档案识别系统需求说明

**文件修改记录**

|  |  |  |
| --- | --- | --- |
| 修改时间 | 修改内容 | 作者 |
| 2016/12/28 | 精简化 | Truman |

目录

[档案识别系统分析设计文档 1](#_Toc470860336)

[文件修改记录 2](#_Toc470860337)

[1. 引言 4](#_Toc470860338)

[1.1. 产品描述 4](#_Toc470860339)

[1.2. 名词术语 4](#_Toc470860340)

[1.3. 产品功能 4](#_Toc470860341)

[1.3.1. 系统管理 4](#_Toc470860342)

[1.3.2. 档案导出 4](#_Toc470860343)

[1.3.3. 档案识别 4](#_Toc470860344)

[1.3.4. 识别，导出模板设计 4](#_Toc470860345)

[2. 业务流程 4](#_Toc470860346)

[2.1. 角色说明 4](#_Toc470860347)

[2.2. 系统管理 4](#_Toc470860348)

[2.2.1. 系统登陆 4](#_Toc470860349)

[2.2.2. 系统配置 5](#_Toc470860350)

[2.2.3. 日志管理 6](#_Toc470860351)

[2.2.4. 用户管理 7](#_Toc470860352)

[2.2.5. 服务管理 8](#_Toc470860353)

[2.3. 档案导出 9](#_Toc470860354)

[2.4. 档案识别 10](#_Toc470860355)

[2.5. 识别，导出模板设计 11](#_Toc470860356)

[2.5.1. 系统登陆 11](#_Toc470860357)

[2.5.2. 识别模板设计 12](#_Toc470860358)

[2.5.3. 导出模板设计 13](#_Toc470860359)

[3. 交互原型 15](#_Toc470860360)

[3.1. 档案识别模板设计 15](#_Toc470860361)

[3.1.1. 识别模板管理-用户-登录 15](#_Toc470860362)

[3.1.2. 识别模板管理 15](#_Toc470860363)

[3.1.3. 识别模板管理-新建-模板信息 16](#_Toc470860364)

[3.1.4. 识别模板管理-新建-选择档案图像 17](#_Toc470860365)

[3.1.5. 识别模板管理-编辑 17](#_Toc470860366)

[3.1.6. 识别模板管理-删除 18](#_Toc470860367)

[3.1.7. 识别模板管理-菜单-文件 19](#_Toc470860368)

[3.2. 档案导出模板设计 20](#_Toc470860369)

[3.2.1. 导出模板-登陆 20](#_Toc470860370)

[3.2.2. 导出模板-设计 20](#_Toc470860371)

[3.2.3. 导出模板-管理 21](#_Toc470860372)

[4. 其他 21](#_Toc470860373)

[4.1. 接口需求 21](#_Toc470860374)

[4.2. 非功能和质量需求 21](#_Toc470860375)

[4.3. 方便和实用性需求 21](#_Toc470860376)

[4.4. 安全与保密性需求 22](#_Toc470860377)

[4.5. 稳定与可靠性需求 22](#_Toc470860378)

[4.6. 灵活性与可扩展性需求 22](#_Toc470860379)

[4.7. 界面友好需求 23](#_Toc470860380)

# 引言

## 产品描述

项目拟研发纸质文档识别系统，该系统可以将纸质文档的电子扫描件通过档案识别系统，生成结构化数据，并可将已有的结构化数据生成为PDF格式的电子文档服务，档案识别以WEB服务的形式部署，由用户提交文档扫描件服务对图片进行识别，并生成结构化数据供用户随时提取或以PDF格式导出。

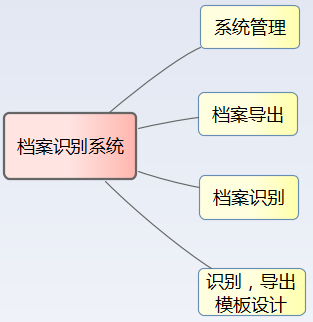
## 名词术语

1. 结构化数据

可供程序进行操作的文本数据，如：{"name":"张三"}。

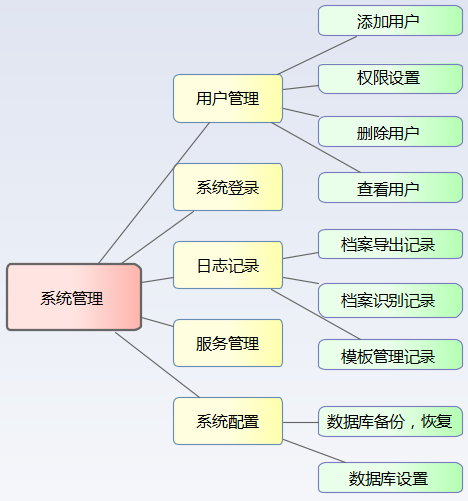
## 产品功能

档案识别系统，由：1、系统管理；2、档案导出；3、档案识别；4、识别，导出模板等子系统组成，每个子系统都可以独立运行。



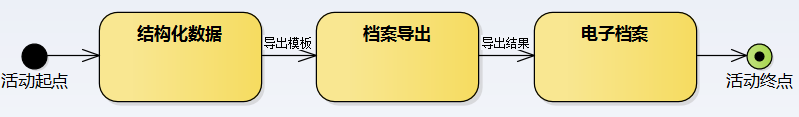
### 系统管理

档案识别系统的管理后台，负责识别服务的运行控制，系统数据库配置，用户权限管理，日志记录等工作，由：1、系统配置；2、权限管理；3、日志记录；4、系统登陆；5、管理等模块组成。



