

自然语言处理平台

用户手册

第1版

版本控制信息

| 版本 | 日期 | 新增/修改 | 说明 | 撰写人 |
| --- | --- | --- | --- | --- |
| V1.0 | 2020/02/23 | 新增 | 创建 | 王成军 |
| V1.1 | 2020/04/09 | 修改 | 升级系统说明 | 顾嘉晟 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

目 录

[一 简介 1](#_Toc85821179)

[二 注册登录 1](#_Toc85821180)

[2.1 注册 1](#_Toc85821181)

[2.2 登录 1](#_Toc85821182)

[2.3 登录问题 2](#_Toc85821183)

[2.3.1 忘记密码 2](#_Toc85821184)

[2.3.2 退出系统 2](#_Toc85821185)

[三 管理大厅 2](#_Toc85821186)

[3.1 项目统计 2](#_Toc85821187)

[3.2 项目列表 3](#_Toc85821188)

[四 项目管理 3](#_Toc85821189)

[4.1 新建项目 3](#_Toc85821190)

[4.1.1 分类项目 3](#_Toc85821191)

[4.1.2 抽取项目 4](#_Toc85821192)

[4.1.3 分词项目 4](#_Toc85821193)

[4.1.4 实体关系项目 5](#_Toc85821194)

[4.2 修改项目信息 5](#_Toc85821195)

[4.3 删除项目 6](#_Toc85821196)

[4.4 收藏项目 6](#_Toc85821197)

[五 文档标注 6](#_Toc85821198)

[5.1 标注管理 7](#_Toc85821199)

[5.2新建标注任务 7](#_Toc85821200)

[5.3查看标注任务 8](#_Toc85821201)

[5.4 标注 9](#_Toc85821202)

[5.4.1 分类标注 9](#_Toc85821203)

[5.4.2 抽取标注 10](#_Toc85821204)

[5.4.3 分词标注 11](#_Toc85821205)

[5.4.4 实体关系标注 13](#_Toc85821206)

[5.5 标注数据导出 14](#_Toc85821207)

[六 模型管理 15](#_Toc85821208)

[6.1 新建模型 15](#_Toc85821209)

[6.2 模型训练 15](#_Toc85821210)

[6.2.1 分类模型训练 15](#_Toc85821211)

[6.2.2 抽取模型训练 16](#_Toc85821212)

[6.2.3 分词模型训练 18](#_Toc85821213)

[6.2.4 实体关系模型训练 18](#_Toc85821214)

[6.3 模型上线 19](#_Toc85821215)

[6.4 模型删除 20](#_Toc85821216)

[6.5 模型评估 20](#_Toc85821217)

[七 模型预测 22](#_Toc85821218)

[7.1 上传预测文件 23](#_Toc85821219)

[7.2 查看预测结果 23](#_Toc85821220)

[7.3 删除文档记录 24](#_Toc85821221)

[7.4 预测详情页 24](#_Toc85821222)

[7.5 预测结果导出 24](#_Toc85821223)

[7.6 预测失败常见原因 25](#_Toc85821224)

[八 管理中心 25](#_Toc85821225)

[8.1 用户管理 25](#_Toc85821226)

[8.1.1 新增用户 25](#_Toc85821227)

[8.1.2 编辑用户信息 26](#_Toc85821228)

[8.1.3 删除用户 26](#_Toc85821229)

[8.2 语料管理 26](#_Toc85821230)

[8.3 导出管理 26](#_Toc85821231)

[8.4 比对规则管理 27](#_Toc85821232)

[九 自定义模型 28](#_Toc85821233)

[附录1 (标注指南) 29](#_Toc85821234)

[附录2 (新手指南) 31](#_Toc85821235)

[附录3 (用户常见问题) 32](#_Toc85821236)

# 一 简介

自然语言处理平台是一款集【项目管理】、【语料标注】、【模型管理】、【模型训练】、【模型评估】、【模型预测】等多功能于一体的系统，可大幅度提高对各种文档处理和管理效率的系统。该系统共有【管理大厅】、【文本分类】、【文本抽取】、【分词】、【实体关系】、【通用工具】、【管理中心】等模块。

# 二 注册登录

## 2.1 注册

系统使用非开放式账户注册，用户可以通过SSO账号验证登录或者超级管理员用户可以在“权限管理”页面创建新用户。

## 2.2 登录

被创建的用户可以通过输入账号、密码来登录系统(如下图)。请按以下步骤登录系统 。

1）访问：<http://nlp-platform-preview.datagrand.com/>

2）输入您的账户名、密码；

3）勾选记住用户名，将会把当前用户名、密码保存下来。（登出账户后，可以方便再次免输用户名密码，重新登录）；

4）点击登录按钮；



## 2.3 登录问题

### 2.3.1 忘记密码

按照用户的角色不同找回密码的方式也不同：

1）普通用户：向管理员用户寻求帮助找回密码

2）管理员：向项目维护人员寻求帮助找回密码

### 2.3.2 退出系统

点击下图的【登出】按钮后页面跳转至登录页面，用户退出系统。



# 三 管理大厅

## 3.1 项目统计

在管理大厅的顶部可以看到分类项目、抽取项目、实体关系项目、分词项目的项目数量、标注任务数、模型数量、已标注百分比、已审核百分比的统计。



## 3.2 项目列表

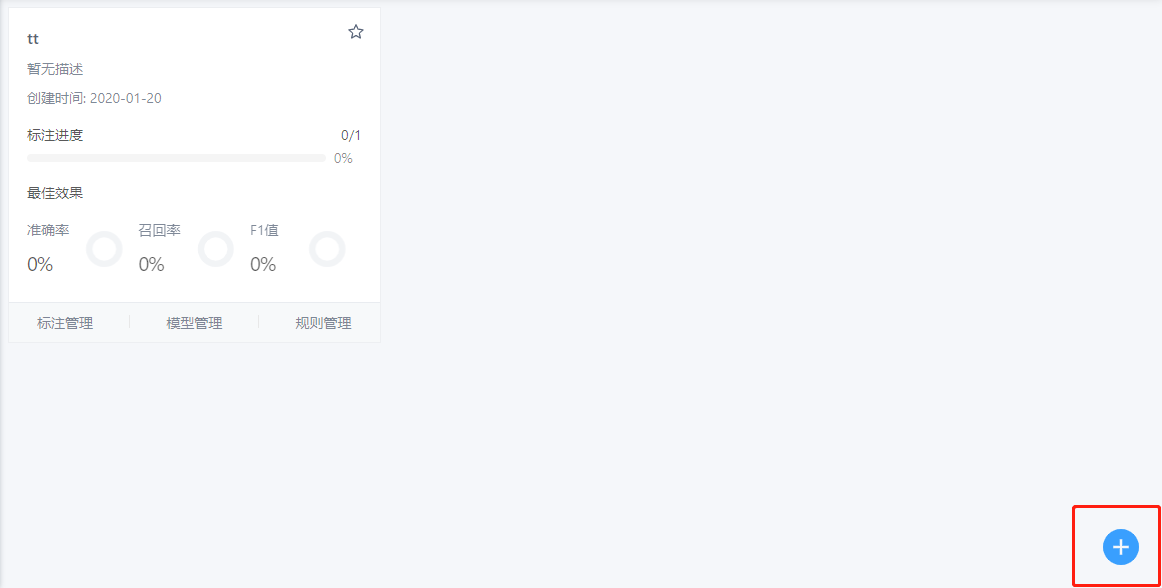
在管理大厅的底部可以看到分类项目、抽取项目、实体关系项目、分词项目的详细项目情况列表。



