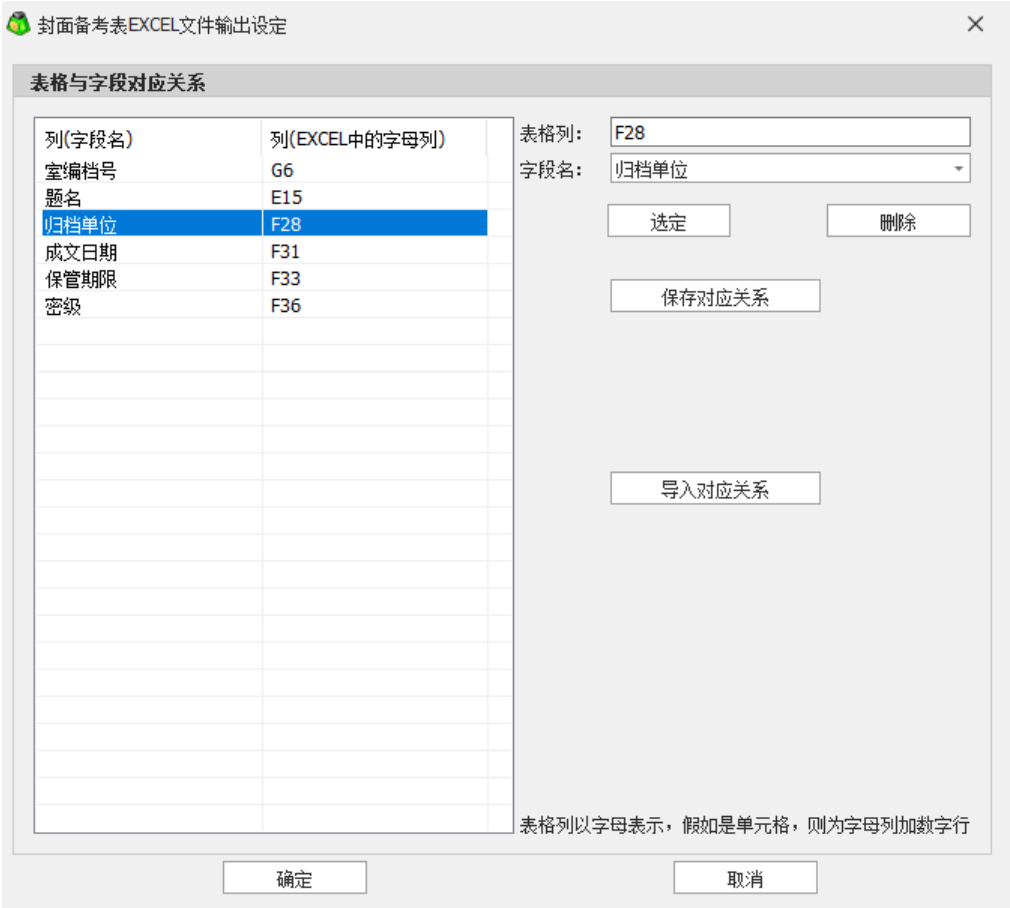


本模块是将在著录环节中录入的数据，包含双层 PDF 导入到全文检索 ES 中，导入后即可在数据浏览-全文检索中检索到

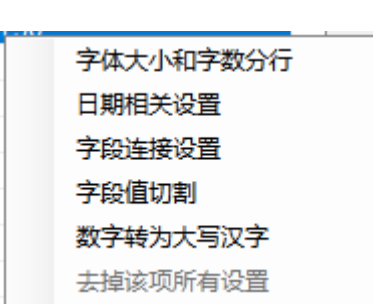


导入成功后在本界面会显示出相应档号被全文检索识别出的文字数据

在表格列输入模板导入的位置，如果导入位置是列的则只输入列所代表的字母列，如果是指定的某个单元格的则输入字母加数字行；在字段名下拉选择数据库中对应的字段，即提取依据，每输入一条逻辑点击确定



高级设置：对输入的任意一条逻辑鼠标右击，有关于字体、日期、字段连接、字段切割和数字转换大写功能，具体操作方式如下：



2.4.1 字体大小和字数分行

字体设置

☐ 按指定字数分行 ☒ 按字数设置字体大小

指定字数: 18 间隔行数: 1 最大行数: 4

字体大小设置

字数大于	字体大小
15	11
17	9
25	8
30	6

字数大于: 25 字体大小: 8

选定 删除

确定 取消

分行逻辑为按指定的字数分行；还有一种为按字数设置字体大小即根据字数的数量进行的缩字逻辑：

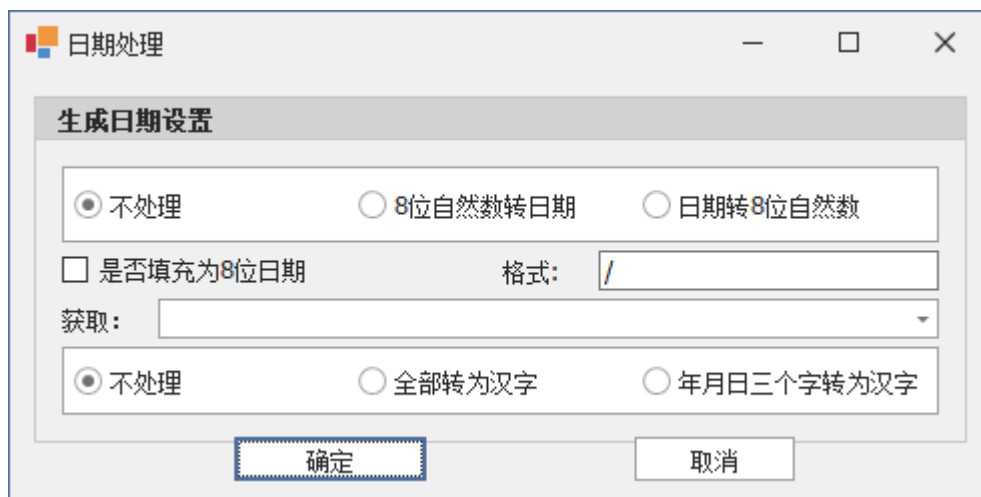
勾选按指定的字数分行：输入数字数量，即多少字后换行，间隔行数和最大的填充换行行数。

勾选按字数设置字体大小：设置字数数量与字体大小的逻辑，在 excel 中字体的大小有 11、9、8、6 这几种，根据实际字数的数量去设置

2.4.2 日期相关设置

当源数据与模板要求的日期格式不一致时，需要此功能转换，目前有 8 位自然数和带符号日期格式互相转换的功能，当需要 8 位自然数转日期时，勾选“是否填充位 8 位日期”，并选择相应的日期连接

格式，默认为“/”；当需要日期转为 8 位自然数时，选择获取下拉框内源数据日期的形式即是否有年、月、日数据，然后根据需要选择时全部转为汉字还是只有年月日这三个字转为汉字



日期处理

生成日期设置

☒ 不处理 ☐ 8位自然数转日期 ☐ 日期转8位自然数

☐ 是否填充为8位日期 格式: /

获取:

☒ 不处理 ☐ 全部转为汉字 ☐ 年月日三个字转为汉字

确定 取消

2.4.3 字段连接设置

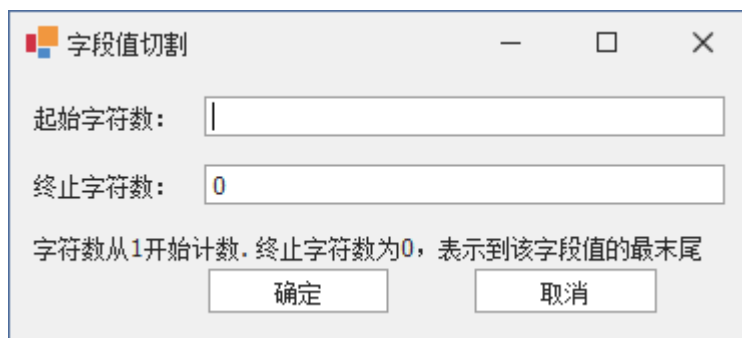
用于在填充时对指定字段的前后添加固定文本使用，选择需要利用的字段，也可由阿拉伯数字转为汉字，在前后文本中输入固定的文字，并可指定在本字段的前后位置，点击确定





2.4.4 字段值切割

这个功能就是把源数据字段内的文字在指定字符数后切割出去，输入开始切割所在的字符数，终止则填切割区间的最后一个字符数



3、数据收集设置

本模块是著录模块的前置设置，包含数据著录架构、表字段元数据、审核显示字段

3.1 数据著录架构

为数据著录界面的树状结构的前置设置，对节点右击由增删改节点名称功能，另外有【编目审核链接】，即设置该节点下所有数据在著录完成后提交审核的流程；【数据著录链接】，将节点与 3.2 表字段元数据关联，代表点击本节点著录的档案类型，注意【设置/清空审核流程链接】只能在父节点上设置，多个子节点可共用一个审核流程链接，而子节点只能【设置/清空著录链接】，审核逻辑会根据各子节点的著录连接设置向上寻找对应的最近审核流程链接执行审核逻辑，若没有审核逻辑的则自审核（没有其他角色审核，自己审核），即著录后保存时自动审核完毕





在本模块新增过的父/子节点后，需要安全保密员在系统设置中给予指定角色的树架构赋权，才能让角色看到并使用相应节点下进行著录

3.2 表字段元数据

该模块的为档案著录字段的前置设置，包括此设置的名称、著录时的展示类型、默认著录行数、对应的数据库、顺序号字段、档号字段、图像文件名字段、图像路径字段，图像字段设置后即可一边著录一边上传图像，界面上带*的字段为必填项

表字段元数据设置向导

← 表字段元数据设置

表的元数据设置

数据录入设置名称*：

文书卷内

输入表格展示类型*：

gridView模式

cardView模式

输入表格缺省行数*：

10

数据录入表表名*：

文书卷内

顺序号字段：

附件序号

临时档号字段*：

室编档号

盒号字段：

图像文件名字段：

图像名称

图像路径字段：

图像路径

上一步 >

取消

22

选择著录后导出的卷内、封面、备考表模板，在 1.3EXCEL 模板设置中设置

表字段元数据设置向导

← 表字段元数据设置

Excel输出模板选择

目录表格输出列表	封面表格输出列表	备考表表格输出列表
<input checked="" type="checkbox"/> 文书卷内 <input type="checkbox"/> 业务案卷 <input type="checkbox"/> 业务卷内 <input type="checkbox"/> 城建卷内 <input type="checkbox"/> 城建案卷 <input type="checkbox"/> 文书案卷	<input type="checkbox"/> 文书封面	

下一步 > 取消

服务器存储路径为 FTP 默认路径，不用更改，父级数据录入设置为著录的卷内有案卷需要著录的则需要选择，会在著录卷内保存时跳出案卷著录界面，审核显示字段设置见审核显示字段设置，存在替代数据表为选择数据库中设置的著录代号数据库

表字段元数据设置向导

← 表字段元数据设置

替代或者样例数据表设置

服务器存储路径: /temp

父级数据录入设置:

审核显示字段设置: 审核显示2资产管理

四性检测设置: 数据表资产管理

☐ 存在替代数据表

替代数据表设置

数据表表名:

select字段:

where子句:

orderby子句:

下一步 > 取消

这里可以对著录的数据设置是否要进行四性检测，假如设置了，在著录数据保存时会进行四性检测并弹出检测报告，关闭检测报告后会弹出提示框提示是否继续保存。

下一步为详细的著录字段设置，首先可对默认的字进行删除或增加一行的设置

设置	字段名称	字段类型
设置	附件文号	varchar
设置	责任者	varchar
设置	题名	varchar
设置	附件日期	varchar
设置	起始页号	varchar
设置	文件页数	varchar
设置	备注	varchar
*		

点击字段左侧的设置按键，可对每一个字段的进行详细的字段属性设置。对于该字段的输入，可以多种多样，比如下拉框等。如下图：



分别设置字段的中文显示名称、宽度百分比、是否为必填项、值的最大长度。



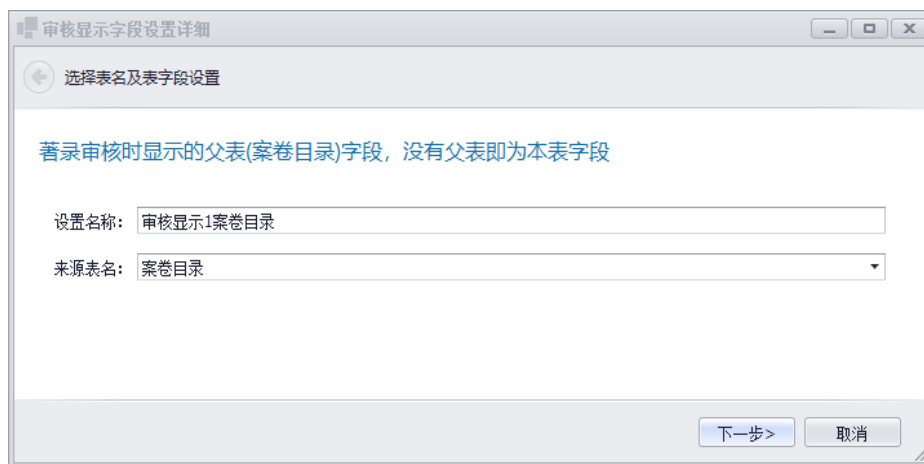
另外还可以利用源数据类型下拉字段中的数据字典(在系统设置的数据字典设置中设置)、或者卷内的件数、最小最大日期、最大页数自动生成。

3.3 审核显示字段

该模块是为了著录信息在审核完成时显示字段的前置设置，设置数据审核后显示的对应数据库、并明确要显示的字段名称与对应的字段宽度百分比。



首先填入设置名称并选择来源表名，如下图所示：



审核显示字段设置详细

选择表名及表字段设置

著录审核时显示的父表(案卷目录)字段, 没有父表即为本表字段

设置名称: 审核显示1案卷目录

来源表名: 案卷目录

下一步> 取消

点击“下一步”后显示全部字段，选择要显示的审核字段名称，录入对应对应的字段中文名和字段宽度百分比，如下图所示，对于原来就是中文名称字段的，对应的字段中文名可以不填，系统自动将原字段当作中文名称字段。



