# 一、模块名称

智能化文本AI识别模型

# 二、模块功能或需求描述

## 概述

智能化AI识别模型旨在利用智能化和机器学习技术，自动化的识别和分类各种类型的文本数据，提升效率和准确度，减少人工干预。

## 主要功能

文本分类；识别文本中的特定实体；捕获事件行为，获取指定的操作结果和行为动作；

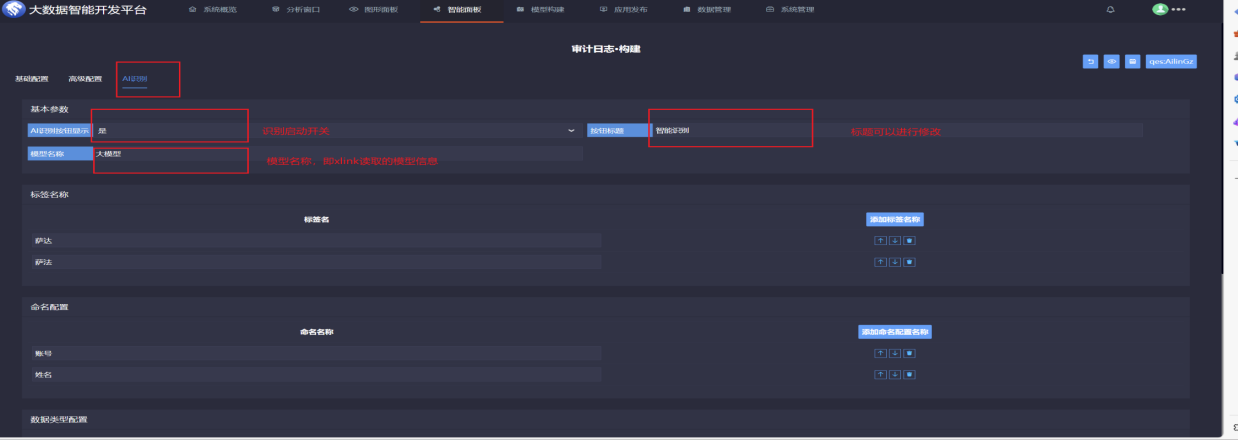
# 三、模块设计

## 1.流程设计

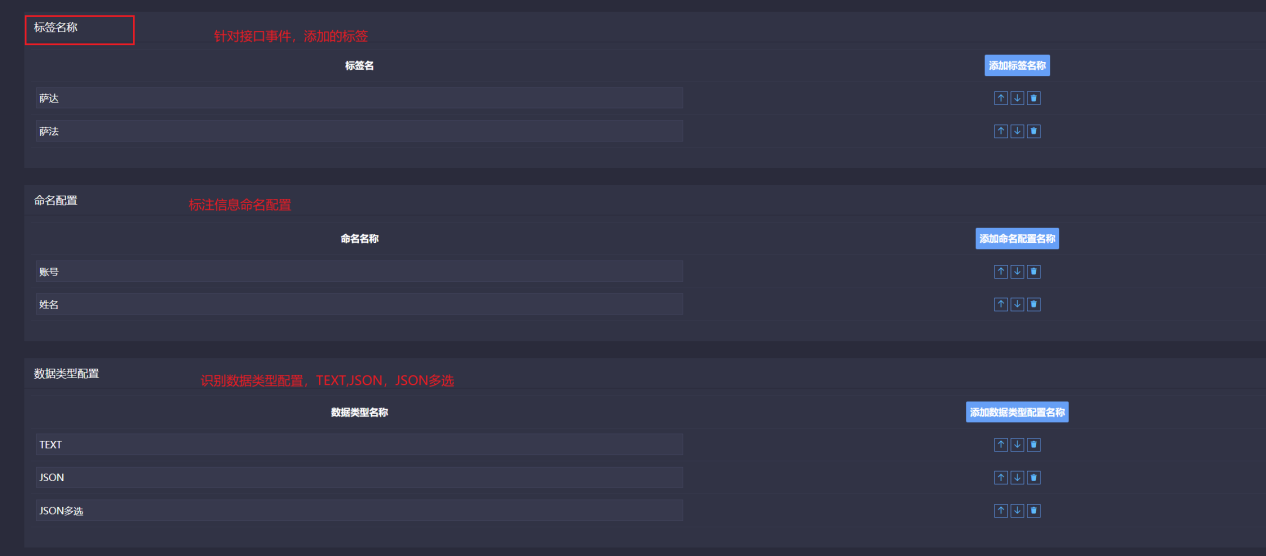
### ① 后台配置

**1.