# 四 项目管理

## 4.1 新建项目

在文本分类、文本抽取、分词、实体关系各主菜单下的【项目管理】页面点击添加按钮新建项目。



### 4.1.1 分类项目

新建分类项目时输入项目名称、备注说明、分类标签。



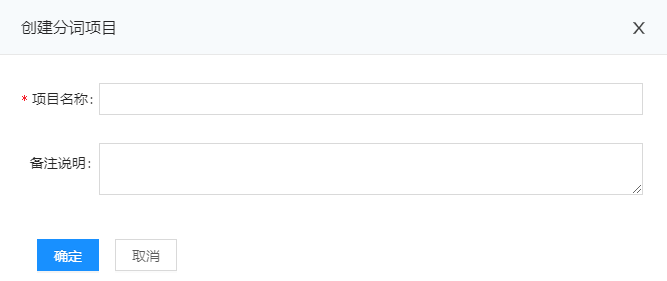
### 4.1.2 抽取项目

新建抽取项目时输入文档类型名称、备注说明、生效条款。



### 4.1.3 分词项目

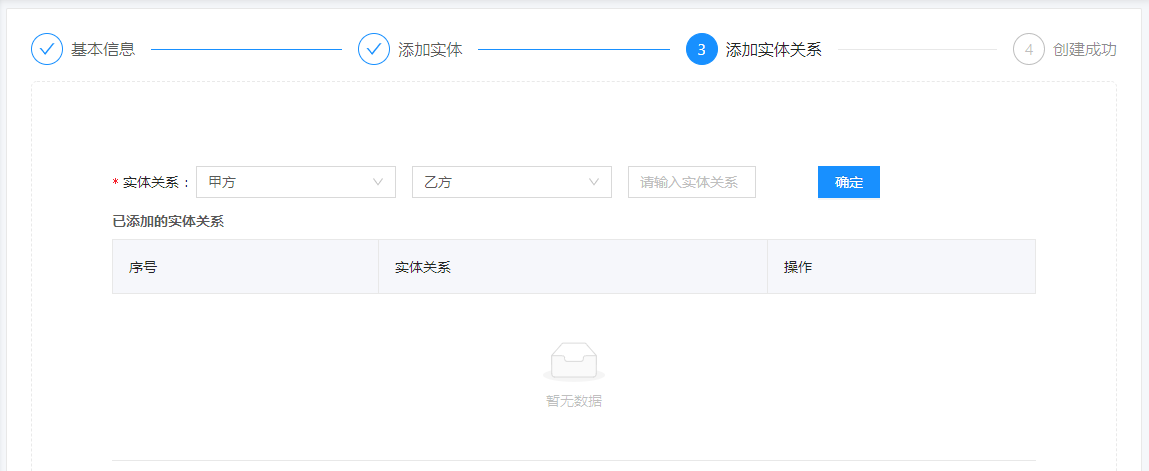
新建分词项目时输入项目名称、备注说明。



### 4.1.4 实体关系项目

新建实体关系项目时添加项目基本信息—添加实体—添加实体关系





## 4.2 修改项目信息

在项目信息卡上点击【修改】按钮进行修改。



## 4.3 删除项目

在项目信息卡上点击【删除】按钮进行删除。

## 4.4 收藏项目

在项目信息卡上点击【收藏】按钮进行收藏，收藏项目置顶显示。

# 五 文档标注

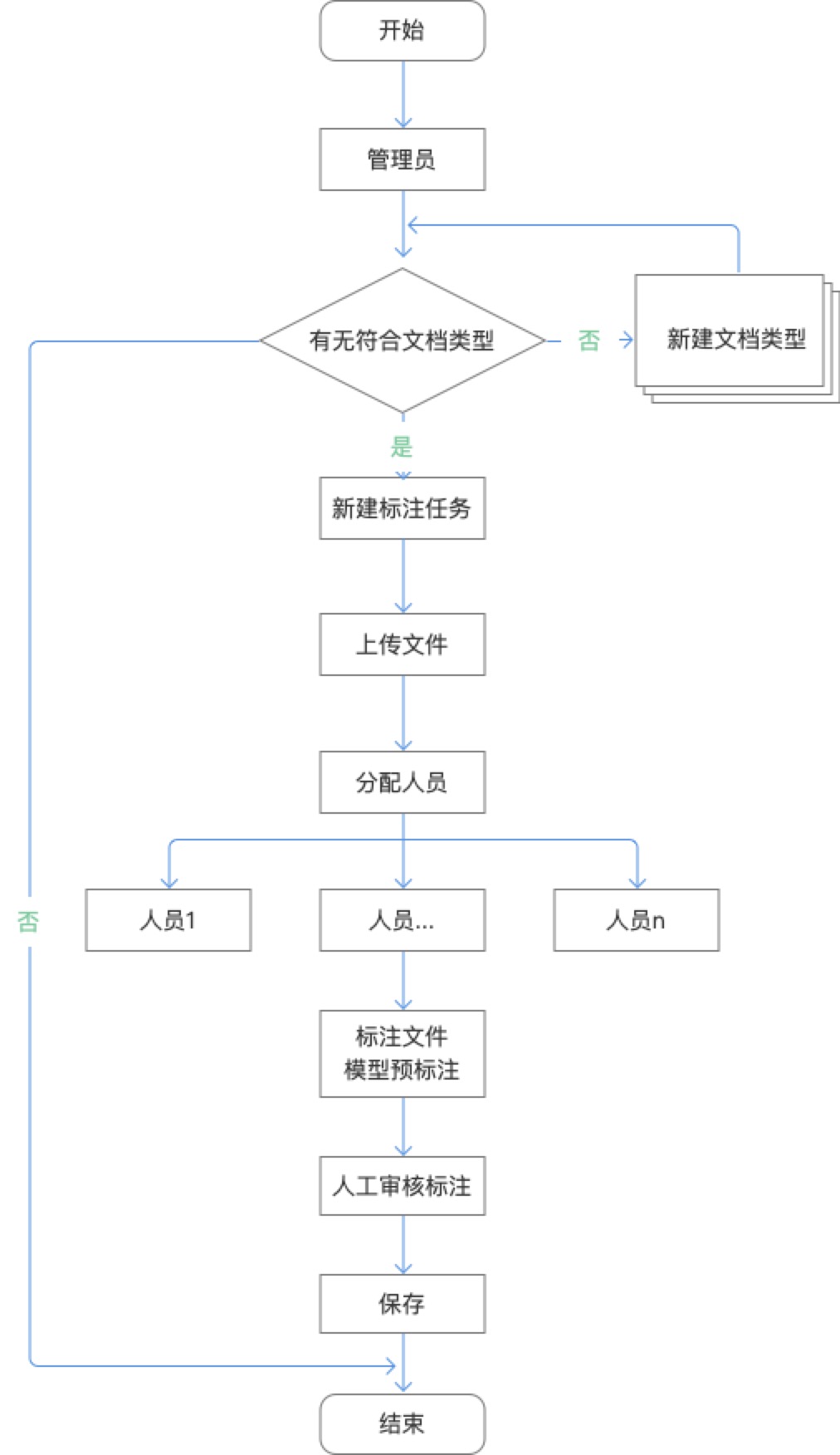
文档标注即根据用户事先定义好的特定信息对需要标注的文档进行标注。被标注的文档将作为模型训练的语料，文档标注的越精确、数据越多训练出来的模型精度也就越高。

文档标注模块分为管理员、审核员和标注员三种角色，系统会根据角色的不同而分配给用户不同的权限。

管理员：可以创建文档类型、新建标注任务、将任务分配给不同的标注员、审核员、标注文件；

审核员：仅可以审核分配给自己的标注文件。

标注员：仅可以标注分配给自己的标注文件。

步骤：项目管理—标注管理—新建标注任务&上传文件—查看任务文件—查看文件进入标注详情—进行标注—保存

## 5.1 标注管理

标注管理模块的功能是让用户为标注的文件创建任务，以方便用户管理不同文档类型的标注文件。



## 5.2新建标注任务

进入“标注任务管理”菜单界面，点击“新建标注任务”，进入新建弹框列表下。

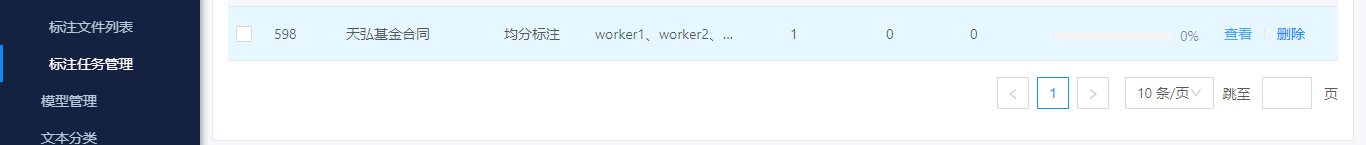
在“新建标注任务”列表下文本框内输入相应的值，上传需要标注的文档，点击“确定”,在页面上方弹出“创建成功”提示语即可。