审核显示字段设置详细

选择表名及表字段设置

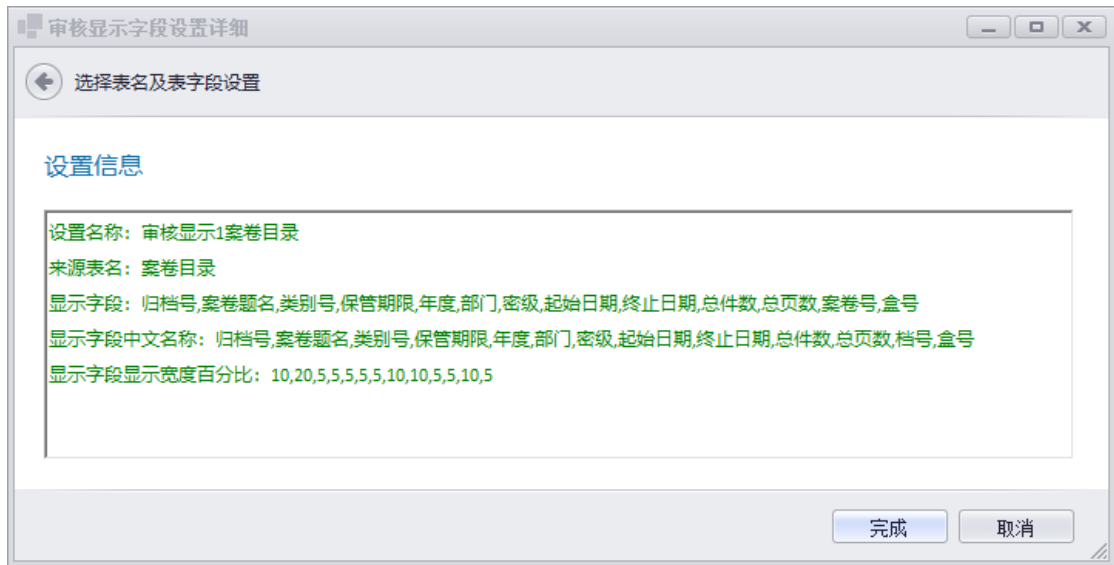
设置显示字段、显示字段的中文名、值的宽度

Enter text to search... 查找

是否选择	字段名称	字段中文名称	字段值宽度
<input type="checkbox"/>	全宗号		0
<input checked="" type="checkbox"/>	归档号	归档号	10
<input checked="" type="checkbox"/>	案卷题名	案卷题名	20
<input checked="" type="checkbox"/>	类别号	类别号	5
<input type="checkbox"/>	分类号		0
<input checked="" type="checkbox"/>	保管期限	保管期限	5
<input checked="" type="checkbox"/>	年度	年度	5
<input checked="" type="checkbox"/>	部门	部门	5
<input checked="" type="checkbox"/>	密级	密级	5
<input checked="" type="checkbox"/>	起始日期	起始日期	10
<input checked="" type="checkbox"/>	终止日期	终止日期	10
<input checked="" type="checkbox"/>	总件数	总件数	5
<input checked="" type="checkbox"/>	总页数	总页数	5
<input type="checkbox"/>	整理人		0
<input type="checkbox"/>	责任者		0
<input checked="" type="checkbox"/>	案卷号	档号	10
<input type="checkbox"/>	整理日期		0
<input type="checkbox"/>	检查人		0
<input type="checkbox"/>	检查日期		0
<input type="checkbox"/>	实物存放位置		0