基本参数**

**(模型名称每个查询分析务必名字不一样，否则会识别到其他查询分析的模型)**

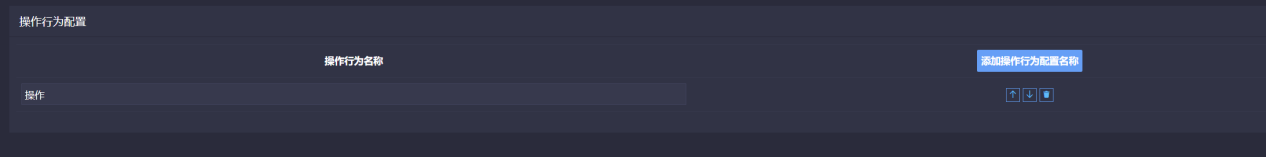


**2.标签名称、标注文本命名、数据类型配置**



**3.操作行为配置**

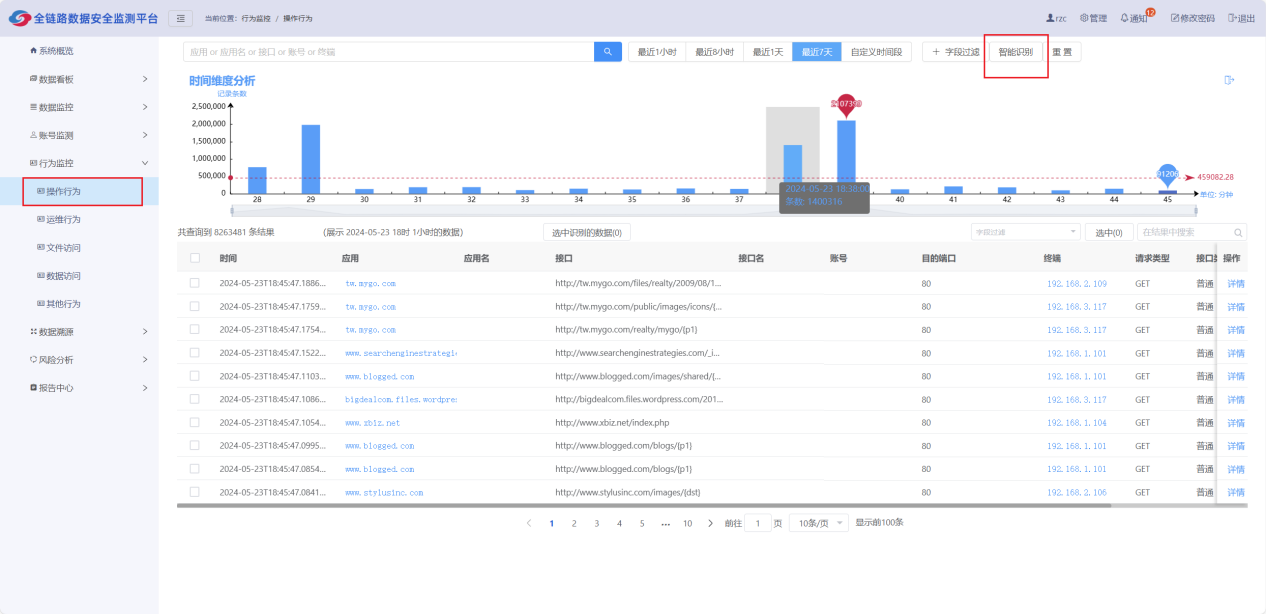
对标注的文本添加操作行为的分析。

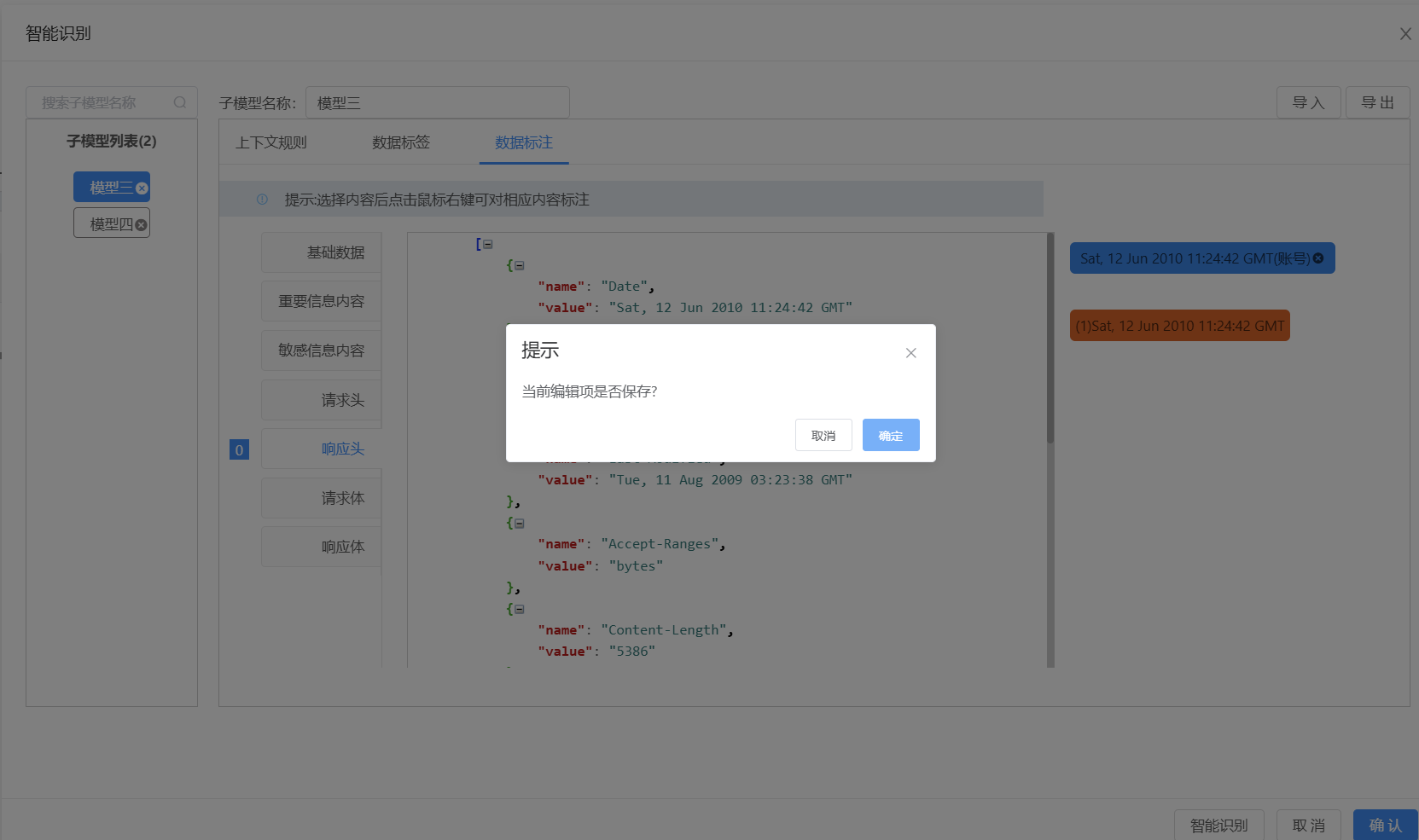


### ② 前台配置及操作方法

#### 界面标签

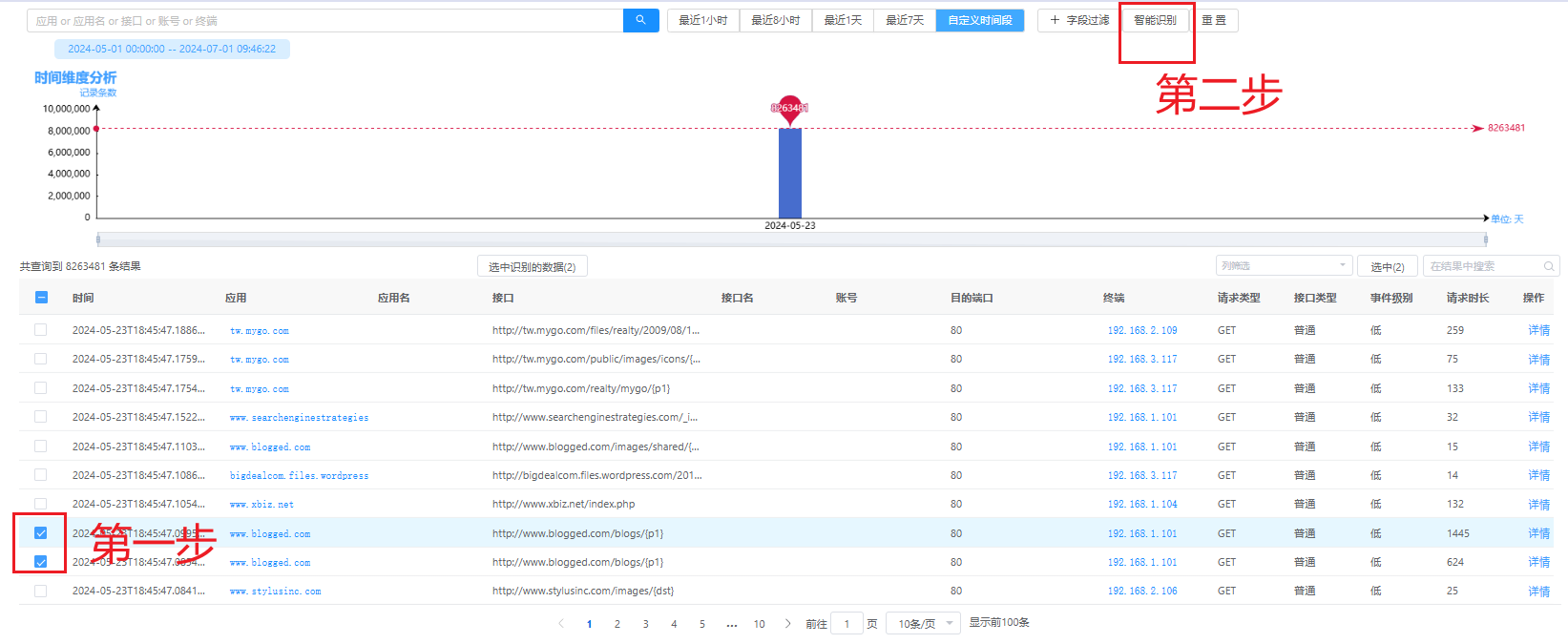
**（找到对应查询分析的导航，后台配置开启按钮之后，界面会显示自定义的标签名右上角所示）**