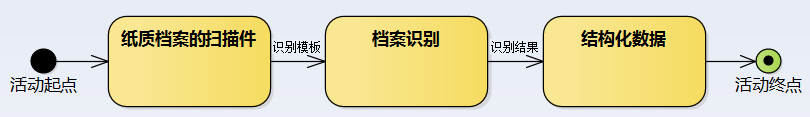
### 档案导出

以WEB接口的形式提供服务，直接面向所有用户，将数据库中的结构化数据根据指定的档案导出模板生成PDF文件，相当于是档案识别的逆向形式。



### 档案识别

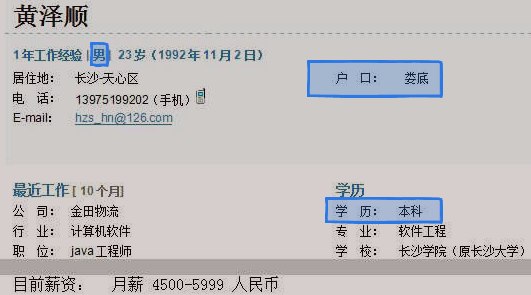
以WEB接口的形式提供服务，直接面向所有用户，将提交的纸质文档的图像根据指定的档案识别模板提取结构化的数据，相当于是档案导出的逆向形式。



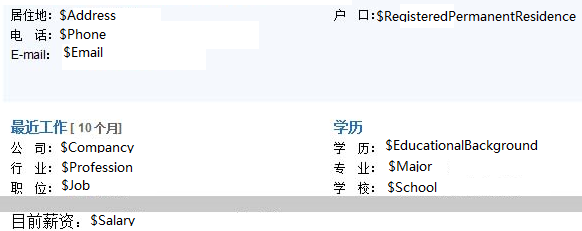
### 识别，导出模板设计

识别，导出模板设计实际为两个不同的部分：

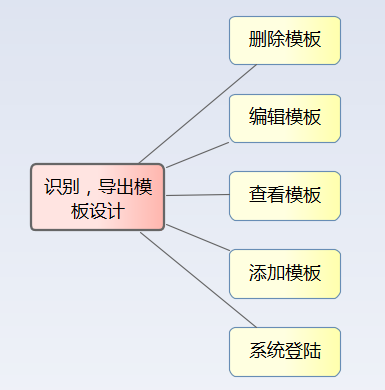
1. 识别模板设计是对每一个类型的档案文档进行区域定义，每一个蓝色区域代表一个字段，通过将文档划分为不同的区域来引导档案识别程序对各区域内的文字进行识别，并转化为结构化的数据。如下图是一份个人简历的识别模板：



1. 导出模板设计是创建一份用来填充数据的文档排版，其中用具有特殊标记的单词来对内容进行标识，具有特殊标记的单词与数据库中的结构化数据对应，供导出文档的时候将结构化数据显示在模板文档中。如下图是一份个人信息的导出模板：



无论是识别模板，还是导出模板，都由：1、删除模板；2、编辑模板；3、查看模板；4、添加模板、5、系统登陆等模块组成。



# 业务流程

## 角色说明

1. 系统管理用户

拥有系统的最高权限，可以对所有子系统进行访问。

1. 模板设计用户

拥有对模板进行管理的权限，由此用户来进行模板的设计以及管理。

1. 普通用户

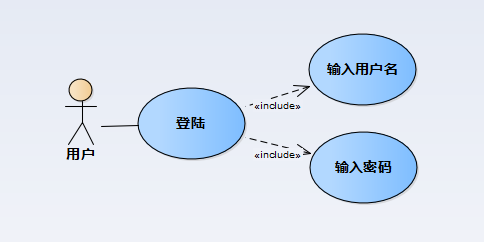
此用户是档案识别，档案导出的直接服务对象，默认属于临时的匿名用户，由”IP地址 + 端口”进行标识。