下一步> 取消

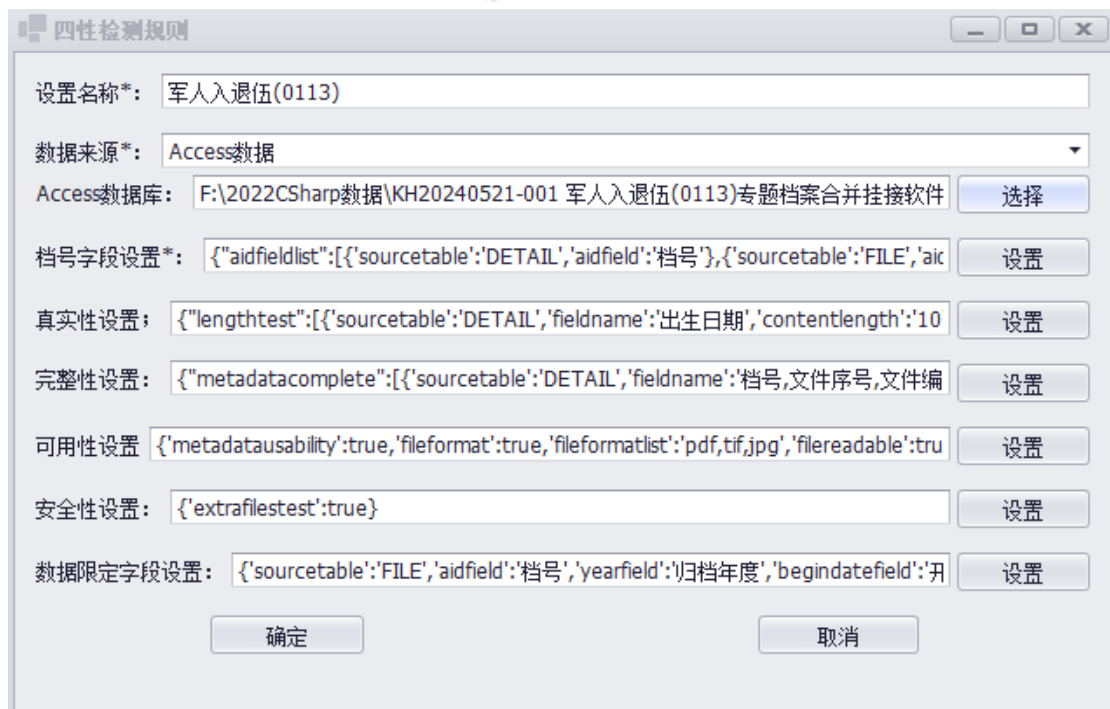
下一步就是显示设置信息，点击完成按钮结束设置



4、四性检测设置

4.1 四性检测规则设置

点击“增加”按钮为增加新的四性检测规则，点击“修改”如下图：

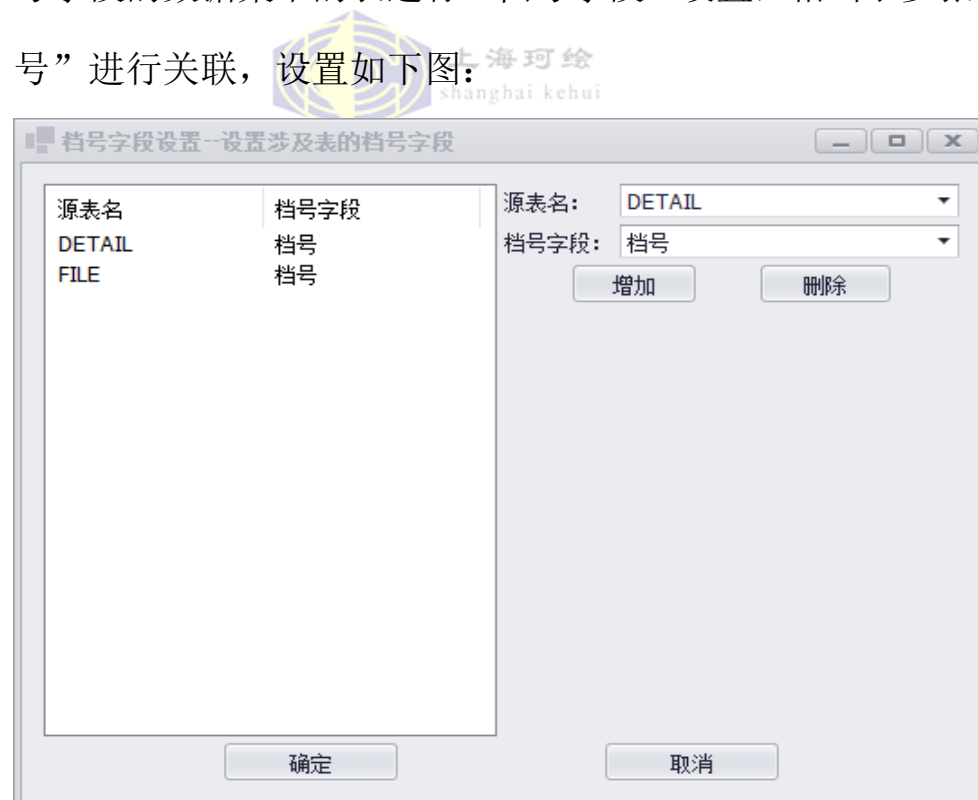


四性检测主要针对两类数据进行检测，一类是存储利用库中的表的

数据，一类为成品 Access 数据库（针对浦东新区档案局的要求）。

如上图，数据来源两类，一类数据表，一类 Access 数据。

四性检测时检测的也是 DataSet 数据集数据（多张相关联的表数据组成的数据集合），当然也可以针对单一表数据集进行检测。先对有档号字段的数据集中的表进行“档号字段”设置，相当于多张表以“档号”进行关联，设置如下图：



没有“档号”字段表，不需要进行设置，比如建设档案中可能的“main”、“unit”等表。

四性检测分为“真实性”、“完整性”、“可用性”、“安全性”，下面分别介绍配置。

“真实性”配置分为“长度检测”、“类型检测”、“值域检测”、“特殊字符检测”、“重复值检测”、“数字摘要字段”、“数字签章检测”7 大类。

“长度检测”如下图，主要是对字段的值进行内容长度的检测。界面上只有“增加”、“删除”两个按钮，要做修改，则先点击左边的ListVIEW 中的条目，然后先删除本条目，修改后再点“增加”即可。



“类型检测”，对指定表的指定段，进行类型检测。字段类型分为：日期型、字符型、整数型、浮点型等档案数据常用类型。如下图所示：



“值域检测”，是对指定字段的值的范围进行了限定，如下图：

源表名	字段名称	字段值域	值域类
FILE	档案类型	传统载体档案,电子...	1
FILE	聚合层次	卷,件	1
FILE	载体类型	{"fieldcondition": {...	2
FILE	已数字化	{"fieldcondition": {...	2
FILE	已数字化	{"fieldcondition": {...	2

源表名:

字段名称:

字段值域:

增加 删除

字段条件设置

确定 取消

字段的可能值以英文逗号隔开，对于有条件的字段值，即同表中某字段为某值，则本字段为某值，比如，上海档案“FILE”表中“档案类型”字段为“传统载体档案”时，则“已数字化”字段填为“是”，如下图：

假如字段:

比较符:

假如字段值:

则字段:

则字段值:

确定 取消

比较符列表如下图：

假如字段: 档案类型

比较符: (字符)为

假如字段值: (数值)等于
(数值)大于
(数值)小于
(数值)不等于
(字符)为
(字符)包含

则字段:

则字段值:

“特殊字符检测”如下图，在对指定字段值进行特殊字符的检测。

四性检查之真实性

长度检测 类型检测 值域检测 特殊字符检测 重复值检测 数字摘要字段 数字签章检测

特殊字符: <, >, ', ', =, {, }, *, &, %

源表名	字段名称
DETAIL	题名
FILE	题名

源表名:

字段名称:

增加 删除

确定 取消

“重复值”检测，检查字段的值是否有重复，主要是针对题名、身份证号等字段。

四性检查之真实性

长度检测 类型检测 值域检测 特殊字符检测 重复值检测 数字摘要字段 数字签章检测

源表名	字段名称
FILE	档号
DETAIL	身份证号

源表名:

字段名称:

增加 删除

确定 取消

“数字摘要”是对附件文件的一种加密算法，本系统中是 SHA-2 下细分出的一种算法：SHA256，系统对附件文件进行 SHA256 后与表中的数字摘要字段值进行对比。下面的附件路径字段一般指最后一层的文件夹，对于多层目录，自动查找最后一层。对于有“附件路径”全文字段的，选择“附件路径”字段。。附件文件命名可以多种多样，但一般会有一个排序字段，比如“文件序号”字段等。假如有“附件文件名称”字段，则选择“附件文件名称”字段。



“数字签章检测”，对 PDF 文件中的数字签章进行检测，一般用于城建档案中的图纸 PDF 等。如下图所示：

四性检查之真实性

长度检测 | 类型检测 | 值域检测 | 特殊字符检测 | 重复值检测 | 数字摘要字段 | 数字签章检测

签名比对规则

☐ 有数字签章检测

☒ 按签名个数 签名个数: 6 一般情况下, 签名个数大于: 3

签名者检查:

源表名:

附件路径字段:

附件名称字段(排序):

确定 取消

“完整性检测”，分为“元数据项是否完整”、“必填项检测”、“连续性检测”、“附件路径检测”、“附件检测”、“页数是否统一检测”、“档号检测” 7 个小模块。

“元数据项是否完整”，如下图，检测表中的字段是否完整，以英文逗号分隔各字段：

四性检测之完整性

元数据项是否完整 | 必填项检测 | 连续性检测 | 附件路径检测 | 附件检测 | 页数是否统一检测 | 档号检测

字段名称	源表名
档号,文件序号,文...	DETAIL
档号,全宗号,归档...	FILE

源表名:

字段名称:

增加 删除

确定 取消

“必填项检测”，一些字段必填，比如档号、题名等。有些字段是根据其他字段的值来判断是否必填，如下图，可以点击“根据其他字段条件来判断是否必填”按钮：

The dialog box '四性检测之完整性' has several tabs: '元数据项是否完整', '必填项检测', '连续性检测', '附件路径检测', '附件检测', '页数是否统一检测', and '档号检测'. The '必填项检测' tab is selected.

字段名称	源表名	必填条件
档号	DETAIL	
数字摘要	DETAIL	
数字摘要生成...	DETAIL	
题名	DETAIL	
题名	FILE	
档号	FILE	
数字化处理员	DETAIL	{ "fieldconditio...
数字化处理员...	DETAIL	{ "fieldconditio...

On the right, there are input fields for '源表名:', '字段名称:', and '必填条件:'. Below them is a button '根据其他字段条件来判断是否必填', and two buttons '增加' (Add) and '删除' (Delete).

The '字段必填条件' sub-dialog is open, showing the following configuration:

- 假如字段: 文件类型
- 比较符: (字符)为
- 假如字段值: 数字化副本

Buttons '确定' (OK) and '取消' (Cancel) are at the bottom of the sub-dialog.