#### 标注界面设置（新增标注信息）

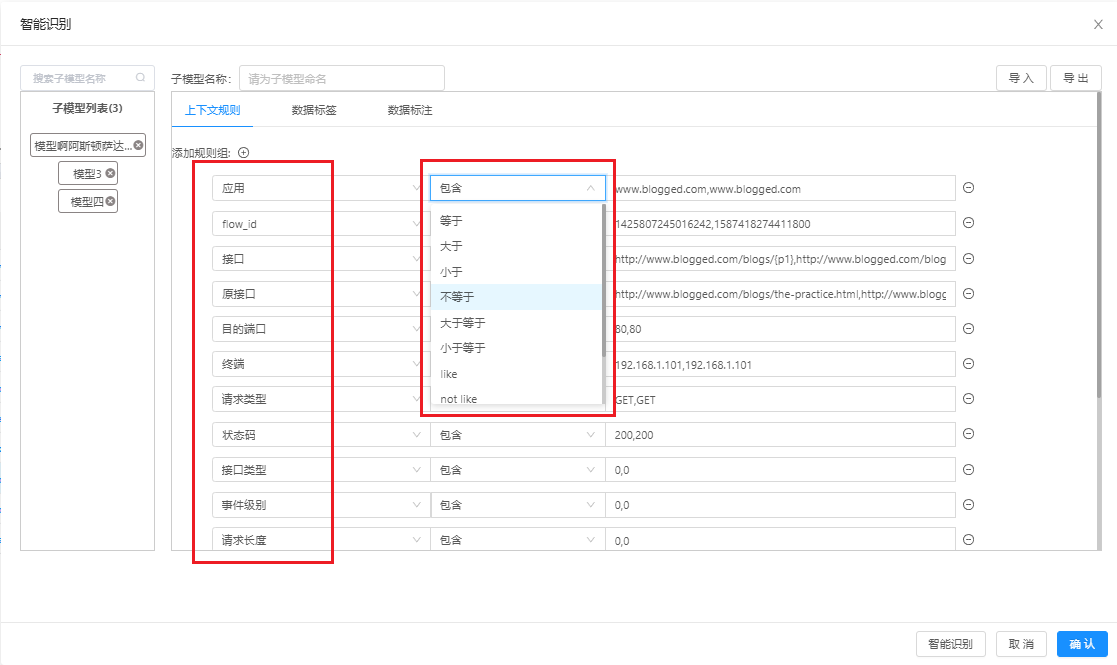
##### （1）选择需要标注的信息（图中选择两条），点击识别按钮



##### 标注三要素（上下文规则、数据标签、数据标注）

###### 1.上下文规则

（根据该查询分析表字段返回的信息，上下文作用是根据规则进行辨别处理的是否为该条信息）



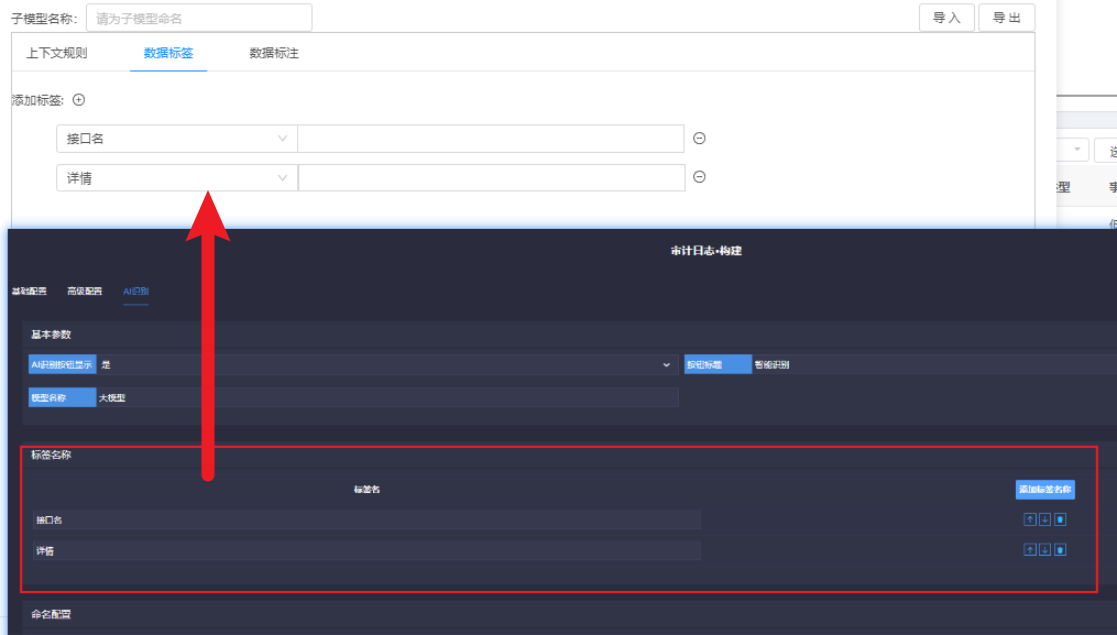
例如：想要对该信息进行详细辨别，可以对上下文规则组进行筛选，选出有用数据信息，

这里只剩下应用和原接口信息，目的是在获取流处理信息进行判断是否为当前的上下文信息，便于识别；

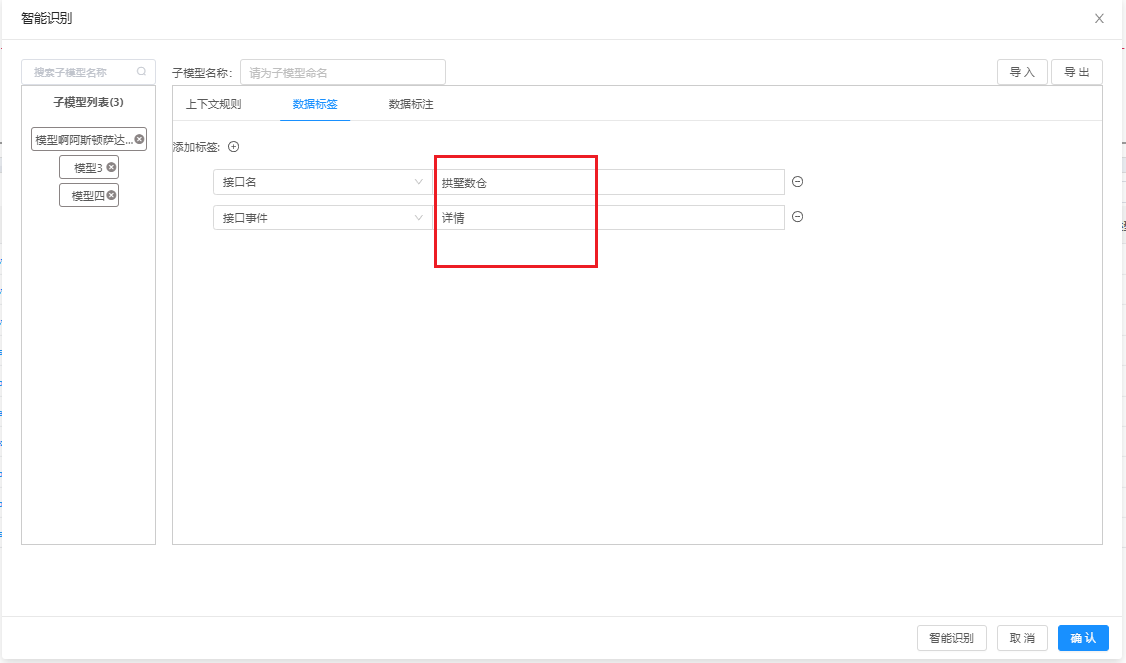


###### 2.数据标签

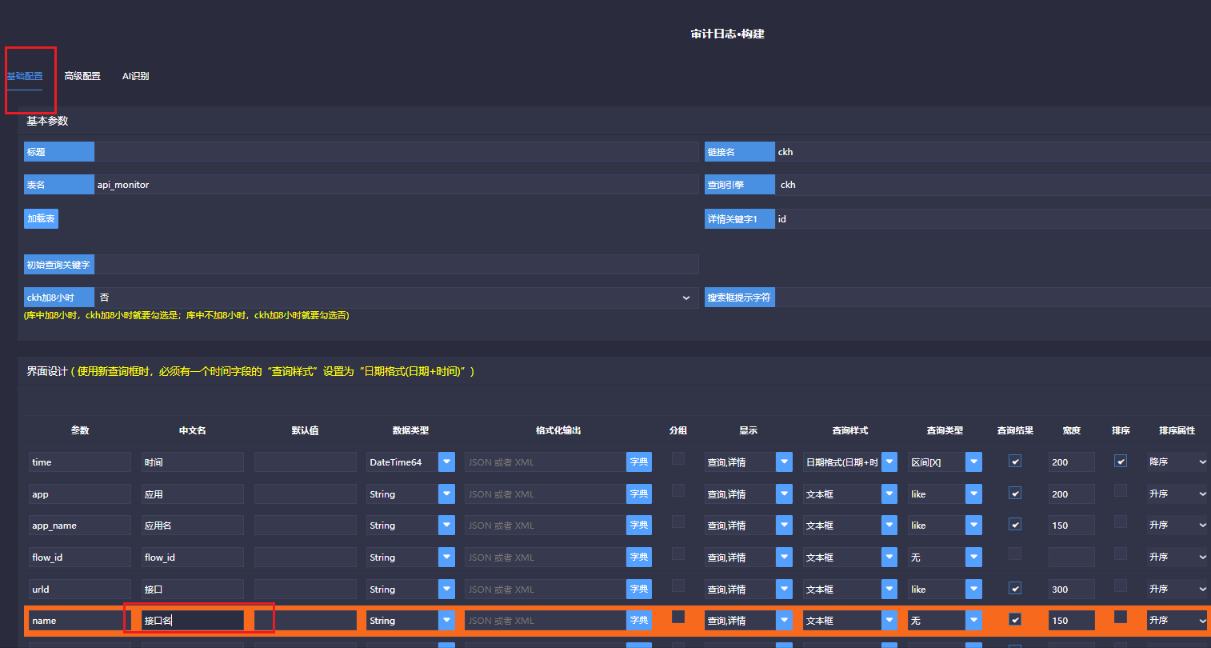
（这里对应后台配置中标签名 配置，配置完数据进行识别时默认展示配置标签，也可以单独新增标签名）



数据标签用于对子模型中Http请求的一个描述，比如接口名：拱墅数仓，洛阳大数据等等；接口事件：详情；到时都会通过识别结果将配置的标签给返回



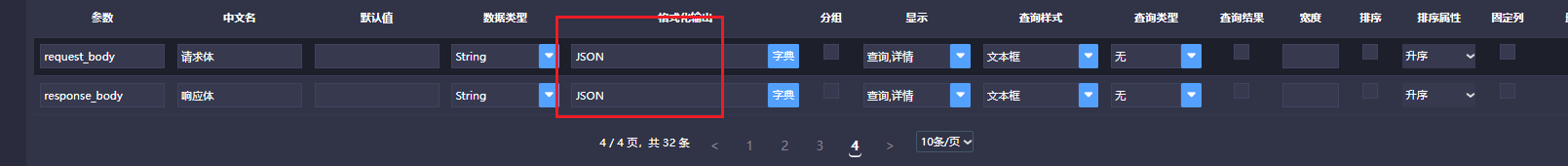
注：如果你在后台配置的基础配置中，给字段的中文名和标签名重复，如（接口名与数据标签重复）会传递字段的参数名称，比如下图，会将“name”返回，而不是返回接口名；如果不想选择标签也可以置为空，不做必须条件。