\* 必填项不能为空，必须输入正确的内容。



## 5.3查看标注任务

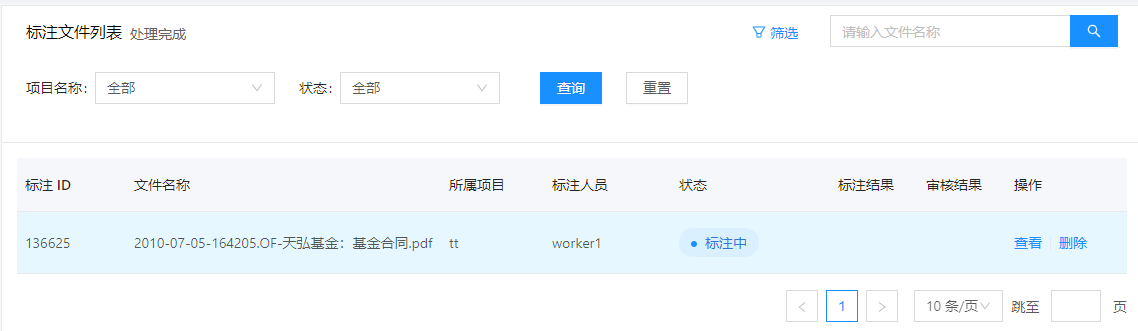
创建一条新的标注任务成功后，点击页面“查看”功能，页面跳转【标注列表】界面；





标注任务管理页面，点击“筛选”功能，会展开隐藏的筛选条件，可根据项目名称进行筛选查看想要了解的任务列表；也可以根据右上角的搜索功能，查看相关‘任务名称’内容。

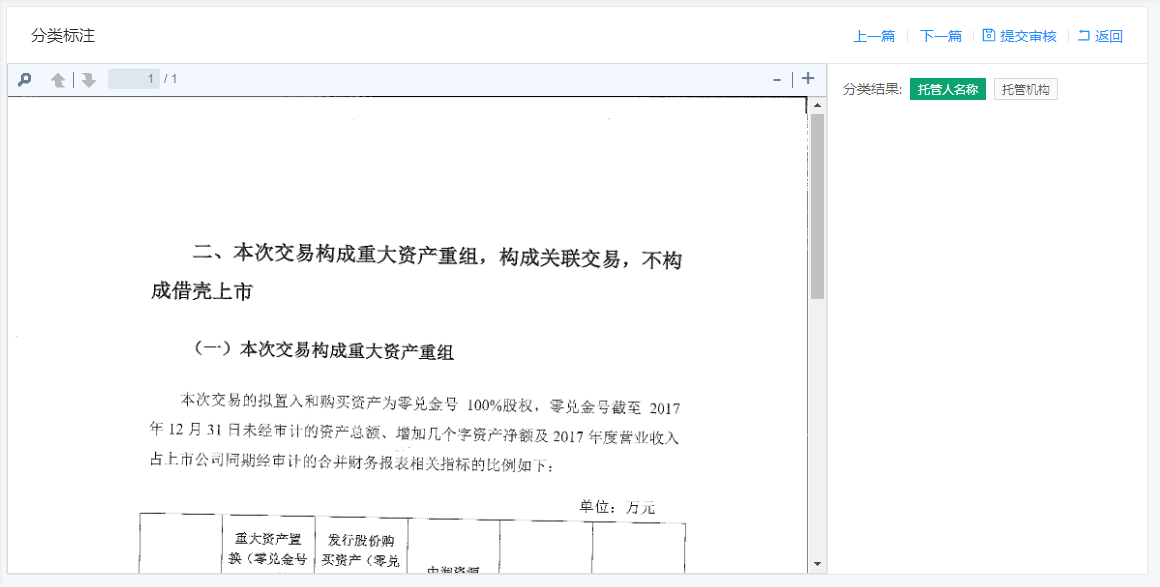
标注文件列表页面，点击“筛选”功能，会展开隐藏的筛选条件，可根据项目名称、状态进行筛选查看想要了解的任务列表；也可以根据右上角的搜索功能，查看相关‘文件名称’内容。



## 5.4 标注

### 5.4.1 分类标注

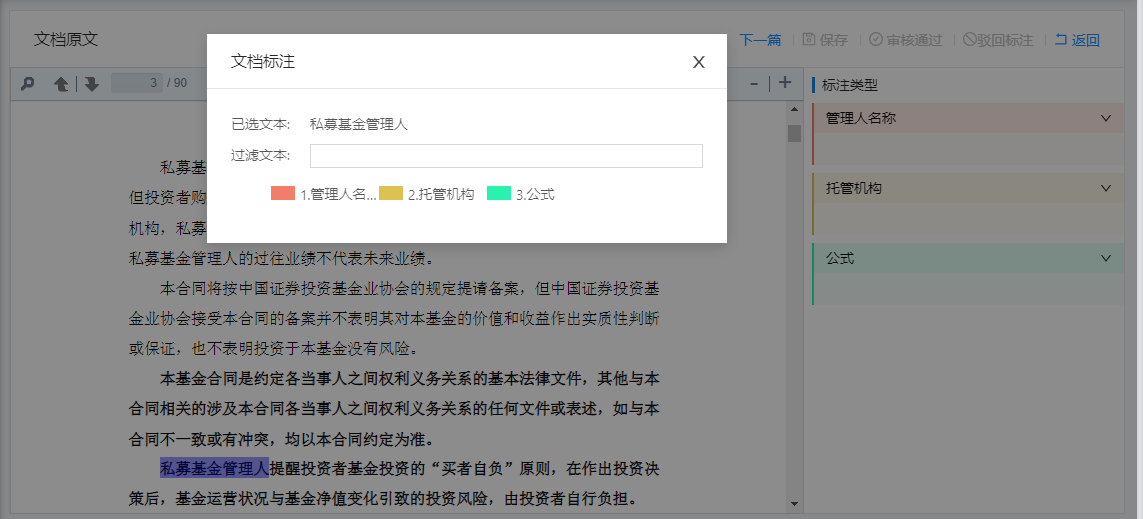
在标注详情页面，模式选择默认是点选，用户用鼠标点击“分类结果”的分类项，被选取的分类则会显示亮色，最后点击右上角的【提交审核】或【下一篇】按钮，被标注的内容则会被保存下来。



### 5.4.2 抽取标注

1.标注功能:

在标注详情页面，默认是划选，用户用鼠标划选好需要标注的内容后，松开鼠标页面会弹出可供选择标注条款，如果可供选择的条款很多，用户可以通过“过滤文本”来筛选，原文中的被选取的内容则会显示与条款颜色相同高亮状态，最后用户点击右上角的【保存】按钮被标注的内容则会被保存下来。



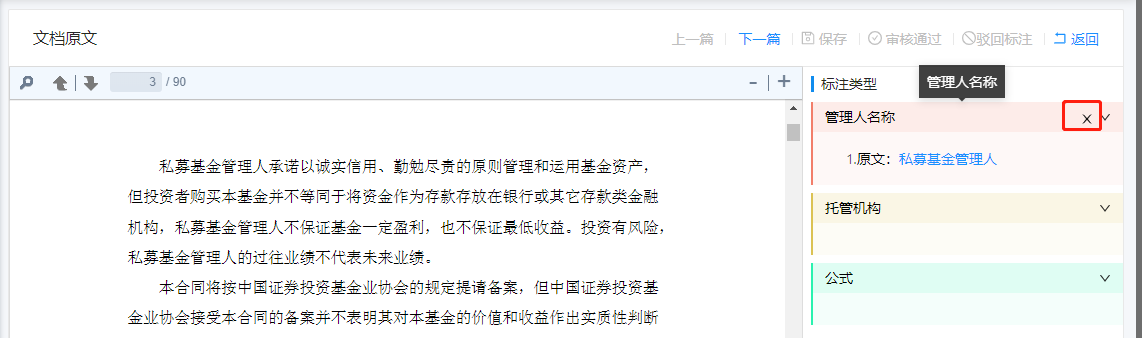
注：为保证每次标注的数据不易遗漏，请常点击保存按钮，以防出现丢失数据现象，造成一定的损失。