上图的含义为：假如字段“文件类型”的值为“数字化副本”，则字段名称“数字化处理员”必填。

“连续性检测”，分为区域连续性和全局连续性。比如“文件序号”，

一般按档号进行区域连续性，全局连续性一般比较少，比如表中的 ID。假如连续性为字符的，可以取最后位数进行连续性比较，比如档号的最后几位是流水号，可以进行连续性检测。下面的 check 列表中，不选“从上至下大小逻辑检查”，则连续性是指持续加 1，比如文件序号。选了，则是下值比上值大就行了，比如起始页号。如下图所示：



“附件路径检测”，对附件的上层目录进行设置，有些是字段值，有些是固定值。对于数据表的，要选“附件路径所在表”和“附件路径字段”，可以进行比对。对于成品 Access 数据库（表中无附件路径字段，假如有，就要选择表和字段），要选如下图，确定 Access 数据库所在层数及 MDB 名称值。



特别值得一提的是，“附件检测”依赖于“附件路径检测”，没有附件路径，就不能定位附件所在位置，至少对于数据表，要选择“附件路径所在表”和“附件路径字段”。

在著录数据和导入数据时设置进行四性检测，系统会自动根据“附件路径检测”中的路径数据，对著录或导入的数据路径进行调整，以适应“附件路径检测”。

“附件检测”，如下图所示，附件名称有多种可能性：

- 1) 有专门的文件名字段，或者是其他字段代替文件名称，比如文件序号，8 位，不足部分以零填充前面。
- 2) 前面是档号，中间连接符，最后是流水号（文件序号），可能是 8 位，也可能是自然数。

四性检测之完整性

元数据项是否完整 必填项检测 连续性检测 附件路径检测 附件检测 页数是否统一检测 档号检测

附件名称(不带后缀)

源表名: ☐ 附件名称字段:

☐ 文件名字段: 8 位 ☒ 文件格式字段:

☒ 字段1: ☐ 转自然数 字段2: ☒ 转自然数 8 位

附件字节数及PDF文件图像

☒ 附件字节数不能超过 MB ☒ PDF内单个图像不能超过 MB

☐ 附件字节数比对字段:

☒ 检查每个PDF文件的页数是否与指定页数字段值相同 页数字段:

☐ 检查PDF文件图像图幅 图幅字段:

附件格式

☒ 每页图像DPI:

确定 取消

假如有附件名称字段，选择后会与实际文件名称进行比对。

附件字节数及 PDF 文件图像，主要针对 PDF 文件且里面是图像合并后形成的 PDF 文件，有附件字节数比对字段、与指定页数比对、图幅比对等。附件格式，主要是 PDF 里面的图像的分辨率。

四性检测之完整性

元数据项是否完整 必填项检测 连续性检测 附件路径检测 附件检测 页数是否统一检测 档号检测

☐ 检查指定表内的【文件页数】+【图纸页数】+【其他页数】=【总页数】

源表名: 总页数字段:

左边字段列表，英文逗号分开

☒ 检查件数与最大文件序号值是否一致

源表名: 件数字段:

对照表名: 文件序号字段:

☒ 检查总页数与文件页数之和是否一致

源表名: 总页数字段:

对照表名: 文件页数字段:

☒ 对照表文件页数中不包括封面卷内目录备考表页数 顺序号字段:

☐ 检查总页数与封口页后的页数是否一致

源表名: 总页数字段:

对照表名: 页次/页号 字段:

确定 取消

“页数是否统一检测”，各类页数的检查，页数内部是否统一等，如上图所示，有 4 种检查方式：

- 1) 表内的页数检查，多个字段之和是否与另一个字段相同。
- 2) 检查案卷目录的件数与卷内目录的最大序号是否一致。
- 3) 检查案卷目录的总页数与卷内目录的页数（每条页数）之和是否统一。假如有封面备考表页数，即卷内目录中有文件序号为零的，是否排除掉。
- 4) 检查案卷目录的总页数与卷内目录的最大页次封口号之后的数字是否相同。

“档号检测”检测档号的情况，两种检测，一种是检测档号是否在相应的案卷目录和卷内目录中都出现过，要是一个档号在案卷中出现，而在卷内没有出现，即为漏录。一种是检测档号本身的组成，将档号以分隔符分开。一般的档号主分隔符为“-”，在二级类目、年度、保管期限中会有二级分隔符，一般为“.”，比如档号组成方式：全宗号-门类代码.二级门代码.年度.保管期限-流水号，举个档号例子：

X349-ZP.01.Y.1998-0001，这里的 Y 是代指保管期限为永久。所以下图中的档号值数组顺序，对于以“-”分隔的，就是一维，只要输入分隔后的位数就行。比如 X349 就是 1，ZP.01.Y.1998，就是 2，0001 就是 3。对于 ZP.01.Y.1998，以“.”分隔的，就是二维数组，比如 ZP，就是 2,1；Y 就是 2,3；1998 就是 2,4。对于 Y，保管期限，可以按“根据其他字段条件来判断数组”，即保管期限为永久时，档号中的这个数组顺序为 2,3 的值是 Y。如下图所示：

四性检测之完整性

元数据项是否完整 必填项检测 连续性检测 附件路径检测 附件检测 页数是否统一检测 档号检测

☒ 同一档号在不同表中存在

☒ 检查档号字段值 档号值切割字符: ~.

档号数组	值类型	字段值	源表名
4	2	{"fieldconditio...	FILE
1	1	全宗号	FILE
3	1	归档年度	FILE
2	3	0113	FILE
4	2	{"fieldconditio...	FILE

档号值数组顺序:

源表名:

字段名:

其他:

根据其他字段条件来判断数值

增加 删除

确定 取消

档号值条件

假如字段: 保管期限

比较符: (字符)为

假如字段值: 永久

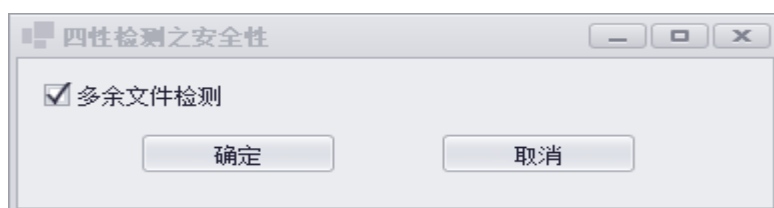
则档号相应值: Y

确定 取消

可用性检测，要先确保完整性检测中的“附件路径检测”和“附件检测”已经设置完成。可用性主要是检测附件是否可读，有没有损坏。元数据可用性检测，是检测各类连接，比如与 Access 数据库能否连接，能否打开数据库中的表等。文件格式是检测附件的格式，比如 PDF,TIF,JPG 等。如下图所示：