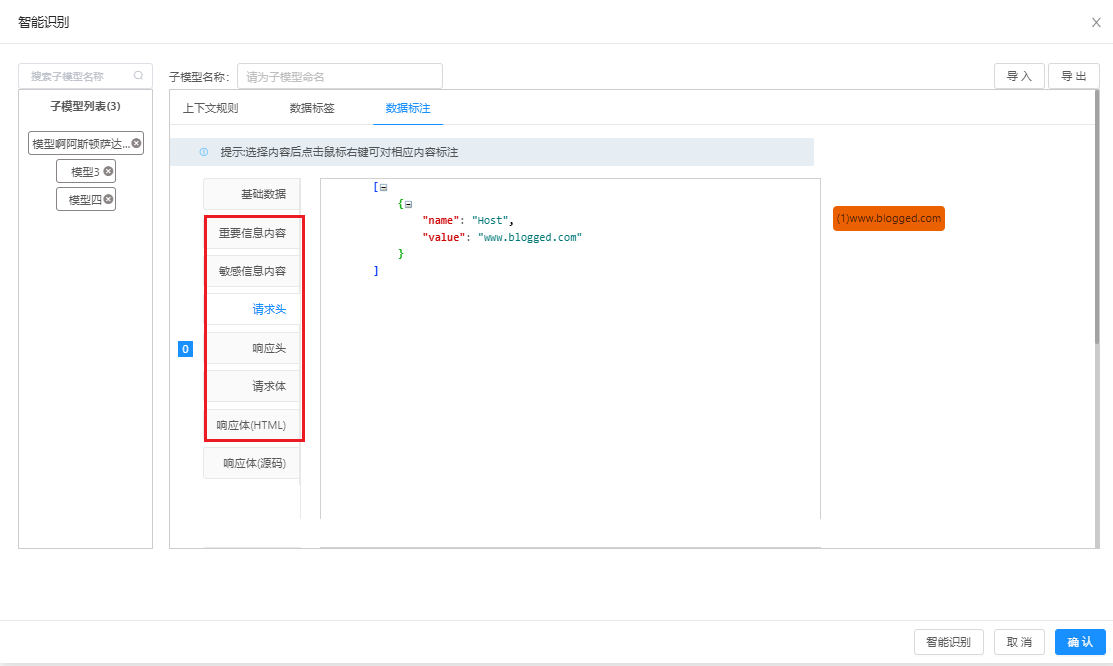
###### 数据标注

（关键一步，对想要进行获取的信息进行标注）；能够进行识别的项是根据后台设置有关;如下图所示：表字段中进行JSON格式化的就可以进行标注

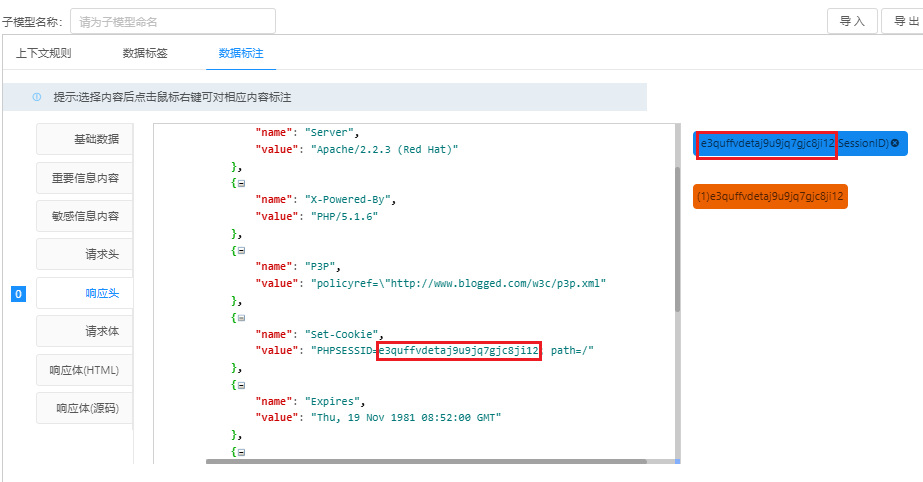




前台界面中能进行标注的除基础数据外，下面的标签都能够进行标注；基础数据中目前只有参数能够进行标注；

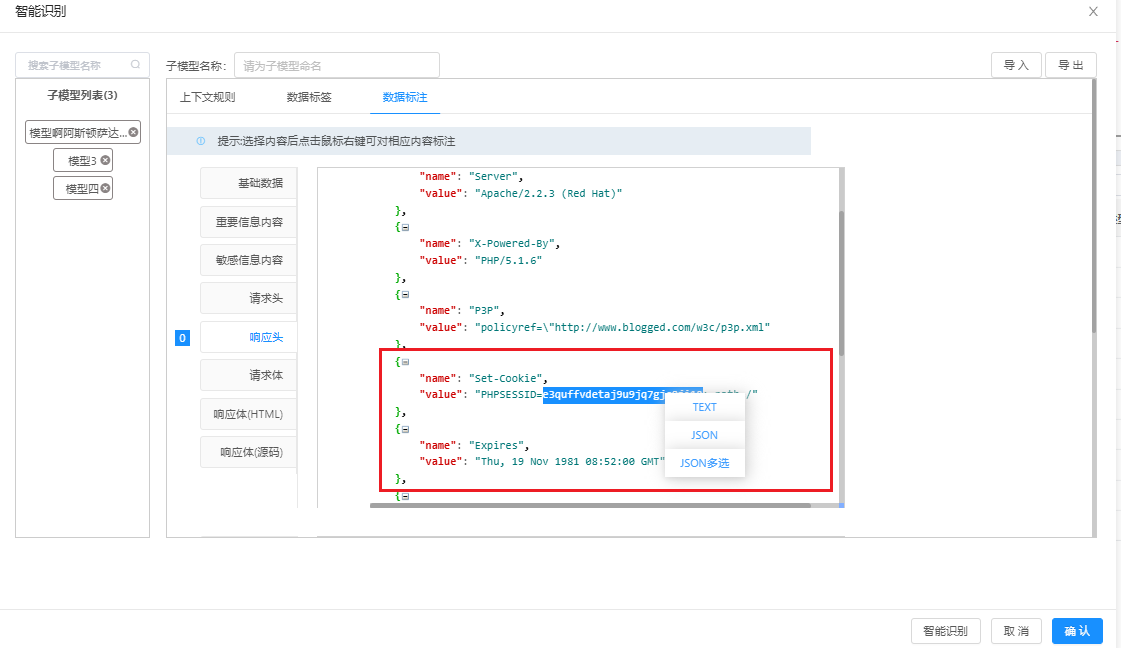


进行标注：例如果想要获取到HTTP请求的SessionID,通过观察，SessionID存在与响应头中，就可以在响应头中进行标注，如下图所示



##### 标注流程

1. 选中需要标注的信息，选择TEXT识别，目前请求头、响应头只能根据TEXT进行识别，且需要标注 name: 、value:后面的值；



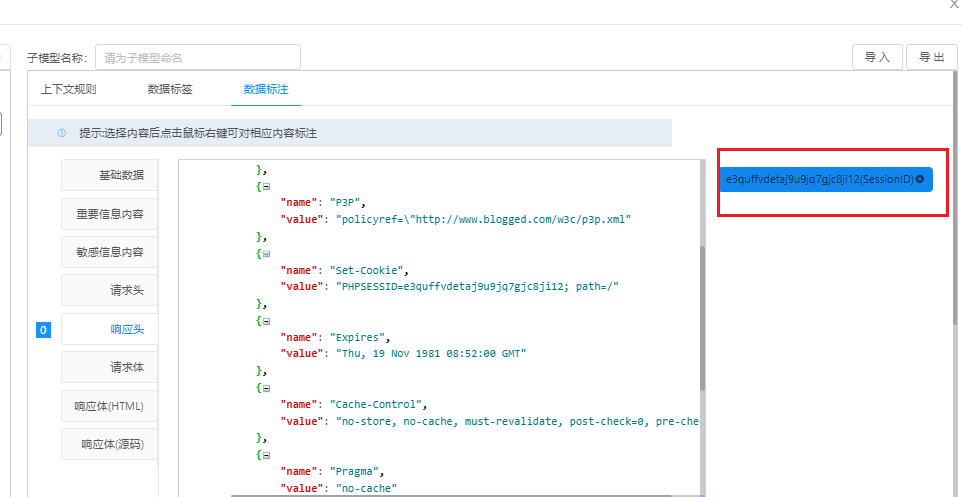
1. 弹出标注信息的配置

命名名称：（自定义名称，如姓名，地址，SessionID）;

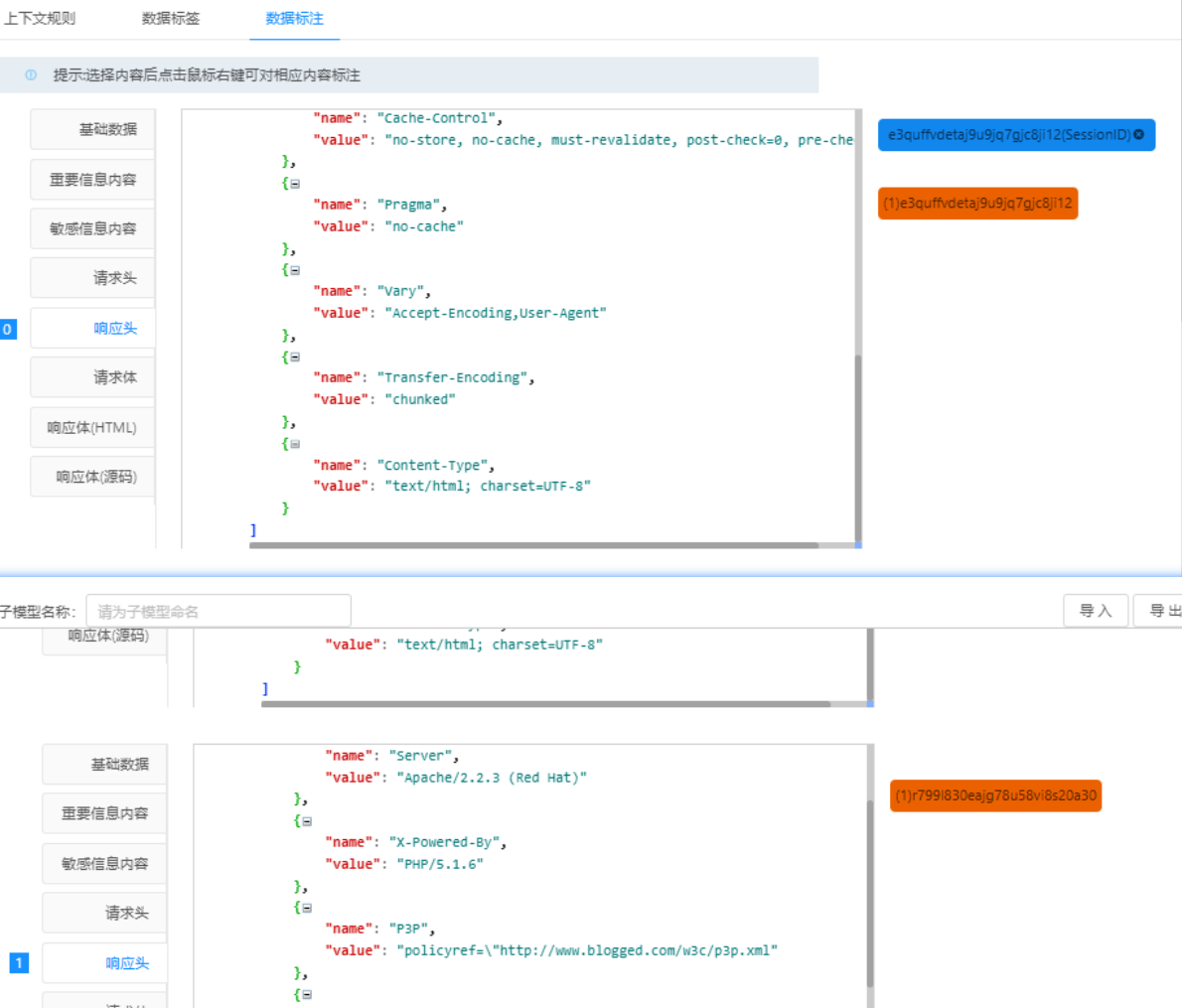
命名标签配置：（后台配置信息，比如目前我们的场景：用户是操作，还是操作之后的返回结果，由于这里是sessionID，所以视为返回结果，具体的根据不同场景进行配置）；