2.取消标注：

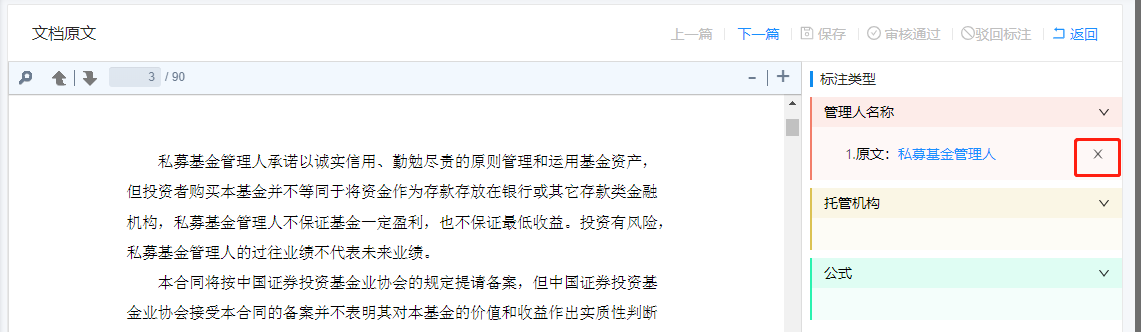
取消（删除）标注有3种方式：

第一种：在文档原文中，双击需要取消标注的内容即可取消标注。对应的标注列表内容也会随即删除。

第二种：在标注类型 - 针对每个条款类型进行删除，点击“X”，会将当前条款类型下所有标注内容删除；同时原文中内容高亮部分也会消失。



第三种：在标注类型每个条款类型内部，逐条删除，点击每条内容后的“X”，将单条删除，同时原文中内容高亮部分也会消失。



3.重复标注：

支持重复标注，可分两种情况：

第一种：同一位置重复标注相同的标注类型；

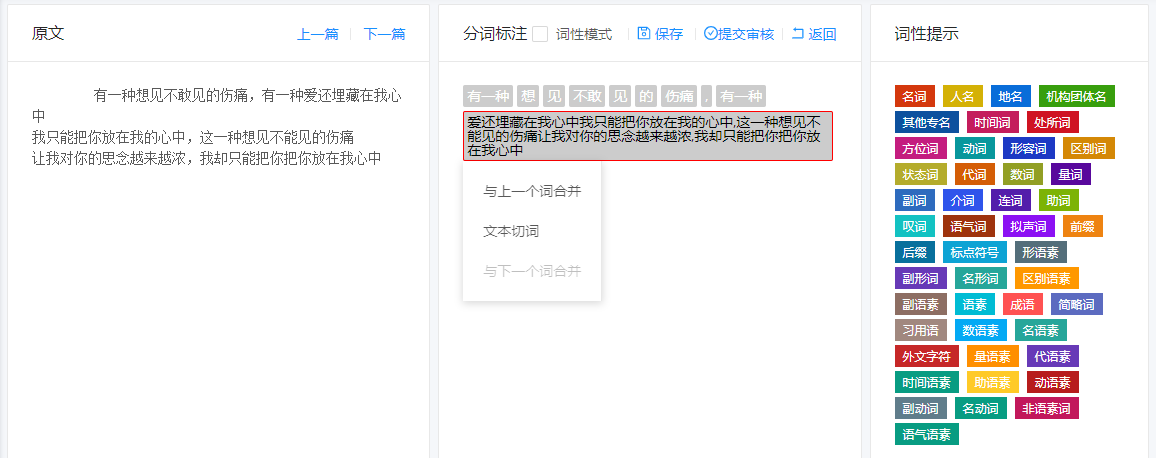
第二种：同一位置重复标注不同的标注类型；

4.标注内容高亮定位：

每次文档内容标注成功后，点击右侧标注类型的条款内容， 即可定位到原文标注内容位置，并高亮显示。

### 5.4.3 分词标注

在标注详情页面有三分屏，左侧屏为原文、中间屏为标注、右侧屏为词性提示，标注分为两种模式【分词标注】和【词性模式】



【分词模式】

在中间屏点击分词的内容，弹出提示框菜单“与上一个词合并”、“文本切词”、“与下一个词合并”，点击“文本切词”进行分词模式，在内容框中通过换行ENTER键进行文本切分。



【词性模式】

在中间屏点击勾选“词性模式”，单击已经切好的分词，弹出词性选择框，点击相应的词性即可标注该词的词性。



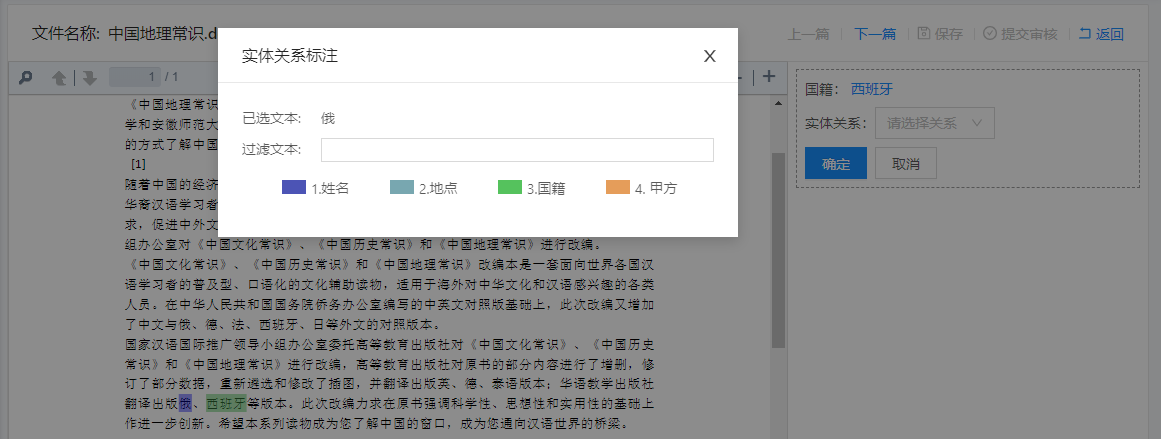
最后点击右上角的【保存】、【提交审核】或【下一篇】按钮，被标注的内容则会被保存下来。

### 5.4.4 实体关系标注

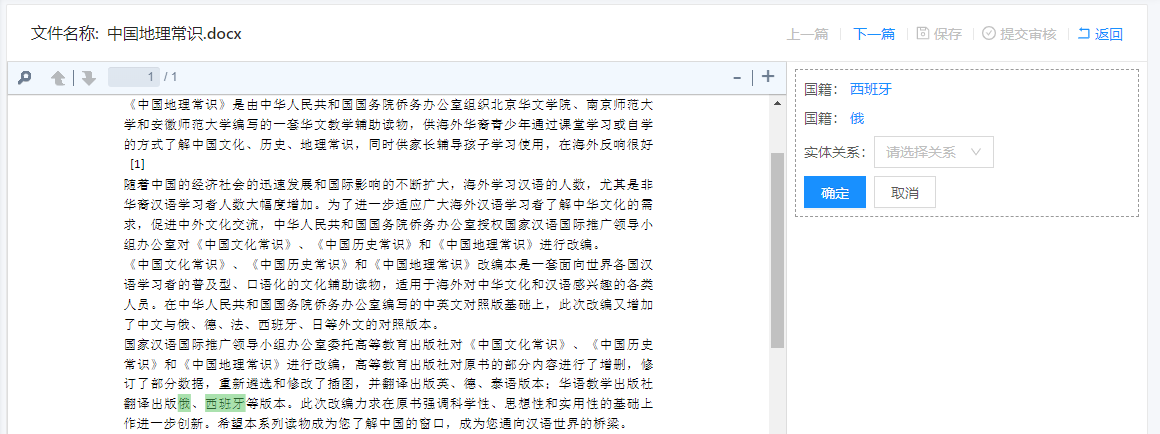
实体关系标注需先标注实体一，再标注实体二，最后标注实体关系。

标注实体实体一：在标注详情页面，默认是划选，用户用鼠标划选好需要标注的内容后，松开鼠标页面会弹出可供选择标注条款，如果可供选择的条款很多，用户可以通过“过滤文本”来筛选，原文中的被选取的内容则会显示与条款颜色相同高亮状态。

标注实体二：标注完实体一后，默认进入实体二标注，标注方式与标注实体一方式相同。



标注实体关系：标注实体二完成后，默认进入实体关系标注，在实体关系下拉框选择实体关系。



## 5.5 标注数据导出

步骤一：从分类、抽取、分词、实体关系各菜单模块中进点击项目管理，进入项目管理页，在项目卡片上点击“标注管理”进入标注管理页。



步骤二、在标注任务管理页勾选需要导出的标注任务，在上方栏的“导出标注”按钮高亮，点击后即可导出并下载标注数据。



# 六 模型管理

模型管理页面是对所有完成模型训练任务的一个列表统筹管理，页面包含新建模型、上线、基本增删改查功能，以及模型评估工作。

## 6.1 新建模型

新建模型功能放在算法模型管理页面，点击侧边栏【模型管理】进入列表页面，点击【新建模型】功能，进行新建操作。



## 6.2 模型训练

### 6.2.1 分类模型训练

基本步骤【基本信息】—【训练数据】—【确认配置】—【创建成功】，基本信息需输入模型名称、文档类型（项目名）、算法类型（内置逻辑回归、SVM、LSTM等算法，也可选择已注册的自定义算法）、算法参数、模型描述。