## 系统管理

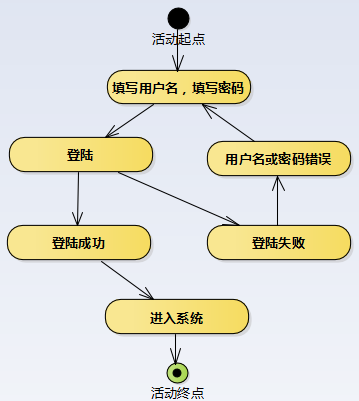
### 系统登陆

进入系统管理时，需要首先验证用户的是否具备管理用户权限，系统登陆模块将完成这项工作。

1. 用例



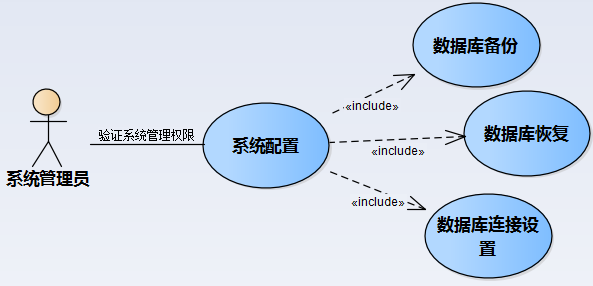
1. 流程



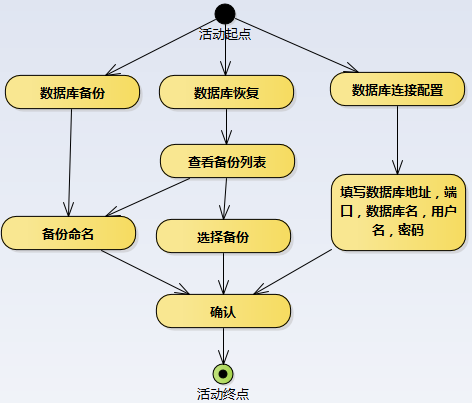
### 系统配置

为了防止数据因为事故发生丢失，可以事先在进行数据库的备份，在数据丢失之后，以便进行恢复。如果数据服务器要进行更换也可以重新配置数据库的连接。

1. 用例



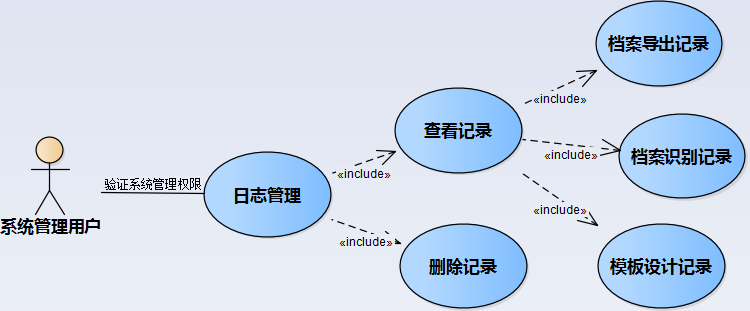
1. 流程



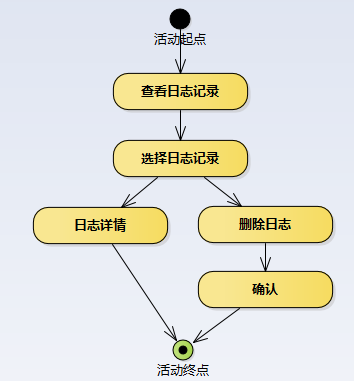
### 日志管理

其主要目的就是跟踪到每一个用户在系统的操作行为，如对数据进行查询、新增、编辑或删除甚至是登录等行为。更进一步的理解可以说是对用户使用系统情况的跟踪，对数据的跟踪防止数据意外删除、更改时有所记录，有所依据，以便对数据的还原，从某种程序上可以保护数据的完整性。