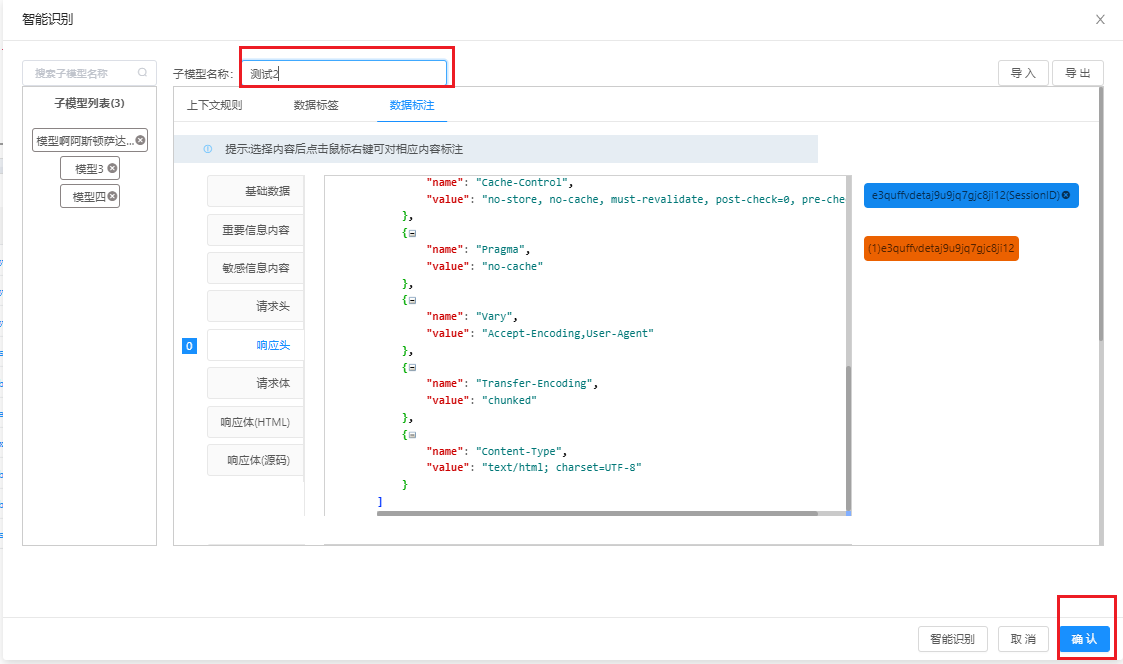
点击确认，右侧会展示标注的信息



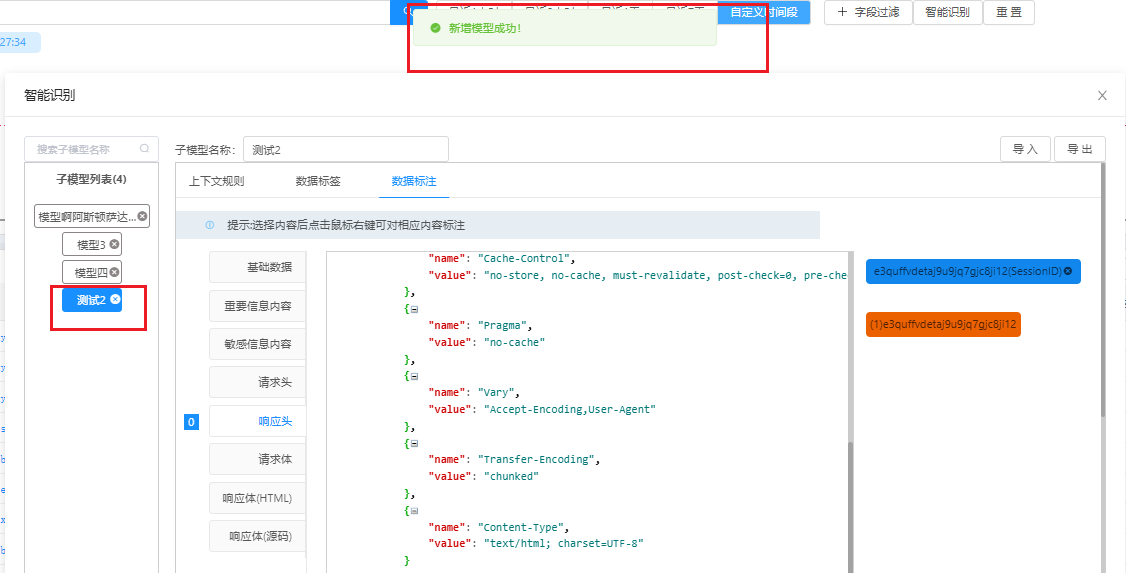
1. 点击智能识别会将识别的结果给展示出来，这里我们选择两条数据，所以两条数据的结果都能够展示出来（前提是需要提取的两个HTTP请求都具有类似的特征，这需要我们进行观察）