### 6.2.2 抽取模型训练

模型训练会分为四个步骤：

步骤一：基本信息

模型名称（必填）； 文档类型（必选）；训练类型（必选）；模型描述（选填），填写相应的模型信息后，点击下一步即可进入选择数据页面；



步骤二： 选择训练数据

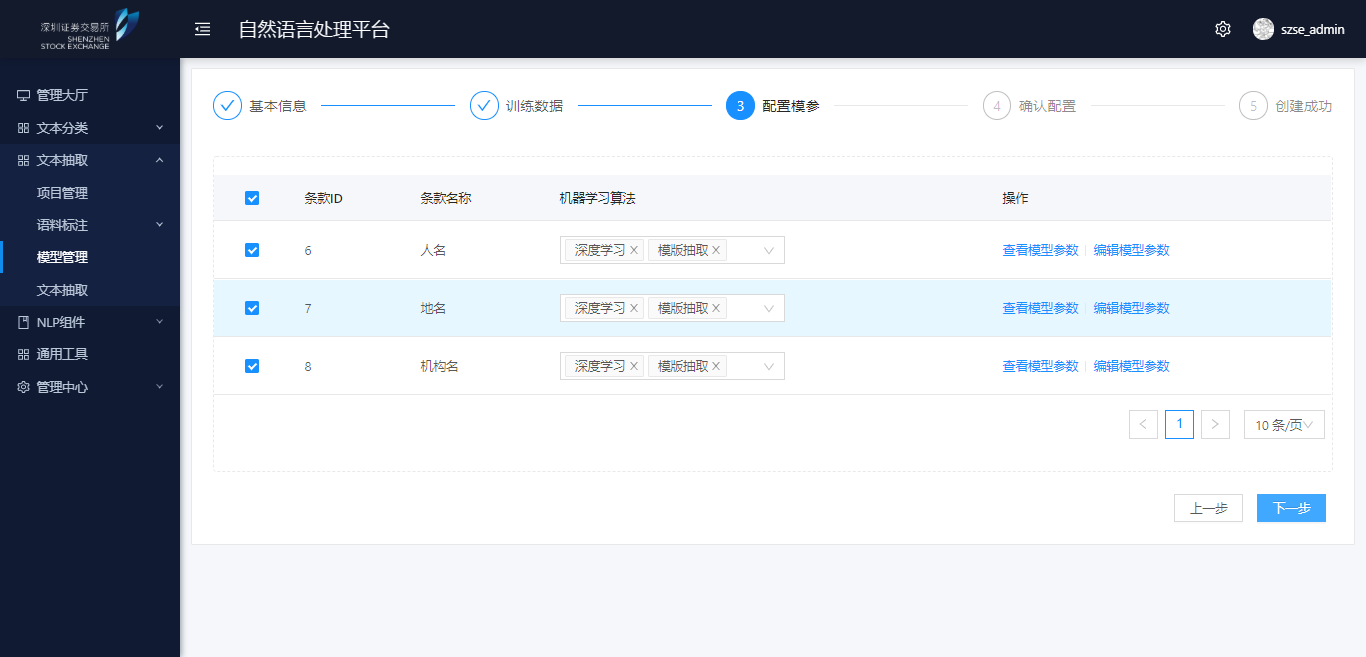
训练数据的来源是该文档类型下对应的不同任务（可通过任务搜索框，快速方便的检索出某一任务），选择某些任务后，在普通用户下，点击下一步即可进入【配置模参】信息页面；

注意: 只有当该任务被标注且审核时才可被选择；

步骤三： 配置模型参数

页面主要功能是对每个文档类型条款采用深度学习、模板抽取两种不同的机器学习算法进行训练；

可对配置模参机器学习算法进行【编辑模型参数】和【查看模型参数】。



步骤四： 确认配置

该步骤是将步骤一、步骤二填写或选择的数据进行确认，若出现错误，点击上一步进行修改，若无误，点击提交，进入创建成功页面。



步骤五：创建成功

当你看到该页面时，意味着你已经新建了一个模型，点击【返回模型管理界面】即可看到新建的模型。

模型状态分为训练成功、训练失败、上线成功。

训练成功： 后台训练成功，生成模型，只有训练成功的模型才可以上线成功；

训练失败： 训练模型失败；

上线成功： 点击列表右侧“操作 - 上线”，即当前模型训练任务正式上线成功，、该模型可用于标注（预标注）、审核、预测等模块中；

### 6.2.3 分词模型训练

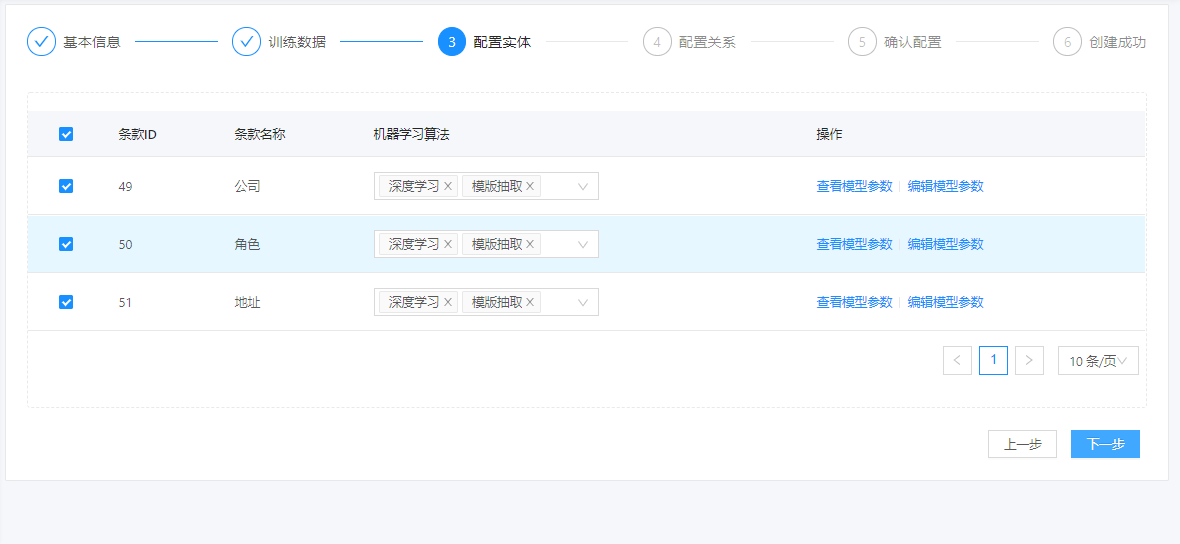
基本步骤【基本信息】—【训练数据】—【确认配置】—【创建成功】，基本信息需输入模型名称、文档类型、算法类型（内置BERT深度学习算法，也可选择自定义算法）、算法参数、模型描述。



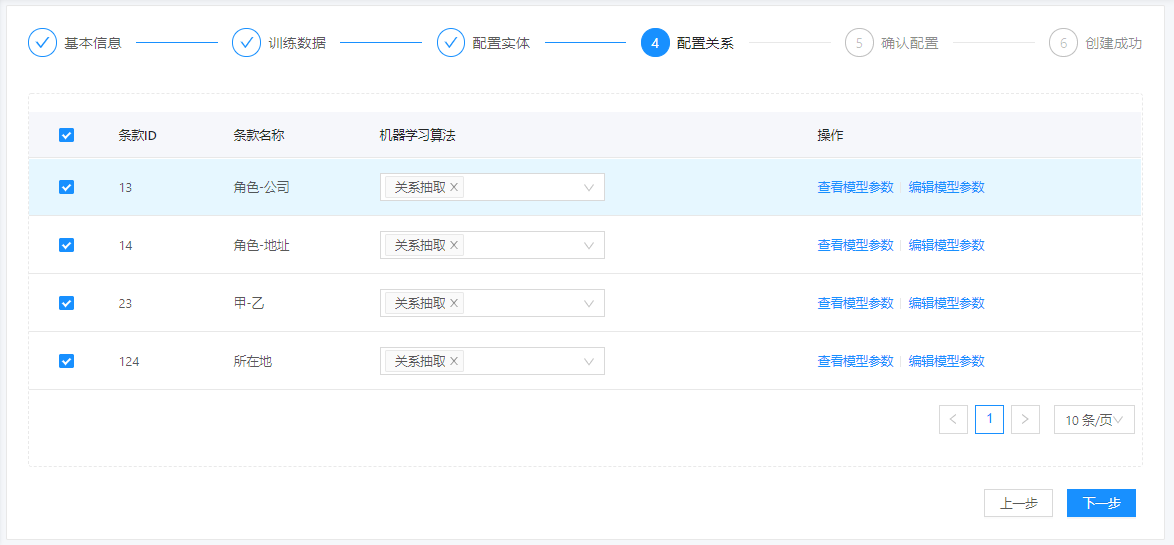
### 6.2.4 实体关系模型训练

基本步骤【基本信息】—【训练数据】—【配置实体】—【配置关系】—【确认配置】—【创建成功】。

【配置实体】页面主要功能是对每个文档类型条款采用深度学习、模板抽取两种不同的机器学习算法进行训练；可对配置模参机器学习算法进行【编辑模型参数】和【查看模型参数】。



【配置关系】可对机器学习算法进行【编辑模型参数】和【查看模型参数】。



## 6.3 模型上线

进入【模型管理】界面，选择需要上线的模型，点击【上线】按钮，即可上线该模型。列表中“上线”灰掉，状态为上线成功，说明当前这条模型训练任务上线成功。



## 6.4 模型删除

进入【模型管理】界面，选择需要删除的模型，点击【删除】按钮，会出现模型删除的确认框，点击【确定】即可删除该模型。目前只支持状态为「训练成功」「已上线」「训练失败」的情况下，才可操作删除，否则删除会灰色显示，不可操作。