1. 用例



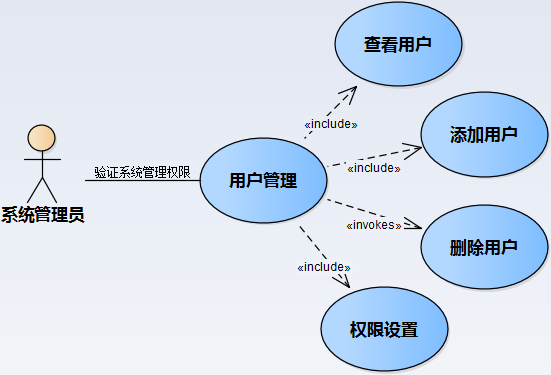
1. 流程



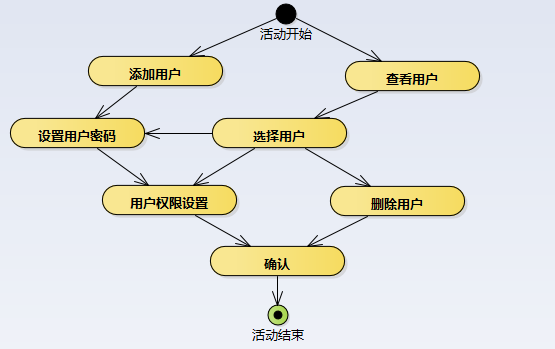
### 用户管理

档案识别涉及多个不同的子系统，每个子系统的职责各不相同，为了整个系统的安全与稳定所以需要统一的用户管理模块，负责不同用户的权限分配。

1. 用例



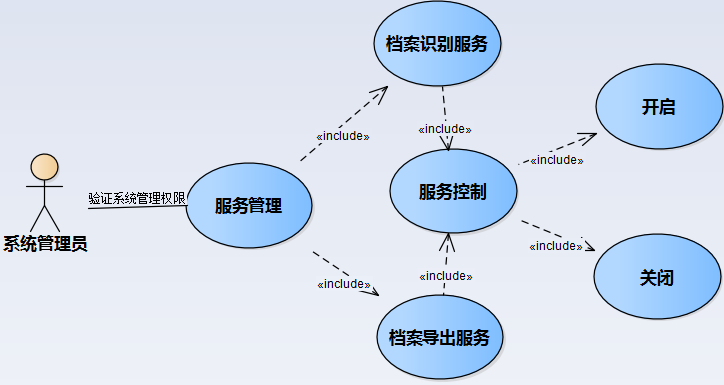
1. 流程



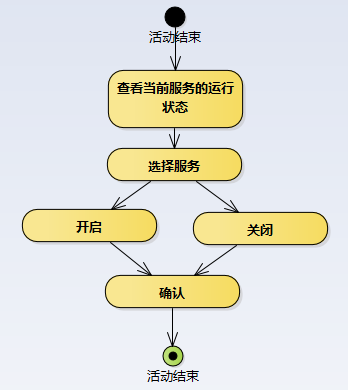
### 服务管理

当系统进行数据库的维护时，为了保证数据的完整性，可以对正在运行的识别服务和导出服务进行关闭，等数据库维护结束时再开启。

1. 用例



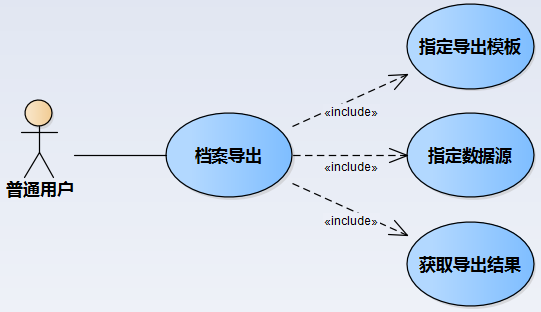
1. 流程



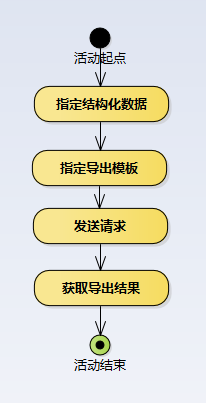
## 档案导出

系统的主要功能之一，将结构化数据以指定的文档排版以电子档的形式进行导出。

1. 用例



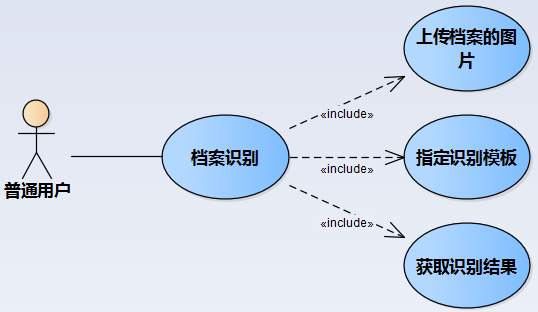