1. 识别结果子模型保存，输入子模型名称，点击确定，就能进行保存



保存成功如下图所示，新增的模型名称会出现在子模型列表中；



#### 标注界面设置（子模型列表及搜索）



#### 标注界面设置（查看子模型）



#### 标注界面设置（删除子模型）

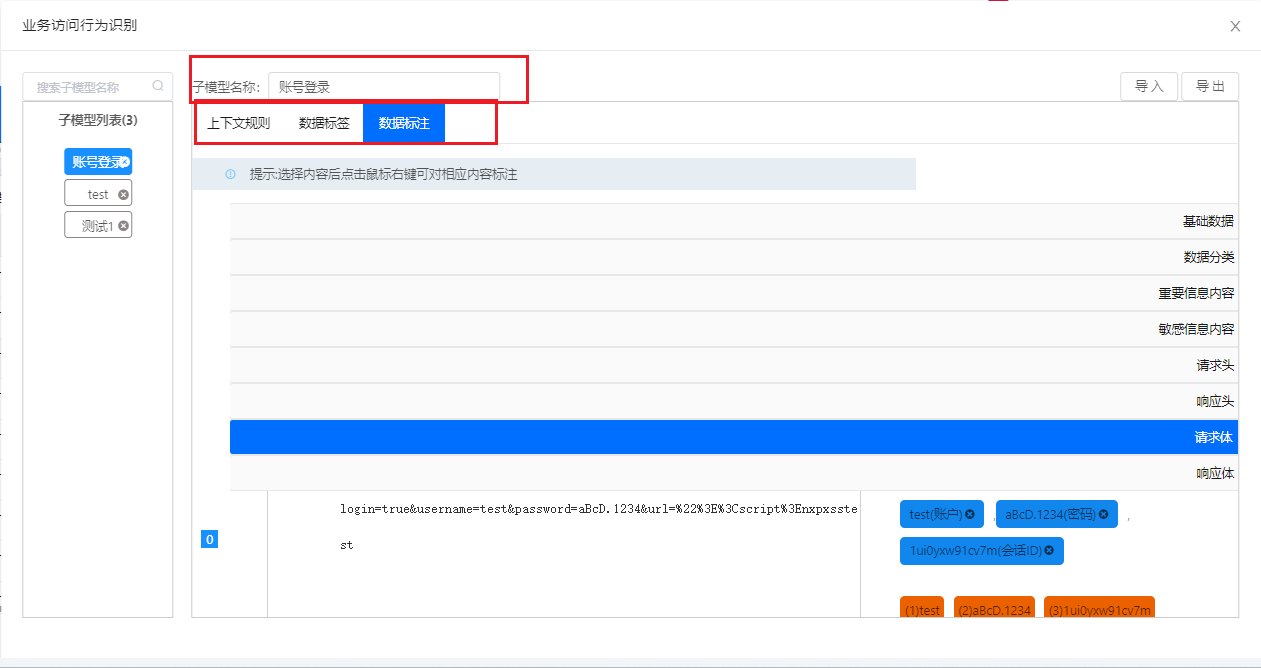
**（点击 x 进行删除，会将该模型删除掉）**





#### 标注界面设置（修改子模型）

针对子模型名称、上下文规则、数据标签、数据标注 都可以 对其进行修改，数据标注修改必须要点击智能识别，修改之后点击保存即可

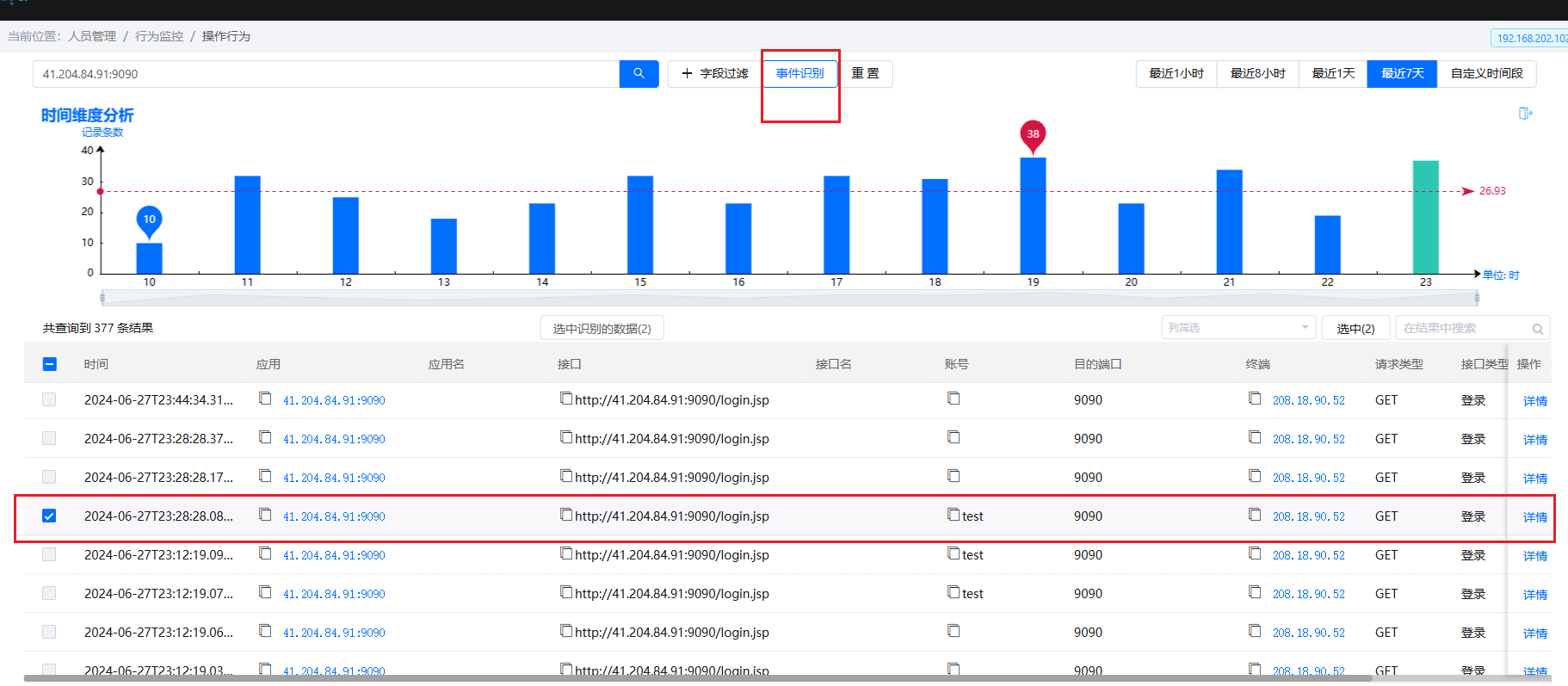


# 模块测试识别及用例

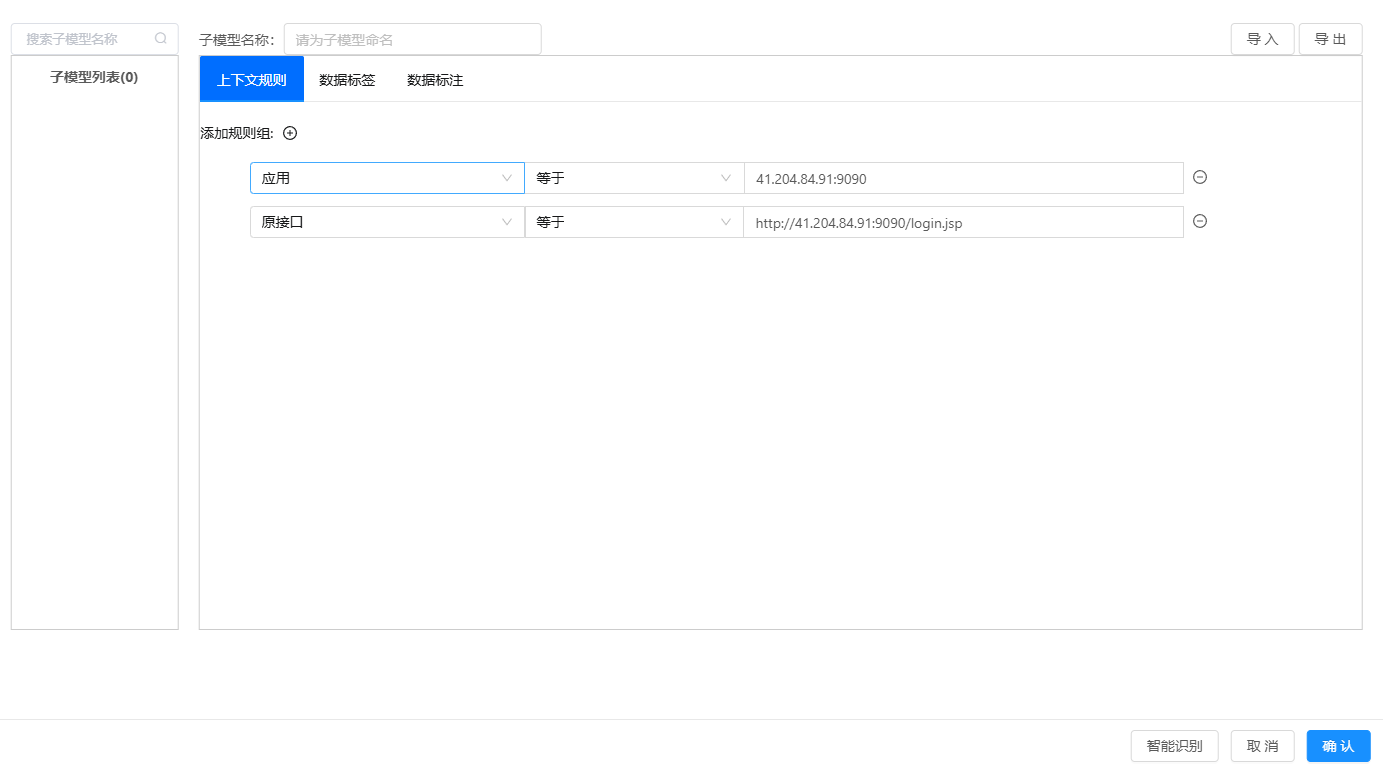
## 用例1

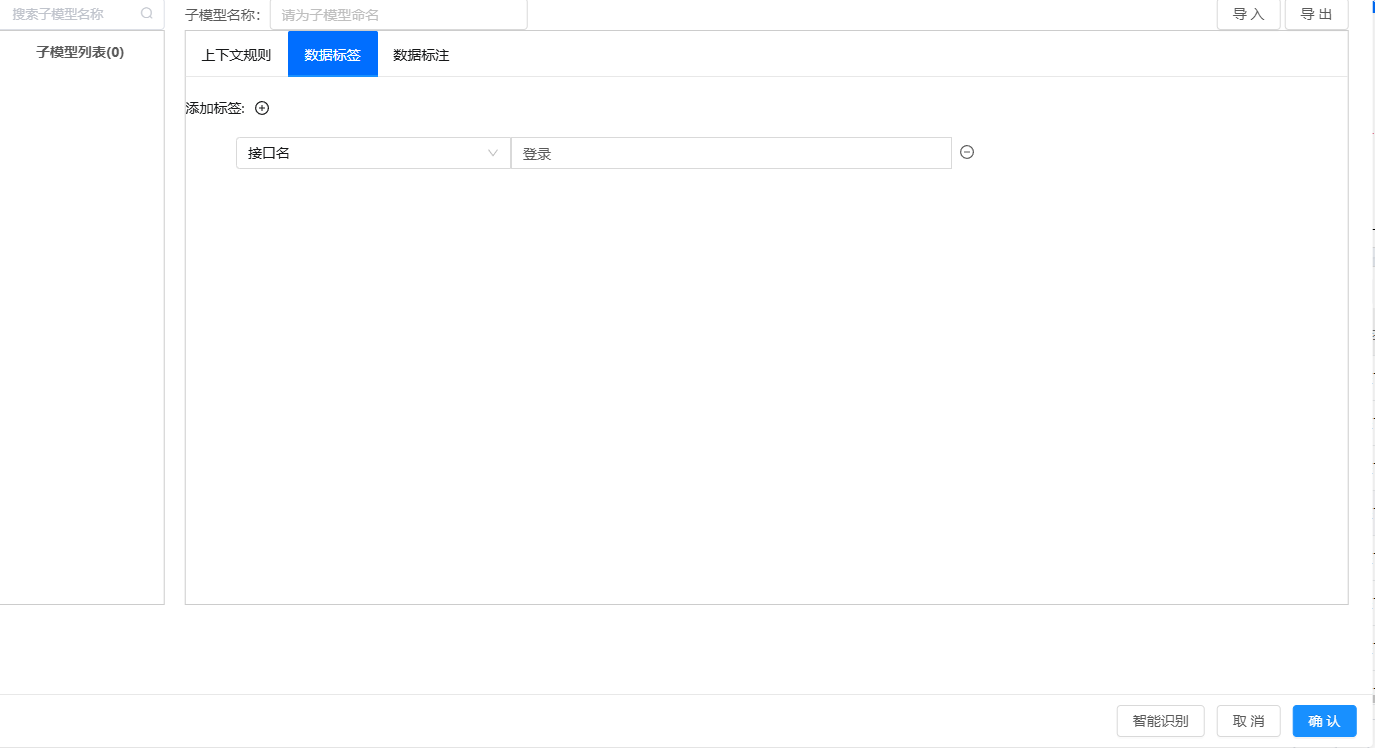
（对登录接口进行标注，获取到相应的账户，Cookie信息，如果登录之后就会获取到相应的Cookie信息，这样就能根据该登录接口去判断对应账户）