## 6.5 模型评估

评估：即评估模型的效果好坏，该指标具有一定的指导意义。

简单来说，评估模型就是创建一个任务，该任务内的文件已标注或审核（且该任务并不是模型训练的数据），系统会调用模型来对该任务下的文件进行预测，并将预测的内容与标注的内容进行指标计算，计算出模型的准确率、召回率、F1值。以下以抽取模型为例详述评估步骤：

进入【模型管理】界面，选择需要评估的模型，点击【评估】按钮。



点击【评估】按钮后，会进入到【模型评估】页面，输入评估任务名称和评估任务描述值，点击“下一步”，进入选择评估数据页面；



该页面会显示出该文档类型对应的任务，选择需要评估的任务，注意只有当该任务被标注且被审核时才可被选择；

 当选择需要评估的任务后，点击“下一步”，即可进入【选择评估字段】界面；勾选需要选择评估的字段类型，点击“下一步”，即可进入【确认并提交】界面；



确认并提交界面，显示的是评估模型信息的基本内容，点击“提交”，将跳转【创建成功】界面，代表评估内容不可修改，如果需要修改，需要重新创建评估任务。

进入【创建成功】界面,即模型评估正在进行中。

点击“模型管理-更多：查看评估结果”功能，进入评估结果列表，即“评估”操作后的评估结果页面；



评估详情页面包含两个部分内容，依次为：

模型信息： 包含文档类型、模型名称、训练时间等模型信息；

整体评估结果：对所有的字段进行整体的计算，包含三个量化的指标，准确率、召回率和F1值（分类项目还具有ROC曲线、混淆矩阵）；



# 七 模型预测

各项目中自定义和内置的算法模型经标注、训练、上线后，可在文本分类、文本抽取、分词、实体关系模块下点击相应的“文本分类”、“文本抽取”、“文本分词”、“关系抽取”按钮进行模型的预测。

模型预测依赖于模型训练，模型训练又依赖于文档标注。因此数据标注的越精确、标注数据量越大，模型精度越高，模型精度越高文档预测的结果也就越好。用户在预测文件之前请务必保证该文档类型的模型已经上线且模型评估的效果较好，否则可能会导致预测结果不理想。

以下以抽取预测为例进行说明。

## 7.1 上传预测文件

在“文本抽取任务页面”右侧打开【新建文本抽取任务】抽屉页面，在当前页面进行文件上传；

任务名称自定义；

项目名称选择在项目管理中建立的项目（项目对应训练的模型）；

可点击“文件上传”控件上传文件，目前文档抽取功能已经支持.pdf、.doc、.docx、.txt格式的文档。

点击确定，页面默认跳转当前文档文档类型的历史记录页面；

点击取消，关闭当前窗口，回到内置文档类型页面；



## 7.2 查看预测结果

在“文本抽取任务页面”选择需要查看的预测任务点击查看进入“文本抽取文件列表”页面，显示所有文档类型的抽取结果信息。

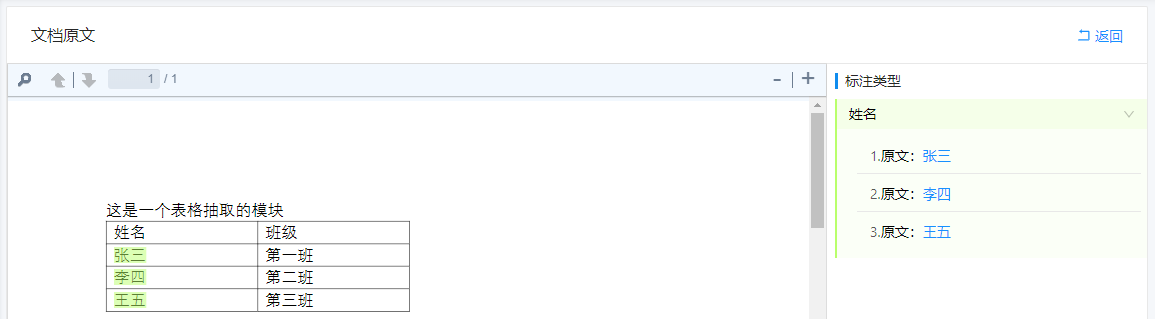


## 7.3 删除文档记录

在文档列表中，选择需要删除的文件，点击【删除】按钮，即可删除文档记录。

## 7.4 预测详情页

抽取详情页面展示了文档抽取的内容，左侧面板为带有标注内容的文档原文，右侧面板为抽取的条款列表，顶部为操作按钮，底部为被抽取的基本信息。



抽取详情页面的左侧为为文档原文，对文档的操作包括： 搜索、翻页/跳转指定页、放大/缩小文件。

## 7.5 预测结果导出

预测成功在预测任务管理页面点击“导出”按钮可导出下载预测结果，有失败或未完成的任务时导出失败。



## 7.6 预测失败常见原因

文档没有抽取任何内容的原因有以下几种：

* 选择的文档类型有误；
* 用于模型训练的标注数据质量不高；
* 用于模型训练的标注数据量太少；
* 已上线的该文档类型的模型效果比较差；

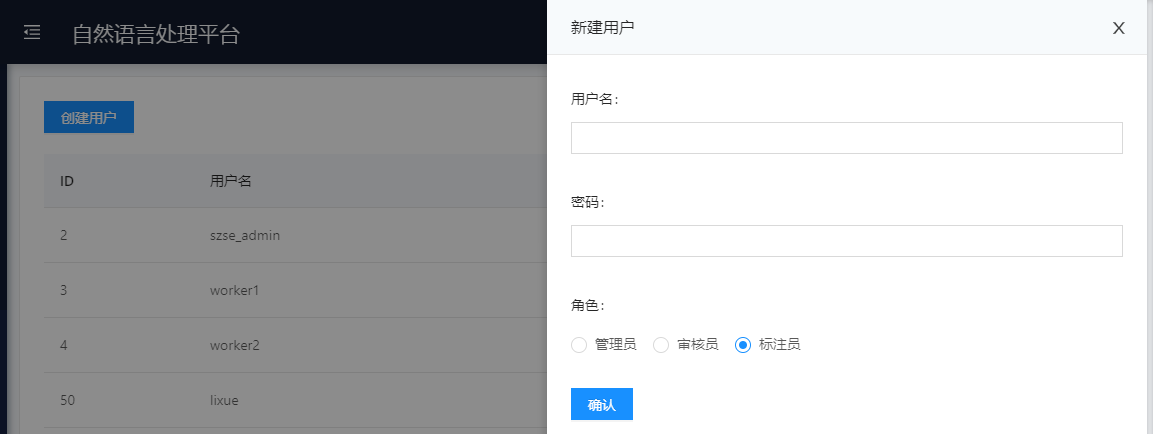
# 八 管理中心

## 8.1 用户管理

管理员用户可以在用户界面可以用户进行增、删、改、查等操作。

### 8.1.1 新增用户

点击用户管理页面中的【创建用户】按钮弹出新建用户界面，管理员填写用户信息并成功保存后创建用户成功。在新增用户界面，管理员的可以给新创建的用户分配角色；不同的角色将会拥有不同的操作权限。



### 8.1.2 编辑用户信息

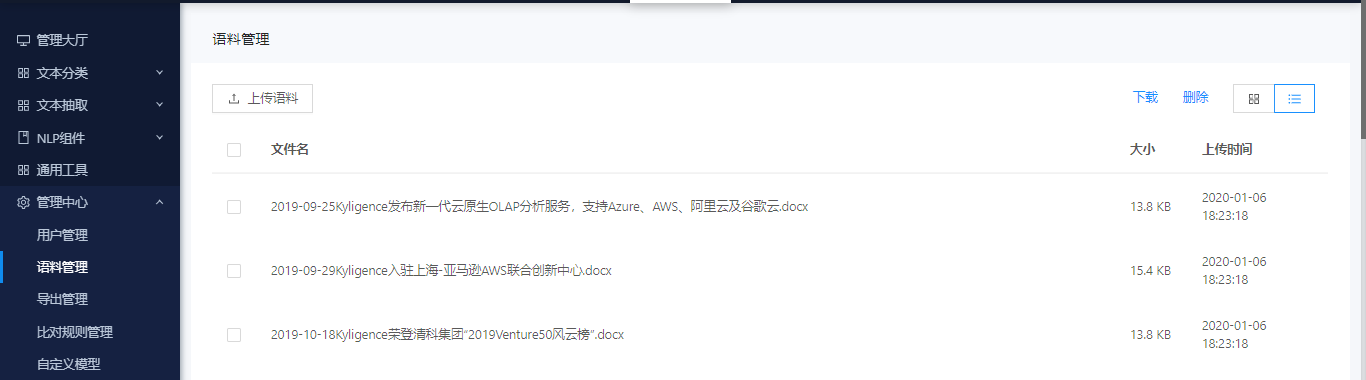
点击用户列表中的【修改】按钮，管理员可以对用户信息进行二次编辑，点击【确定】按钮保存成功后用户信息即被修改成功。