1. 流程



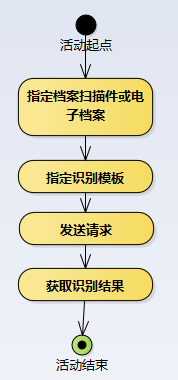
## 档案识别

系统的主要功能之一，对档案的扫描件进行识别，并将识别后的内容转换成结构化数据进行储存。

1. 用例



1. 流程

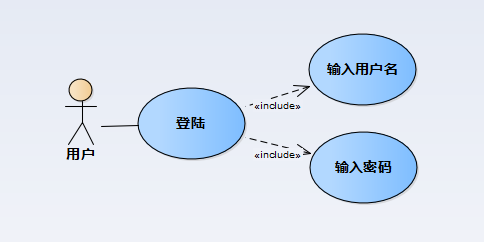


## 识别，导出模板设计

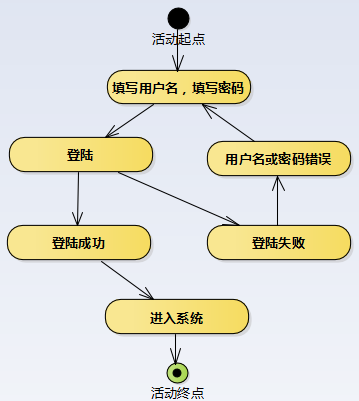
### 系统登陆

进入模板设计时，需要首先验证用户的是否具备管理用户权限，系统登陆模块将完成这项工作。

1. 用例

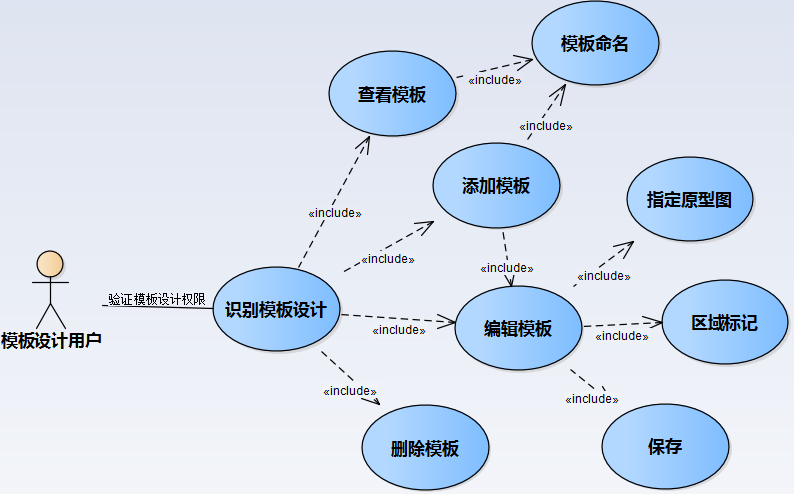


1. 流程

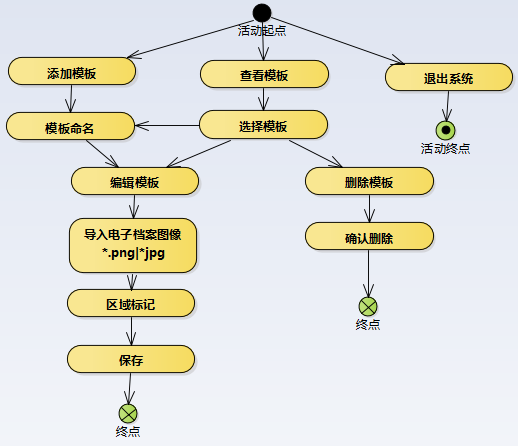


### 识别模板设计

1. 用例

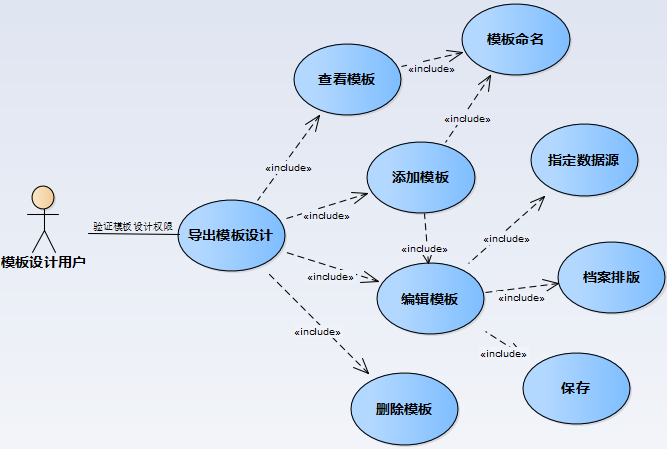


1. 流程

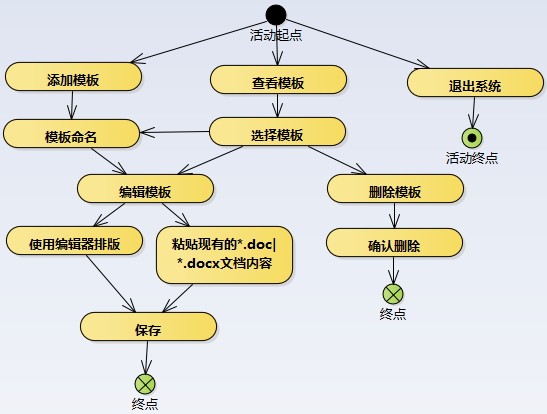


### 导出模板设计

1. 用例



1. 流程