### ①选择标注信息



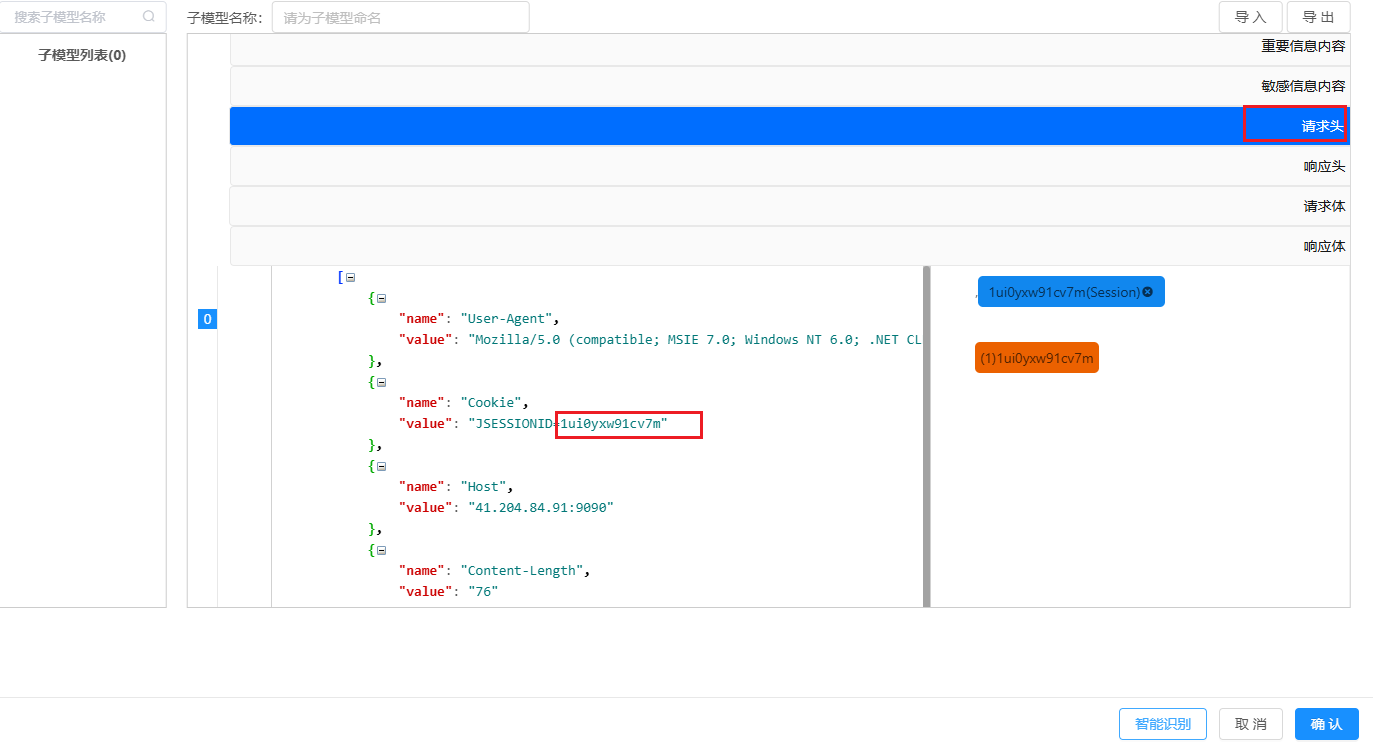
### ②上下文、标签信息处理



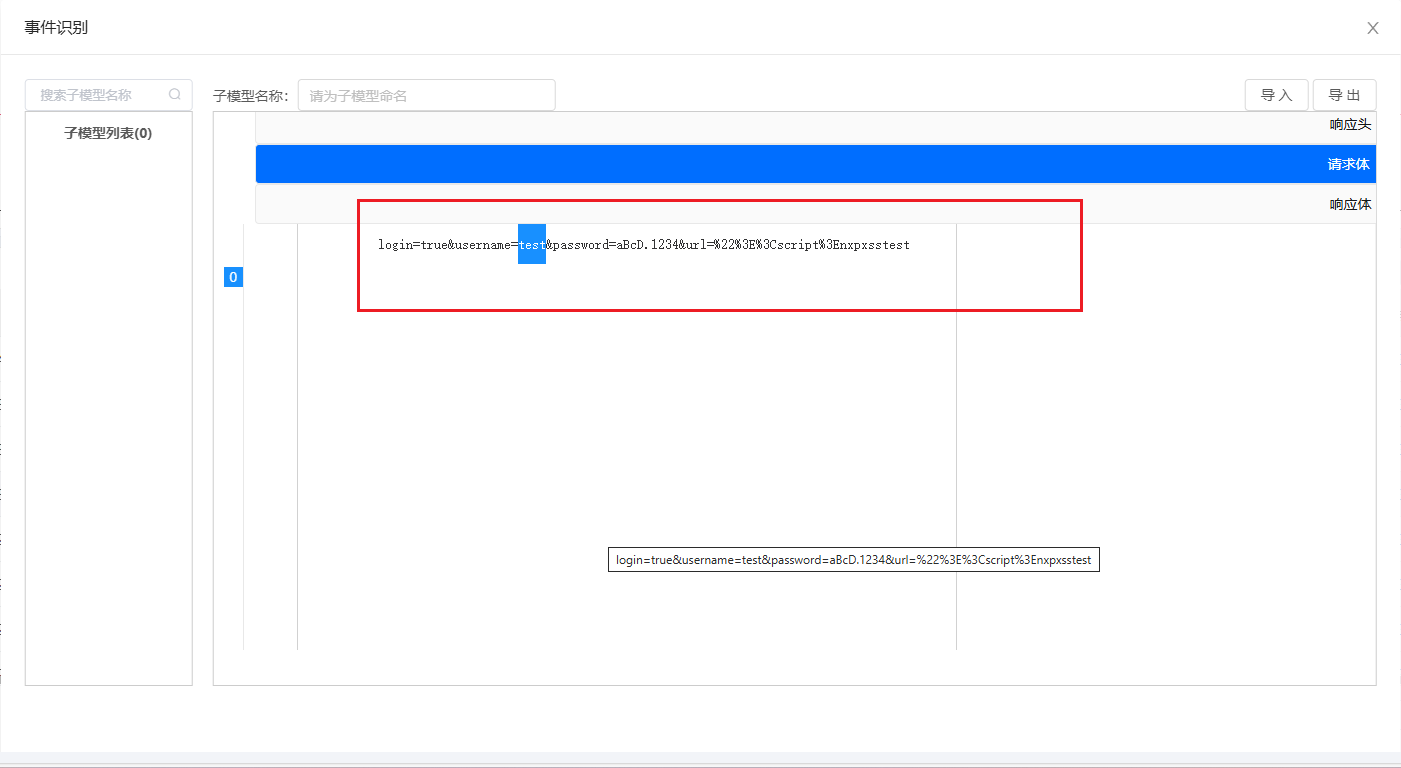


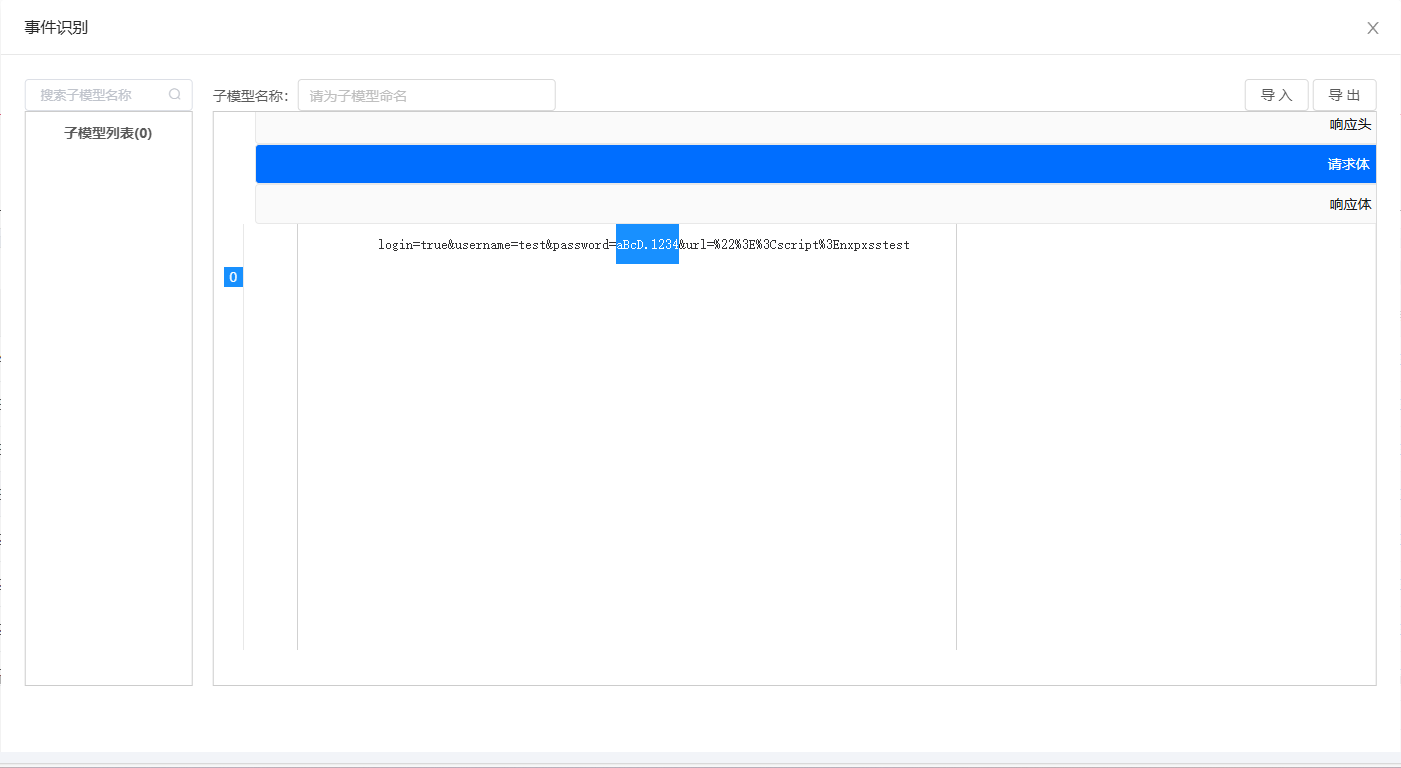
### ③数据标注

从请求头中获取sessionID