安全性检测，主要检测多余文件，如下图所示：



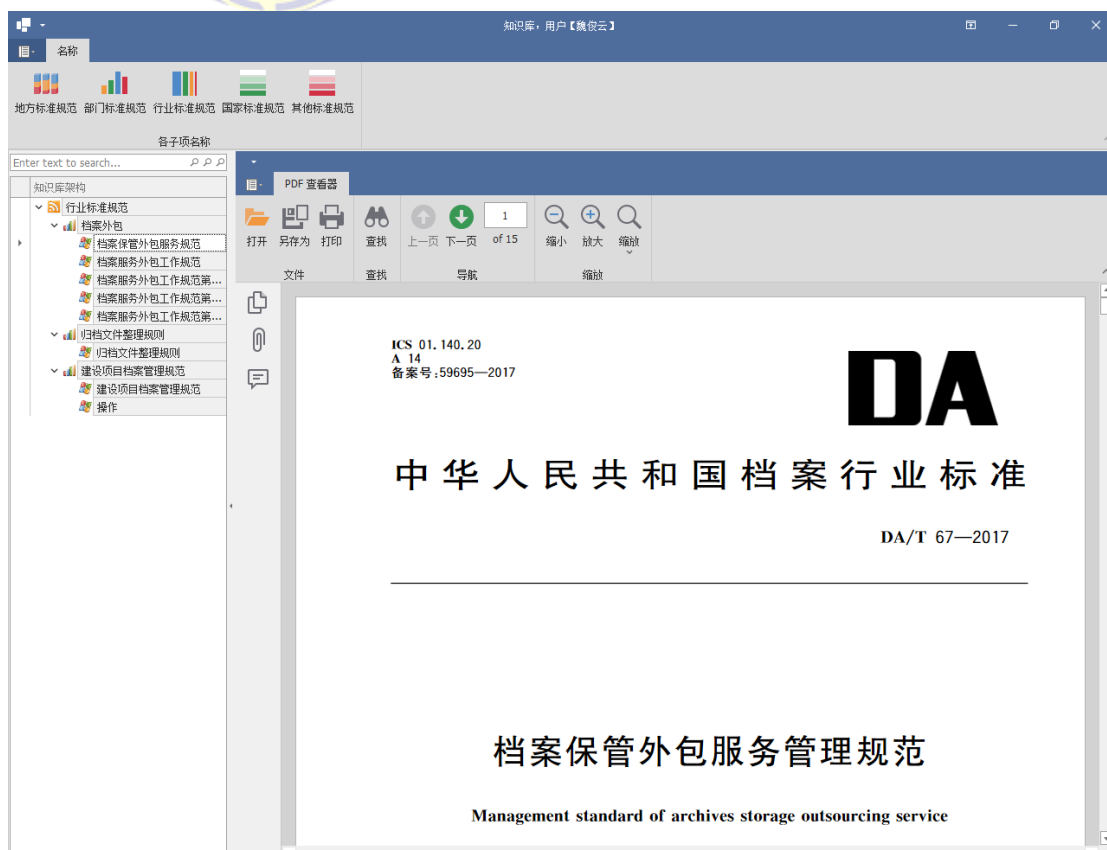
数据限定字段选择，是为了在 TOOL 工具库中进行四性检测时，对数据表或 Access 数据库中的数据进行限定，以防止数据量过大。限定字段主要是三种，一种是档号，档号的最后是流水号，可以对流水号进行限定。一种是年度，int 类型，选择其他 int 类型也可以。还有一种是日期型，比如 起始和终止日期。源表一般选择案卷目录表。



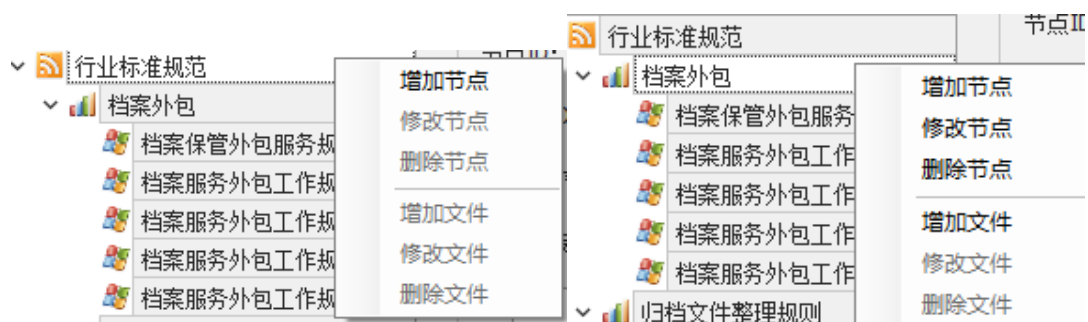
其他导出导入、删除之类的，与【查询浏览】中的【数据集元数据】一致。

二、知识库

本界面为多类型(word、pdf、xls、xlsx、MP4 等主流格式)档案规范性文件文件的汇总模块。由系统保密员在系统设置-知识库设置父模块，且给角色赋权后，相应角色即可在每个规范模块下对节点及节点文件进行增删改操作，并可上传多格式的附件文件，另外还有文件的相关使用控件，比如打印、另存、上一页、下一页、缩小、放大等。



对每个模块下得父节点鼠标右击有节点得增删改功能，若父节点下没有子节点，则也可上传/修改/删除文件；同样作用与子节点，在子节点下上传过文件得，则可执行子节点名称得增删改和文件上传功能



对已经上传得文件鼠标右击，则有文件名称修改和删除功能

新增文件

节点ID:

父节点ID:

文件名称:

文件上传:

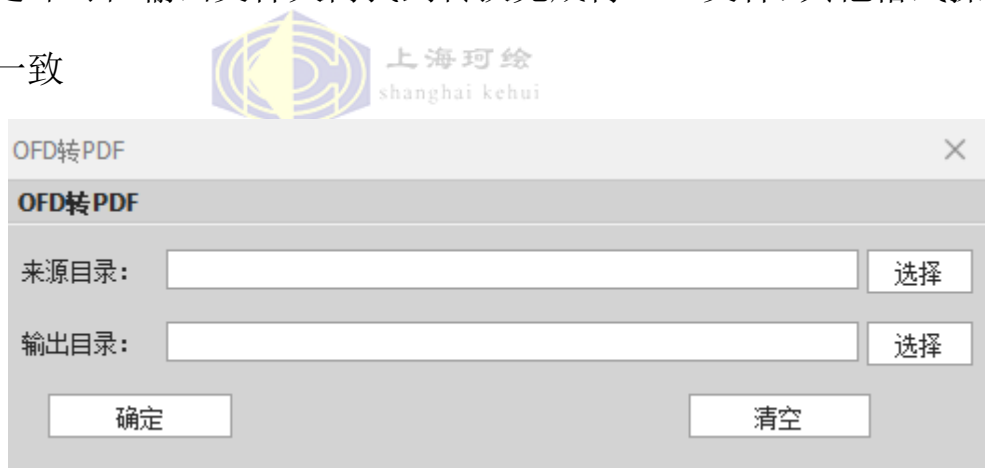
双击文件，弹出文件浏览控件，可执行另存、打印与图像浏览控件



三、工具集合

1、格式转换

格式转换主要是 PDF 与 OFD、JPG、TIF、TXT、XPS 之间的格式互转，以 OFD 转 pdf 为例，选择源数据 PDF 文件夹和输出路径，点击确定即可在输出文件夹内找到转换完成得 PDF 文件，其他格式操作方式一致




2、PDF 工具

包含对 PDF 文件的大小压缩、合并拆分、水印、数字签名与条码的添加、加解密等等



2.1 压缩 PDF 文件

可将 PDF 文件大小内存进行缩小，注意会影像图像质量



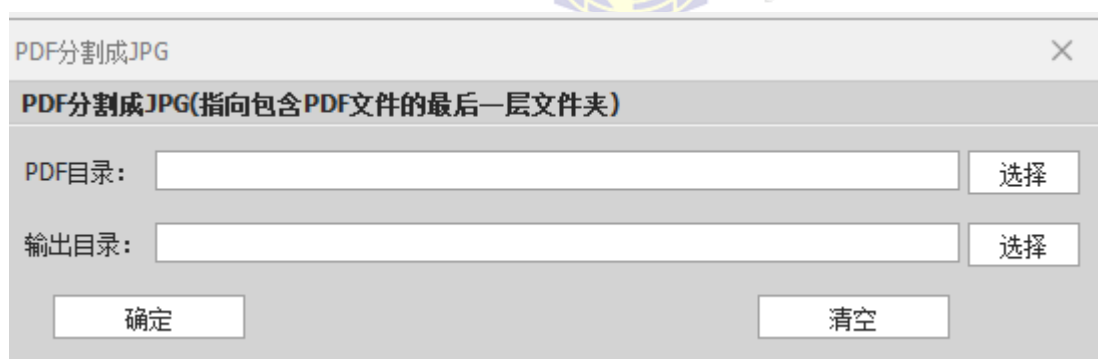
2.2 合并和拆分 PDF

将 PDF 按页数、指定页码区间范围这两种逻辑拆分输出；也可按照 PDF 文件所在的文件夹为单位合并为一个 PDF 文件夹



2.3 PDF 分割成 JPG

将文件夹下的 PDF 文件按单页分割为 JPG 图像



2.4 PDF 添加水印

选择图像或文字水印，再选择源数据 PDF 文件，输出后即是有水印的 PDF 文件

PDF添加水印

水印选择

☐ 图像水印 水印图像:

☐ 文本水印 文本名:

PDF目录:

输出目录:

2.5 PDF 添加条码

系统中有以下多种条码类型 Codabar、Code11、Code32、Code39、Code39-E、Code128-A、Code128-B、Code128-C、Code93、Code93-E，选择条件类型输入条码下的文本内容与位置，即可在 PDF 文件中的每一页添加相应条码

PDF 中添加条形码

条码选择

条码类型: 条码文本: 条码位置:

PDF目录:

输出目录:

2.6 PDF 添加数字证书

自备数字证书，输入其密码，并再签名内容中输入相关文本，即可在 PDF 文件中的每一页添加相应数字证书

PDF添加数字证书

数字证书选择

PDFX数字证书:

选择

证书密码:

签名内容

签名者标签: Digitally signed

签名者: Gary

日期标签: Date:

所在地标签: Location:

所在地: CN

原因标签: Reason:

联系标签: Contact Numb

联系方式: 15021219665

原因: Ensure authenticity

PDF目录:

选择

输出目录:

选择

确定

取消

2.7 PDF 文档加解密

设置打开密码后即可批量对 PDF 文件加解密

PDF加解密

密码设置

打开密码: open

权限密码: permission

PDF目录:

选择

输出目录:

选择

加密

解密

取消

2.8 查找和替换 PDF 文档中的文本

对于可编辑版的 PDF 文件，输入检索文字内容，可批量替换为“替换文本中”输入的文字



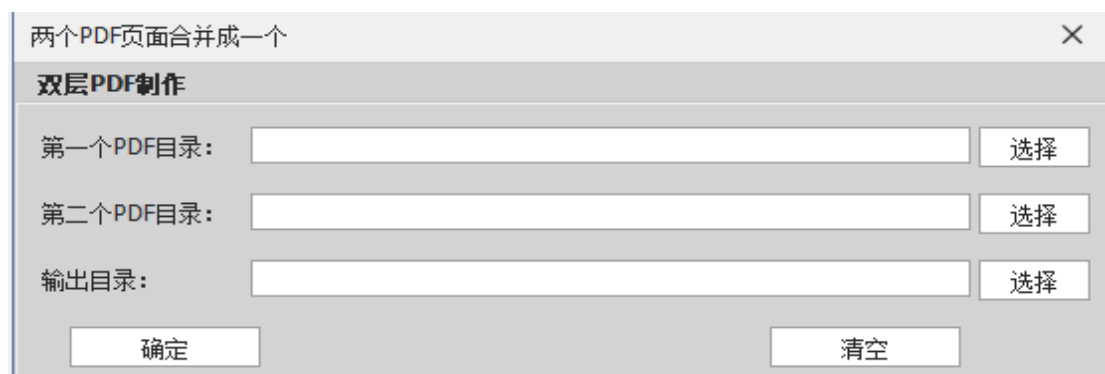
2.9 PDF 添加骑缝章

选择公章图像，可批量在 PDF 图像的骑缝位置添加相应的公章图像



2.10 PDF 重叠页面

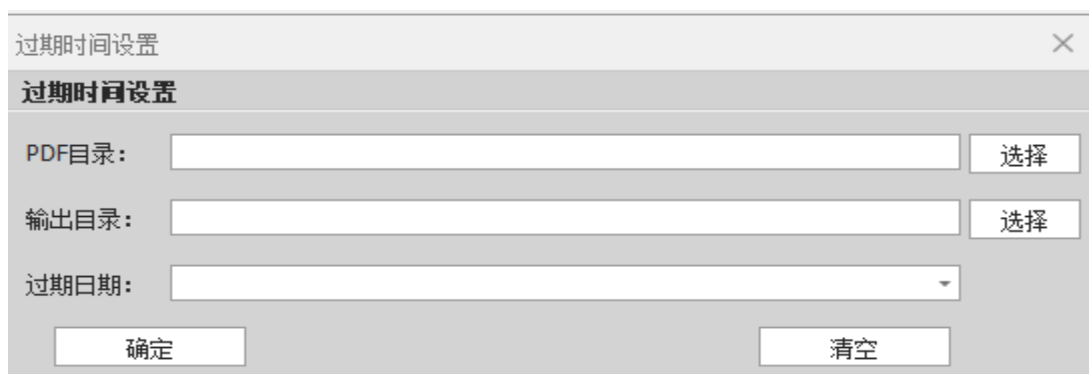
将两个文件夹内的同名文件的 PDF 图像进行重叠转化为双层 PDF



The dialog box is titled '两个PDF页面合并成一个' (Merge two PDF pages into one) with a close button (X) in the top right corner. Below the title bar is a section header '双层PDF制作' (Double-layer PDF creation). The main area contains three input fields, each followed by a '选择' (Select) button: '第一个PDF目录:' (First PDF directory), '第二个PDF目录:' (Second PDF directory), and '输出目录:' (Output directory). At the bottom, there are two buttons: '确定' (OK) on the left and '清空' (Clear) on the right.

2.11 PDF 设置过期时间

可批量将 PDF 设置打开的有效期时间没超过设定时间无法打开



The dialog box is titled '过期时间设置' (Expiration time setting) with a close button (X) in the top right corner. Below the title bar is a section header '过期时间设置' (Expiration time setting). The main area contains three input fields, each followed by a '选择' (Select) button: 'PDF目录:' (PDF directory), '输出目录:' (Output directory), and '过期日期:' (Expiration date). At the bottom, there are two buttons: '确定' (OK) on the left and '清空' (Clear) on the right.