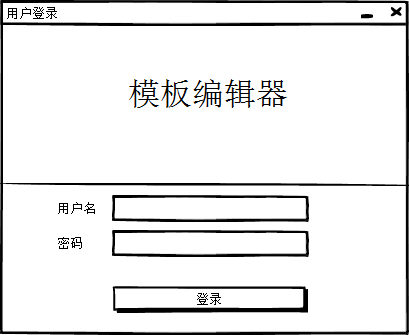


# 交互原型

## 档案识别模板设计

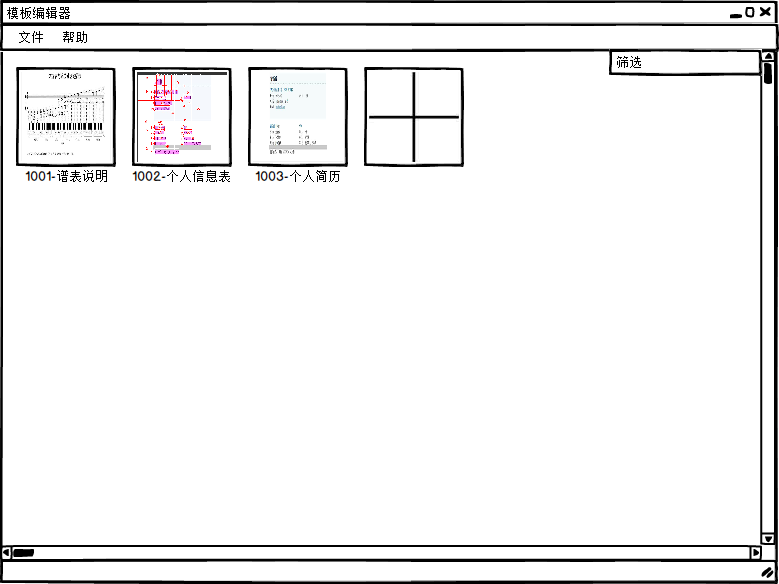
### 识别模板管理-用户-登录

使用模板编辑器需要具备“模板管理”权限用户认证，输入用户名及密码，即可进入管理界面。



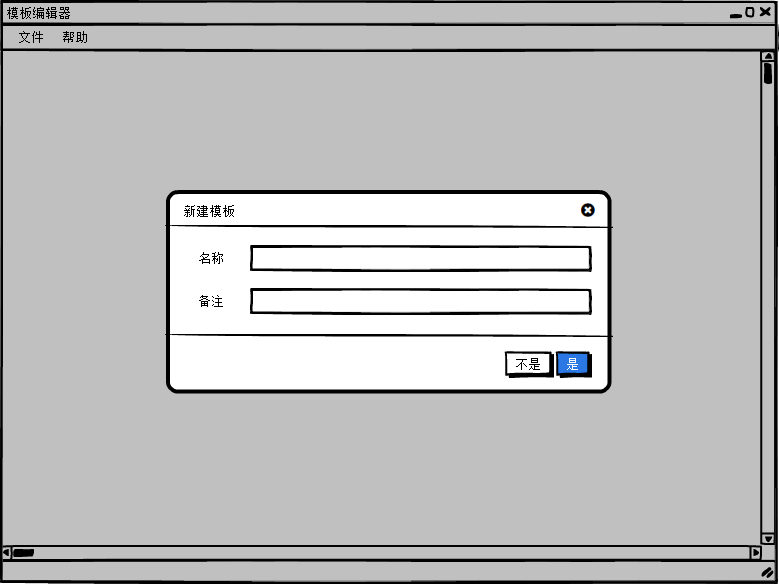
### 识别模板管理

模板管理界面，会展示出当前可供使用的模板，加入筛选关键字，可以搜索出符合规则的模板。点击列表末尾的图标“+”，可添加新的模板。



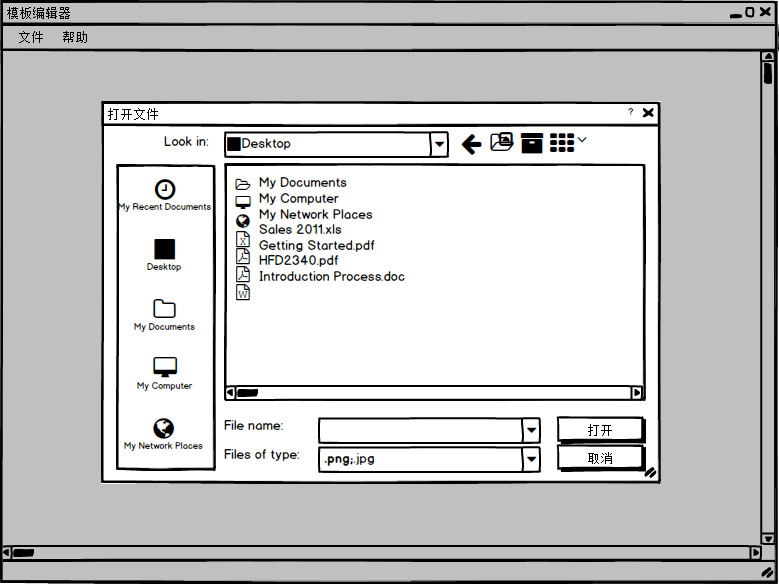
### 识别模板管理-新建-模板信息

新建模板首先要求输入模板的名称，和备注以作标识。



### 识别模板管理-新建-选择档案图像

填写完模板名称和备注后，需要选择一个档案的图像做为模板原型，然后进入编辑界面标识文字提取区域。



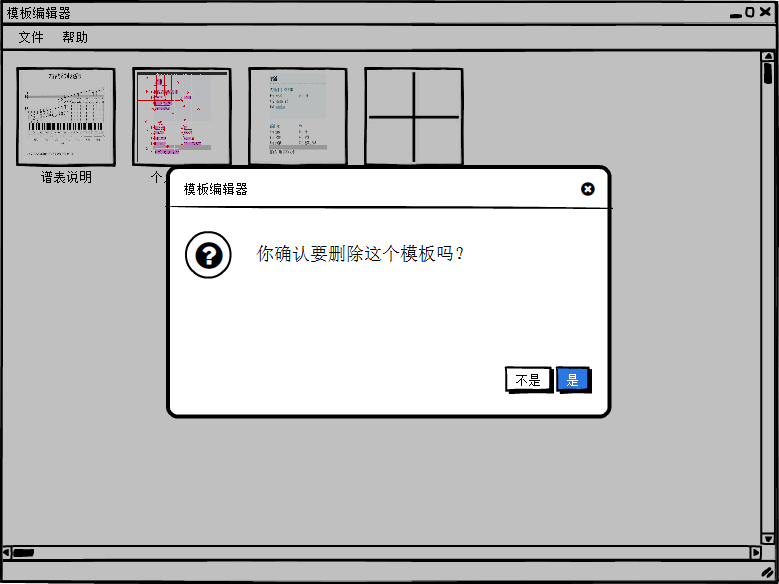
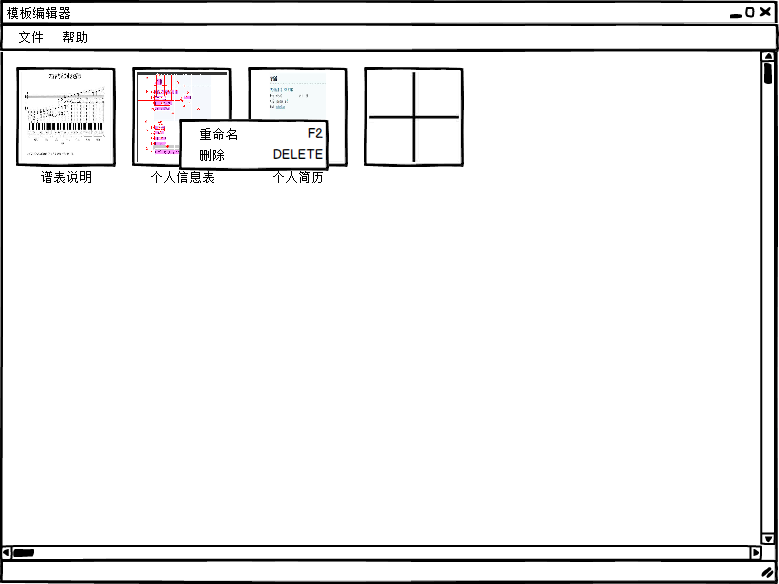
### 识别模板管理-编辑