### 8.1.3 删除用户

管理员点击用户列表中的【删除】按钮，确认删除之后即可成功删除该用户。

## 8.2 语料管理

语料库是指用于基础API、常用模型训练时所使用有标注及无标注的训练语料。语料库管理功能支持上传语料、语料下载、删除等功能。



## 8.3 导出管理

进入导出管理页面，点击“下载”可对导出的标注数据进行下载。



## 8.4 比对规则管理

比对规则配置是针对文档比对的内容进行筛选配置，添加一些列标点或其他特殊字符，在文档两两比对过程中，配置过的符号不可参与比对，提高文档比对结果的准确率。

标点符号是常规（常见）的符号，如下图，点击高亮显示，说明已选中，点击”保存“，被选中的标点符号则可以使用在文档比对过程中；

全选：全选功能，所有的标点符号都被选中，点击”保存“，即可使用；

反选：反选功能，标点符号默认选中或未选中状态，点击反选，会取反获取当前未选中或选中的内容，点击保存，即可使用。

其他字符，可以自定义选择，点击”添加“，编写自己需要比对过滤的字符，点击保存，即可使用。

重置：点击重置功能，会将其清空操作的内容，重新操作。



# 九 自定义模型

自定义模型注册应支持各类型算法的自定义修改，系统提供自定义模型所必须具备的模型接口。自定义模型通过使用规定的模型接口，补充自定义算法后，完成自定义过程。系统应支持自定义模型的训练、评估、自动封装导出等操作。

模型名称：自定义

标识名称：配置需与自定义算法服务中的标识名称保持一致

模型类型：可选择长文本抽取、短文本抽取、分类、分词、实体关系

预测端口：输入自定义算法程序定义的预测端口

评估端口：输入自定义算法程序定义的评估端口

模型描述：模型说明内容描述

模型参数：自定义模型的参数



# 附录1 (标注指南)

“数据决定上线、模型逼近上线”，标注数据的质量对该智能文档处理系统的其他功能有着非常重要的作用，如若标注的数据质量较高，再结合我司累积的各种技术定能达到较高的准确率，帮助用户准确抽取文档中的信息。因此，针对标注数据，我司︎出了标注数据的最高准则，即“字段要清楚、标注要统 一、标注要正确、标注要完全。

一、字段要清楚

管理员在设置字段(即文档条款)时，每一个设置的字段的含义需要明确，切勿含糊不清。例如:文档中有大写金额，有小写金额时，建议设置两个字段，分别为【金额(大 写)】、【金额(小写)】。如此设置，可以方便标注员在标注文档时，不会产生字段模糊不清楚及不知如何标的情况。

二、标注要统一

标注要统一主要指的就是标注的内容边界要统一，并且在每一份标注的文档中都需要严格遵守。

例如:文档 A 中出现 20000 元。 文档 B 中出现 43000 元。

(错误)标注情况: 文档 A---2000 元。 文档 B---43000 元 (文档A未标注元，文档B标注元)

(正确)标注情况: 文档 A---2000 元。 文档 B---43000 元 (文档A标注元，文档B标注元)