2.12 PDF 添加动态图章

批量将 PDF 文件上添加可拖动的图章，填写图纸内容与姓名，系统自动在图纸上添加时间





添加动态图章

PDF目录: 选择

输出目录: 选择

动态图章内容:

图章用户姓名:


确定 清空

3、检查工具

3.1 PDF 图纸检查工具

3.1.1 检查图层和标记

选择需要检查的 PDF 文件夹，设置检查图层数量的最小值，点击开始，可检查每个 PDF 文件的图层是否大于设置的最小图层，并且检查 PDF 文件上是否有标记



PDF图纸检查工具

检查图层和标记

PDF目录: 选择

最小图层数:

开始检查

3.1.2 数字签字检查

选择需要检查的 PDF 文件夹，设置检查每个 PDF 文件的数字签名个数或设置检查是否大于等于多少个数字签名，在签名者检查中输入关键字，可检查数字签名中的属性是否包含关键字

数字签名检查

PDF图纸目录:

签名比对规则

☐ 按签名个数 签名个数: 一般情况下, 签名个数要大于等于:

签名者检查:



3.2 档案数据四性检测

数据源有两种, 一种是存储利用库, 一种成品 Access 数据库。



选择四性检测规则设置和相应的 Access 数据库文件, 对于数据表, 只需要选择规则设置。不选“是否检测全部数据”的话, 下面就是在四性检测设置中的数据限定字段设置的字段名, “【】”的为字段名称。可以限定要进行四性检测的数据量。

档案四性检测

数据准备

四性检测规则设置: 军人入退伍(0113)

Access数据库文件: F:\2022CSharp数据\KH20240521-001 军人入退伍(0113) 选择

☐ 是否检测全部数据

【档号】前缀:

【档号】流水号: 至

【归档年度】: 至

【开始日期】: 【终止日期】:

下一步> 取消

点击“下一步”，选择四性里要检测的，如下图：

四性检测项目选择

			
真实性	完整性	可用性	安全性
<input type="checkbox"/> On	<input type="checkbox"/> On	<input type="checkbox"/> On	<input type="checkbox"/> off

开始检测 取消

点击“开始检测”，即会进行检测，检测完成后会弹出四性检测报告：

四性检测报告

四性检测编号:

b5f86a34-c948-4255-b2f1-72447b0c1cd1

测试人:

魏俊云

测试时间:

2024/6/27 14:38:46

四性检测名称:

军人入退伍 (0113)

真实性测试总数:

31

真实性测试错误数:

17

完整性测试总数:

22

完整性测试错误数:

2

可用性测试总数:

3

可用性测试错误数:

0

安全性测试总数:

0

安全性测试错误数:

0

导出当前数据

Enter text to search...

查找

四性...	检测...	档号...	源表名	检测情况	情况类型
可用性	元数...				成功
真实性	长度...	出生...	DETAIL	【DETAIL】表中字段【出生日期】有值超过指定...	错误
真实性	长度...	身份...	DETAIL	【DETAIL】表中字段【身份证号】的【长度检测】...	成功
真实性	长度...	归档...	FILE	【FILE】表中字段【归档年度】的【长度检测】...	成功
真实性	长度...	归档...	FILE	【FILE】表中字段【归档日期】有值超过指定长...	错误
真实性	类型...	出生...	DETAIL	【DETAIL】表中字段【出生日期】的【类型检测】...	成功
真实性	类型...	文件...	DETAIL	【DETAIL】表中字段【文件序号】的【类型检测】...	成功
真实性	类型...	归档...	FILE	【FILE】表中字段【归档年度】的【类型检测】...	成功
真实性	类型...	开始...	FILE	【FILE】表中字段【开始日期】的【类型检测】...	成功
真实性	值域...	档案...	FILE	【FILE】表中字段【档案类型】的【值域检测】...	成功
真实性	值域...	聚合...	FILE	【FILE】表中字段【聚合层次】的【值域检测】...	成功

可以以 EXCEL 形式导出当前报告，以便进一步分析数据、整改数据。

4、OCR 工具：Umi-ocrV2.1.1

本工具可以将单层文件，比如 JPG、JPE、JPEG、jfif、png、webp、bmp、tif、tiff、pdf、xps、equb、mobi、fb2、cbz 等以上文件进行 OCR 识别，将其转换为双层 PDF 或将图像上的文字识别出另存为指定格式文件中



4.1 主界面

点击 OCR 工具下的 umi-OCR 模块弹出主界面



4.2 批量文档(可生成双层 PDF)

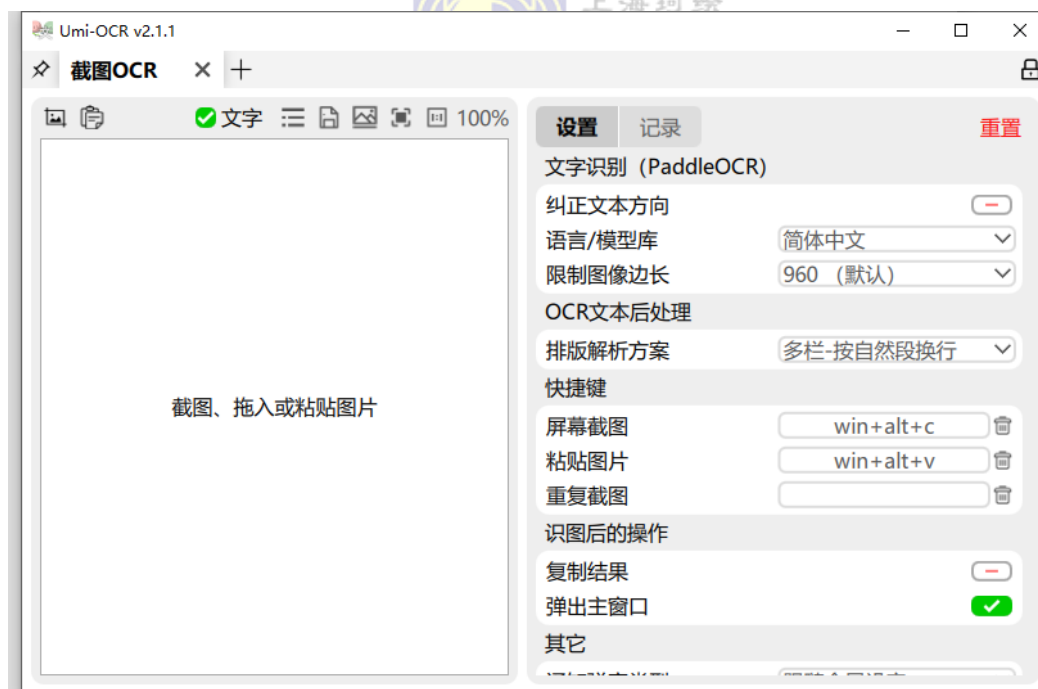
此模块支持 pdf、xps、equb、mobi、fb2、cbz，批量对以上格式的文件进行 ocr，将需要 OCR 的图像或文件夹拖入界面左侧。选择输出文件夹路径，和输出后的文本保存形式，点击开始即可 OCR 识别，生成双层 PDF





4.3 截图 OCR

此模块可将截图或者本地的单个图像进行 OCR 识别，设置右边截图和输出参数，可将识别出的单个图像弹出



4.4 批量 OCR

此模块支持 JPG、JPE、JPEG、jif、png、webp、bmp、tif、tiff 格式文件的批量 OCR 命令，设置输出文件夹与输出文档格式



4.5 二维码

可使用软件界面左上角的截图控件，或者复制二维码图像后点击粘贴控件，软件将自动识别二维码内容，并可对结果复制粘贴另存





联系方式

如果您在使用我们软件的过程中，发现存在某个问题或者感不方便的地方，请您将存在的问题和改进的方案以电子邮件或传真的形式发给我们上海珂绘数码影像科技有限公司技术部（电话：021 - 50317854 / 50316124 分机号：50317854-8002 传 真：021 - 50316124 E-mail: wjy72@sina.com 公司网址：www.shanghaikehui.com 公司地址：上海闵行区新骏环路188号15幢3层），我们将在尽可能短的时间内给您回复，并在短期内提供改进的新版本的 软件产品。同时，欢迎您能够在百忙之中不吝赐教。

（您可以在附表“软件反馈单”中将您的建议填写清楚，回寄给我们，此致敬礼。）

软件反馈单

软件反馈单	
配置式档案数据治理系统 V2.0	
客户信息	客户名称：
	填写人：
	所用软件：
使用评价及建议	
日期：	