在管理界面中，通过点击现有的模板图标即可进行模板编辑，使用鼠标对档案图像的指定区域进行框选，即可标识对该区域的文字进行提取，每个区域都存在对应的标识符，用以结构化储存。



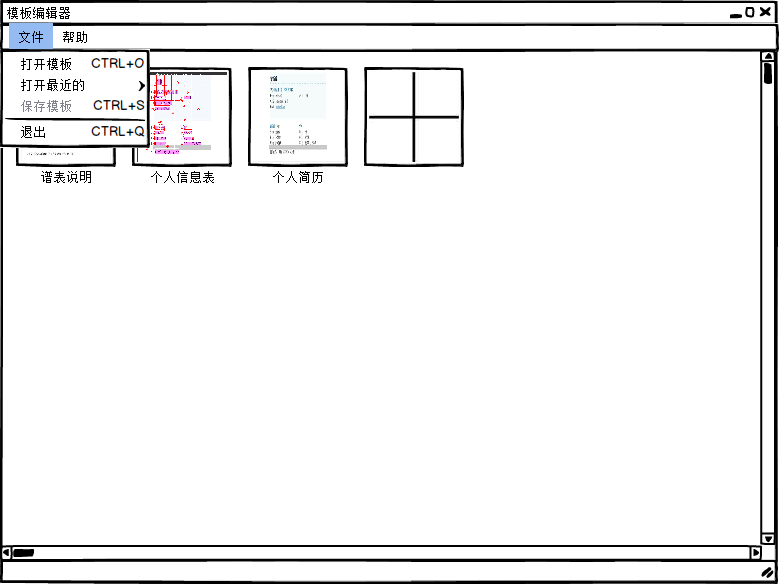
### 识别模板管理-删除

在管理界面中，右键可以对模板进行“重命名”，“删除”等操作。



### 识别模板管理-菜单-文件

用以打开，保存模板文件的菜单。

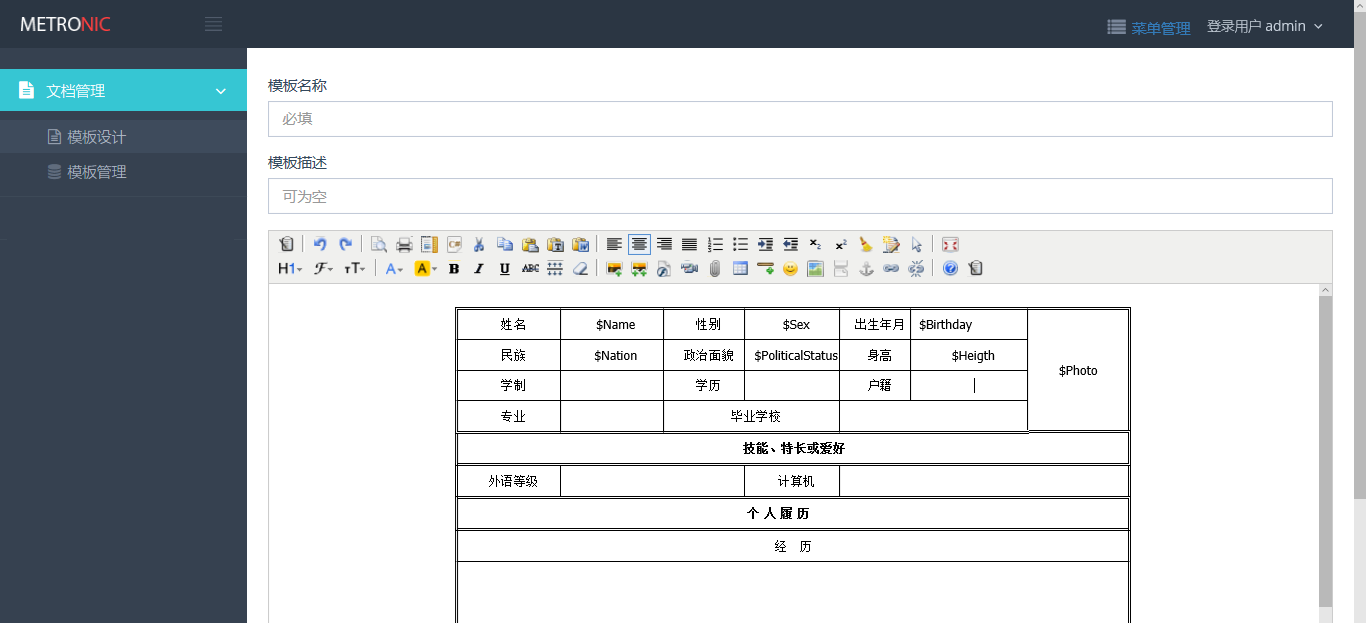


## 档案导出模板设计

### 导出模板-登陆



### 导出模板-设计



### 导出模板-管理



# 其他

## 接口需求

该系统要与原有相关系统保持数据一致，以保证数据的安全性。因此，在设计的时候，需着重考虑本软件和原有相关软件之间的接口，确保能够正常信息交互。

## 非功能和质量需求

除上述需求外，作为需求的补充，需求规格说明还应包括非功能需求，它描述了系统展现给用户的行为和执行的操作等。它包括产品应该遵从的标准、规范和合约； 外部界面的具体细节； 性能要求； 设计或实现的约束条件及质量属性。质量属性是通过多种角度对产品的特点进行描述，从而反映产品功能。该系统需求还应该满足以下非功能特性要求。

## 方便和实用性需求

系统建设要以满足业务需求为首要目标，系统应该方便和实用，采用稳定可靠的成熟技术，保证系统长期安全运行。主要体现为：

1. 界面布局、操作程序、业务逻辑、专业用语等应该符合业务习惯、业务思维规律和业务规程，提高使用工作的质量和效率。
2. 所有交互界面应该按“以人为本”的原则进行设计，交互界面的业务控制逻辑要求“清晰明确，一目了然”，功能按键及其功能按键之间的相互控制和依赖关系，应该由应用软件明确提示和呈现，不得有“隐含晦涩”的标题或提示信息误导用户或让用户不知所措。

## 安全与保密性需求

遵循有关信息安全标准，须具有切实可行的安全保护和保密措施，确保数据永久安全。要从数据存储、传输，用户管理、认证，用户应用模块、数据使用权限分配等方面进行管理，确保业务系统能安全运行，主要表现为：

1. 具有数据整体安全性（防止批量窃取）、私密安全性、统计安全性。防止数据被非法利用和篡改。
2. 应用软件提供支持严格的权限分级管理和授权机制，提供权限管理工具，防止数据非法访问、利用、删改。
3. 在应用软件、数据库、操作系统、网络特性、身份认证等多个方面，综合考虑应用软件的安全要素，确保系统的整体安全性；防止非法用户的入侵；减少合法用户误操作的可能性；保证数据的真实性、完整性、一致性和准确性；对于敏感数据的操作，在应用软件中留有痕迹，提供审计信息，便于追踪和查处。