三、标注要正确

标注的数据要正确，要根据已设置的字段(条款)在文档中正确的标注。

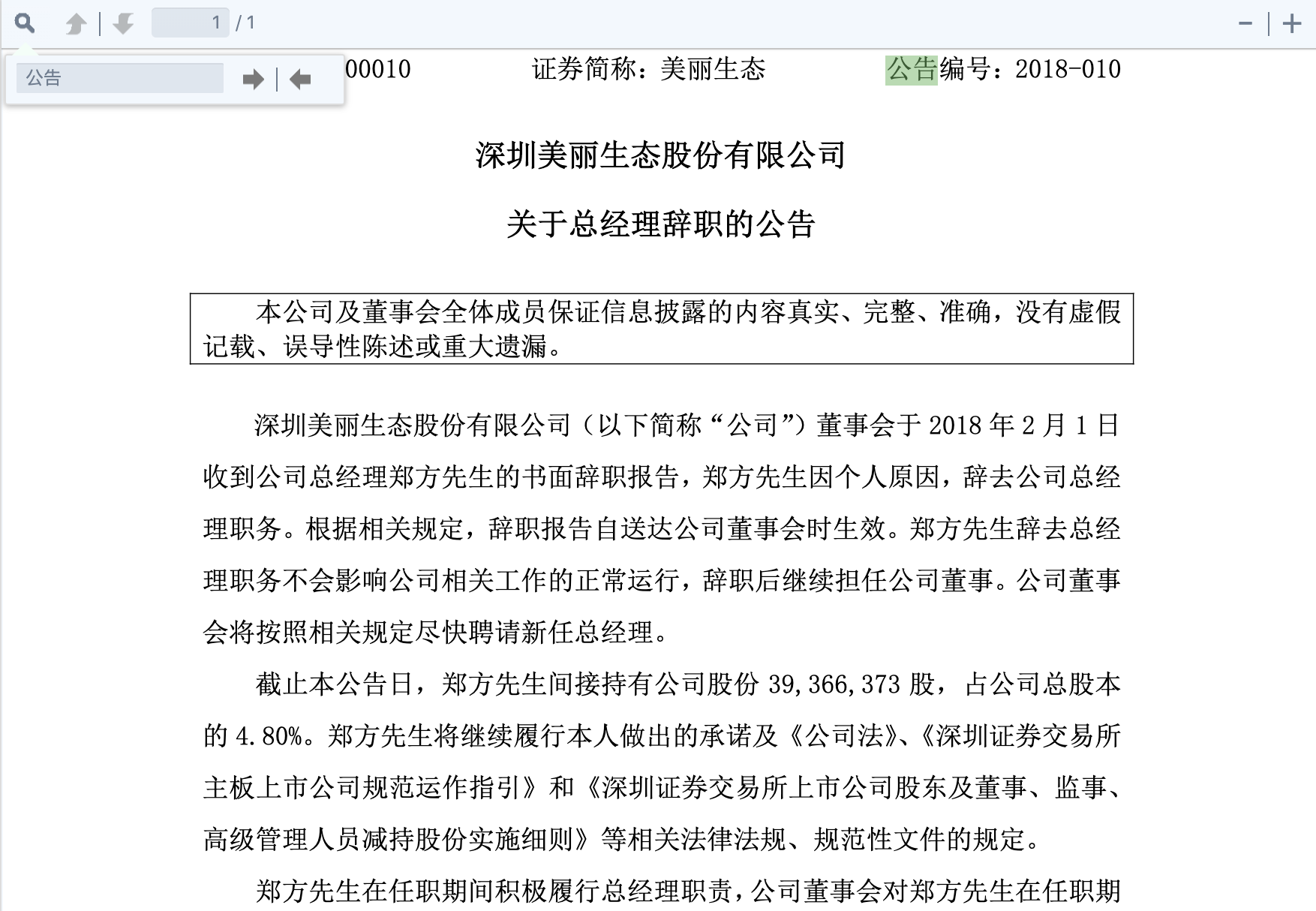
四、标注要完全

标注要完全指的就是对文档中出现的字段(条款)内容都应标注。

五、其他

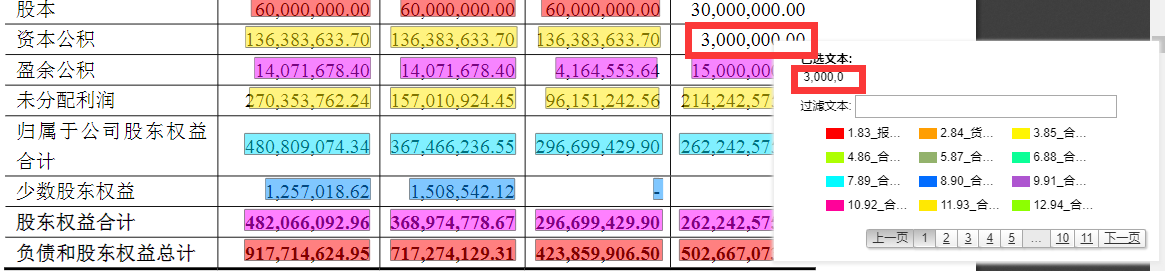
1）字段位置确认

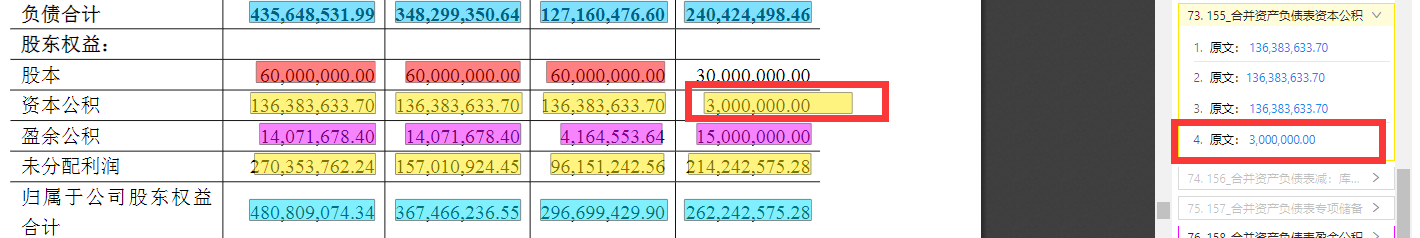
当文档较长时，可分析样本，按章节对字段进行分类，弄清每一部分字段处于文档的哪一章节，这样标注时可以直接根据目录以及页码跳转对长文本的字段进行快速定位；也可总结每一章节部分的字段有无关键词规律，直接通过搜索关键词定位；



2）选中文本确认

选中要标注文本后，一定要核对弹窗的“已选文本”和要标注内容是否一致，因文档原因，选中后的“已选文本”会出现多选中或者少选中的问题。如图1，鼠标全选后“已选文本”与原文内容相比不完整，这时鼠标拖动时往后多选中一些就能标注完整，如图2所示。同理，如“已选文本”选中的内容较多，则拖动时少选中一些。

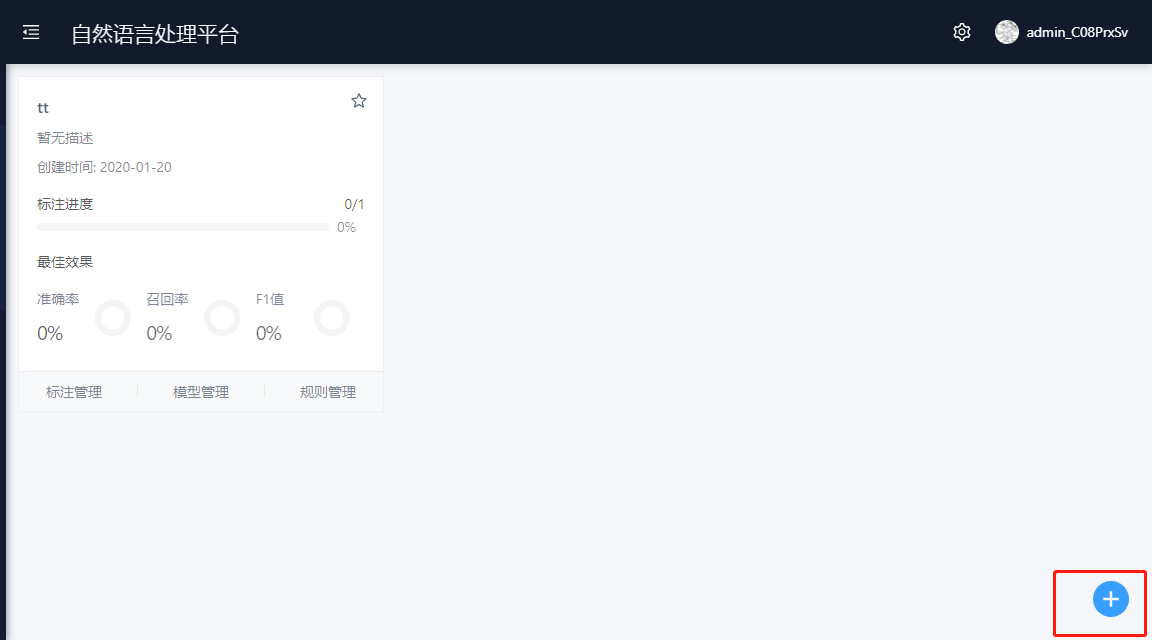




# 附录2 (新手指南)

如果你是一位想自己从头创建文档类型，自己训练模型的话，建议按照如下的操作路径完成；

第一步：新建项目；



第二步： 新建标注任务并上传文件；（选择你刚刚好创建的文档类型）



第三步： 标注任务中的文件并审核；

第四步： [训练模型](#_1）Q：如何新建模型/模型训练？)；通过【模型管理】🡪【新建模型】入口训练模型



第五步：[上线模型](#_6）Q：如何上线模型？)

自此，该系统中就已上线你创建的文档类型的模型，后续可按操作手册完成[文档的抽取](#_文档抽取)、[文档审核](#_文档审核)、模型评估、模型重新训练等功能。

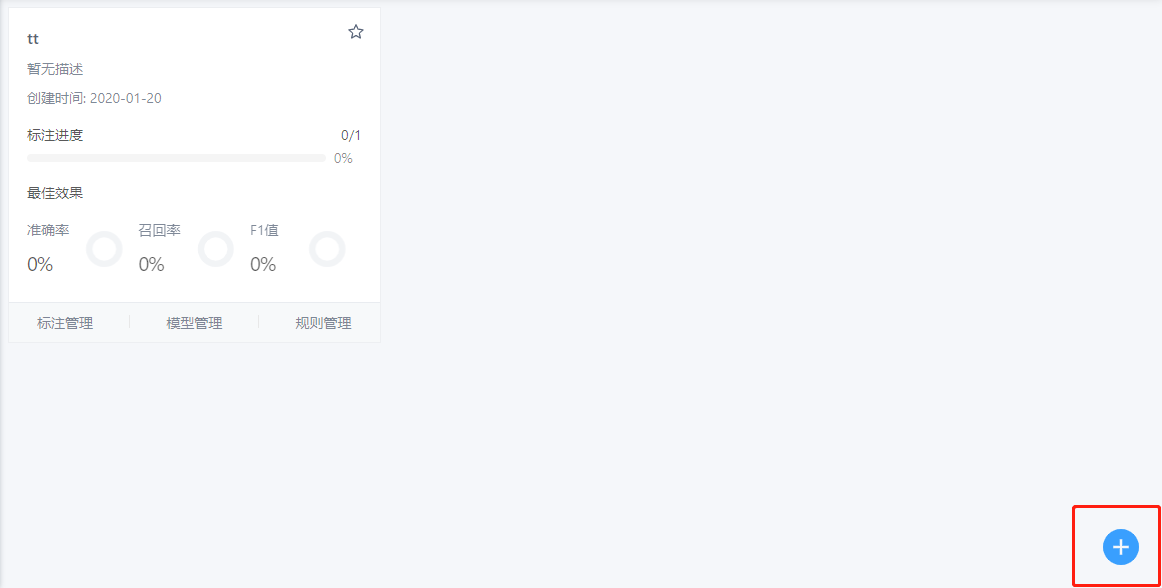
# 附录3 (用户常见问题)

Q1：企业内部账号可以登录吗？

A1：可以，系统已进行SSO单点登录对接，私有化部署后，用户可以输入企业内部账号登录，通过身份认证服务后即可访问。

Q2：新手训练模型从哪一步开始？

A2：系统以项目管理的形式和方法对使用进行管理，分类、抽取、分词、实体关系菜单下均设置有项目管理页，第一次使用可在项目管理页下从新建一个项目开始，在项目卡片上进入标注管理、模型管理等。



Q3：标注文件可以支持哪些格式？

A3：分类、抽取、实体关系标注文件支持上传的格式包括.pdf、.doc、.docx、.txt、.csv，分词支持上传的格式包括.txt、.csv

Q4：标注是否支持多人标注？支持多人标注的方式包括哪些？

A4:是，支持共同标注、均分标注方式

Q5：标注的数据是否必须经过审核？

A5：是，标注的数据必须经过审核员审核才能被勾选用于模型训练、和模型评估

Q6：标注数据、预测数据是否支持导出？

A6：支持，标注数据从项目卡片上点击“标注管理”进入选择标注任务导出标注数据，包括分类、抽取、分词标注数据导出，实体关系标注数据的字段分析导出；分类、抽取、分词预测数据导出

Q7：训练完的数据是否必须经过上线操作才能用于预测？

A7：是，模型训练成功后在模型管理页面点击“上线”后该项目训练的模型即可用于预测

Q8：是否可以自己新建通用工具提供给用户使用？

A8：是，用户可以自行在通用工具页面点击“添加自定义工具”新增新的工具

Q9：标注文件、语料文件等上传是否有大小限制？

A9：上传文件的大小通过nginx限制，具体大小可灵活设置，具体咨询管理员。