4. 制定系统运行、维护和管理的规章制度。

## 稳定与可靠性需求

稳定性和可靠性是最重要的质量特性，系统应该在成本可以接受的条件下，从系统结构、设计方案等方面考虑，使得系统故障发生的可能性尽可能少，对各种可能出现的紧急情况有应急的工作方案和对策。具体体现为：

1. 产品本身应该是稳定和可靠的。
2. 在实施和维护过程中，高度重视需求管理和源代码管理，认真考虑需求变化，制订科学、合理和切实可行的修改方案，避免无序修改，减少可能造成的混乱和损失。所有的修改和完善不得以牺牲系统总体稳定性和可靠性为代价。
3. 在初次提交或经修改后提交给银行技术部门或终端用户之前，先进行测试，并提供较为先进和完善的测试工具。根据软件的实际修改情况，（自动或手工）更新或修改测试用例库，重新进行测试。
4. 通过优化达到提高整个系统的稳定性和可靠性。
5. 根据项目实际情况，制定信息突发事件的紧急预案，组织演习。

## 灵活性与可扩展性需求

要应充分考虑到用户增加和业务扩展，有扩充能力及接口。应用软件的模块化程度要高，对不同业务流程和管理方式的适应能力要强，软件维护方便。贯彻面向最终用户的原则，建立友好的用户界面，使操作简单、直观、灵活，易于学习掌握。具体体现在：

1. 系统的架构要求能够满足业务及其相关的事务的流程及模式变更的需求。
2. 主张应用基于组件编程、领域工程、产生式程序设计等理论和技术进行系统的研发和改造，以增加系统架构的灵活性和稳定性，同时不以牺牲系统性能和可靠性为代价。
3. 应用系统在上线、扩充或升级时，须提供便捷的数据转换工具，将原有数据（包括历史数据）转换到新的结构中，以保护数据资源和确保新旧系统间的平稳过渡。
4. 应用系统在用户界面的灵活性上，需要达到因人而异的程度。
5. 应用软件的模块化程度要高，对不同业务流程和管理方式的适应能力要强，软件维护方便。

## 界面友好需求

1. 交互界面应该按“以人为本”的原则进行设计，要求有足够的“在线帮助”、“日常工作导航”、“常见问题处理指南”等帮助文档。
2. 所有交互界面的业务控制逻辑要求“清晰明确，一目了然”，功能按键及其功能按键之间的相互控制和依赖关系，应该由应用软件明确提示和呈现，不得有“隐含晦涩”的假设，不加提示，误导用户或让用户不知所措。
3. 操作界面的使用和维护界面应该友好、便捷，符合当前流行的模式和习惯。对敏感操作或关键数据，醒目辅以视觉提示，以减少误操作几率。系统应避免冗余数据和操作，并充分利用数据之间的约束关系自动校验和采集数据。为安全起见，不得随意提供数据缺省值。