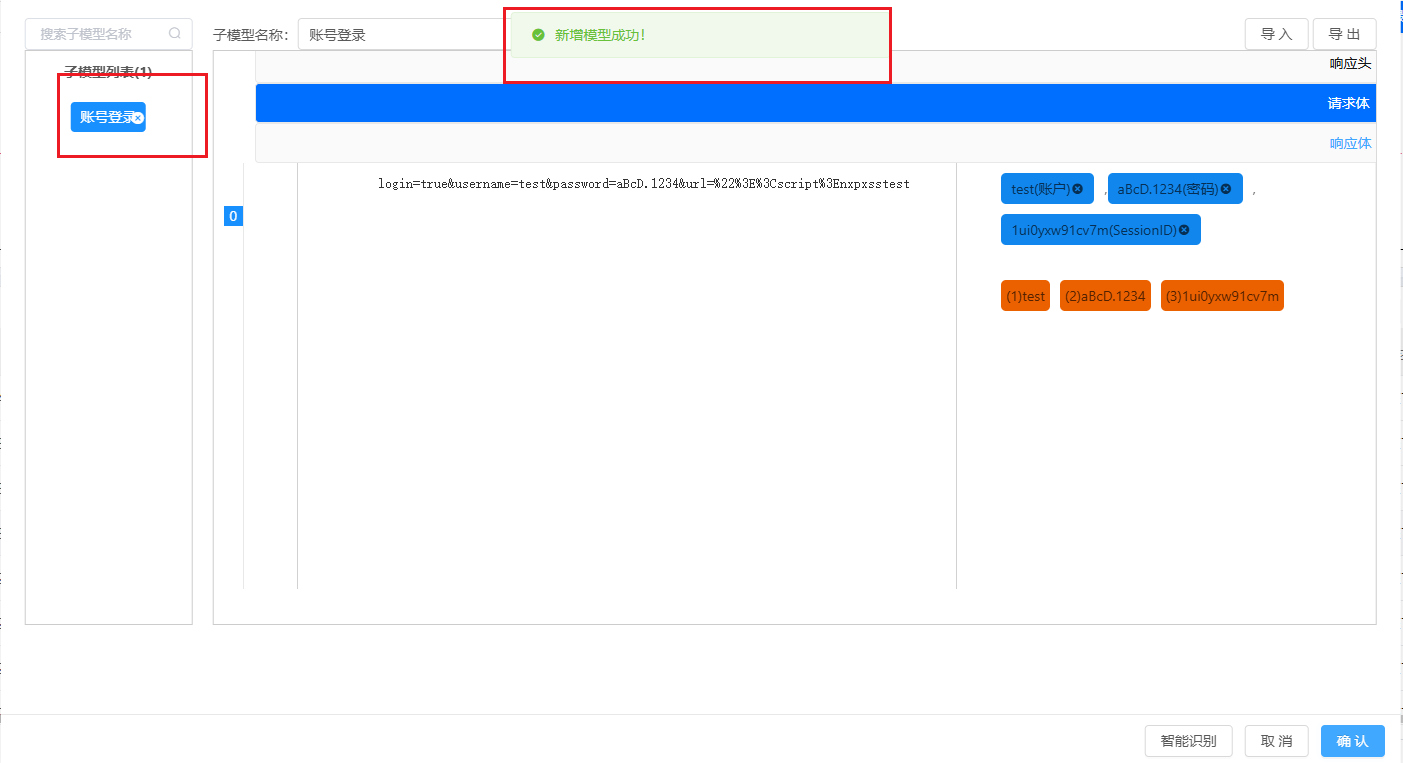
从请求体中获取我们想要的账号与密码





返回识别的结果信息

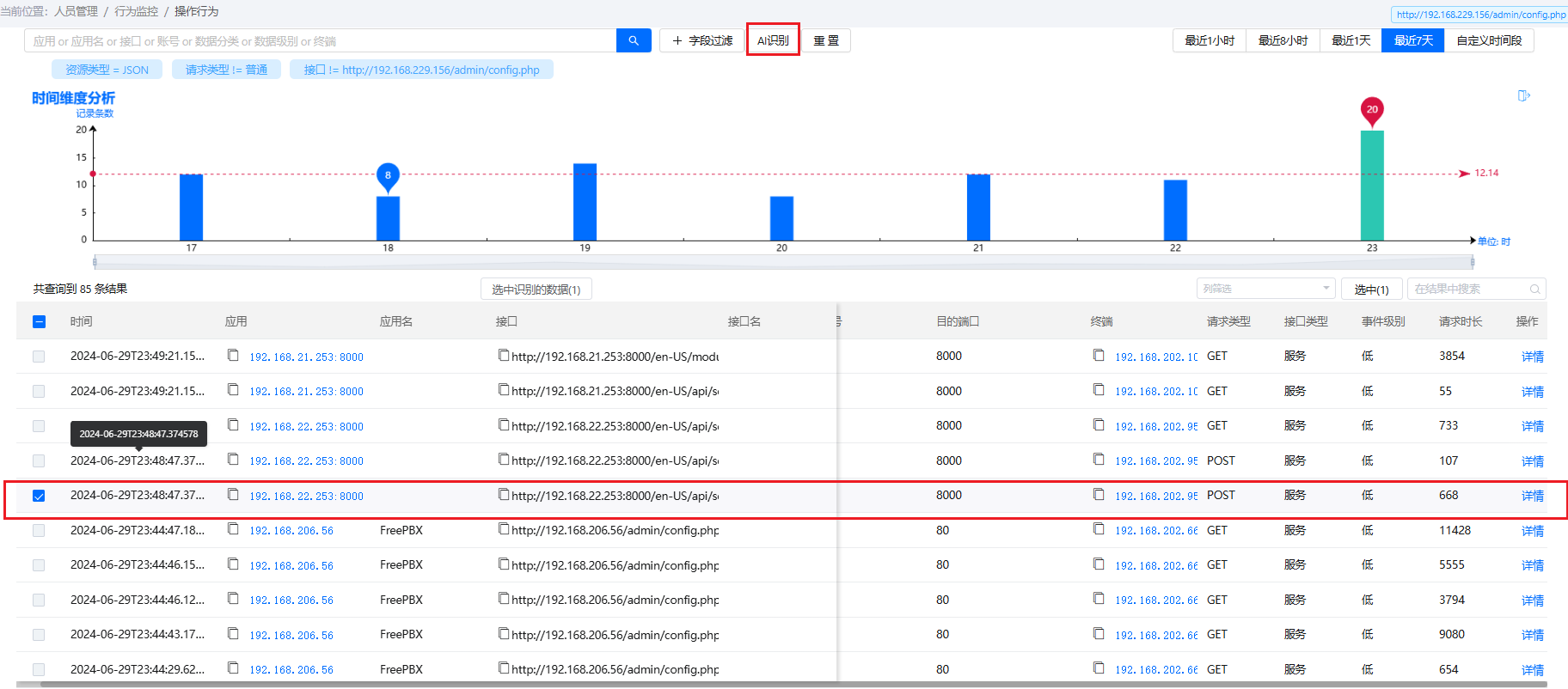
### ④保存数据结果



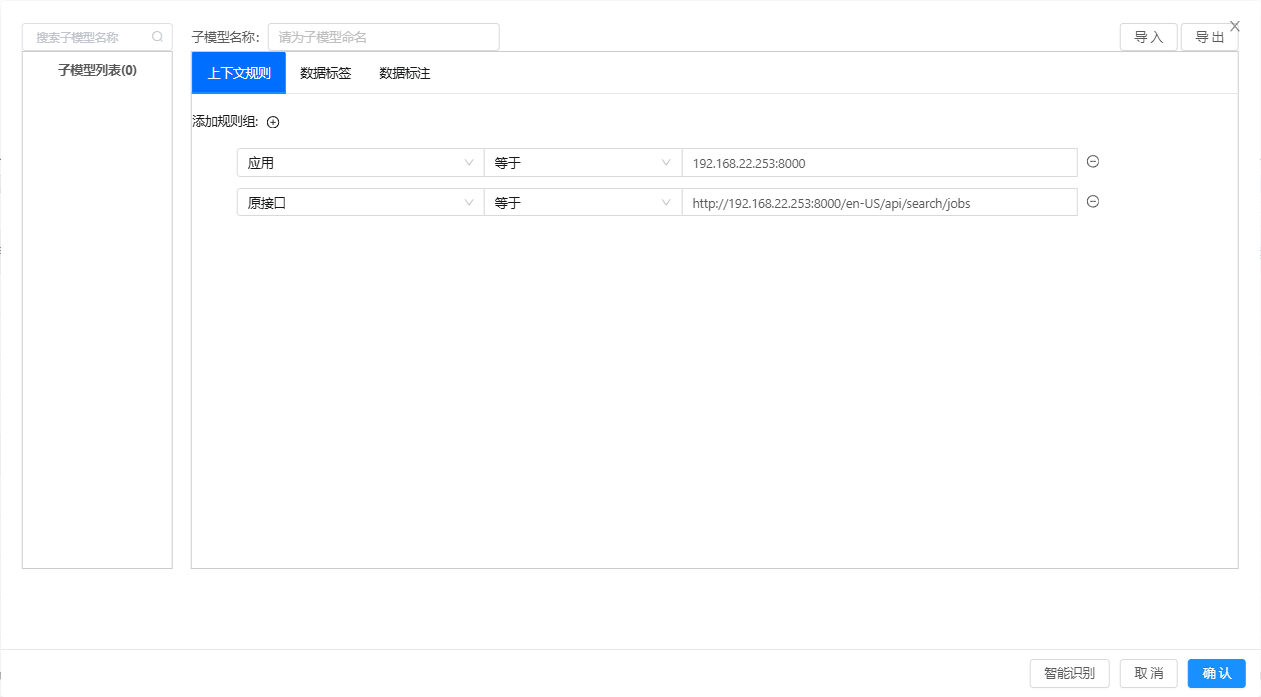
## 用例2

（针对JSON数据的提取）

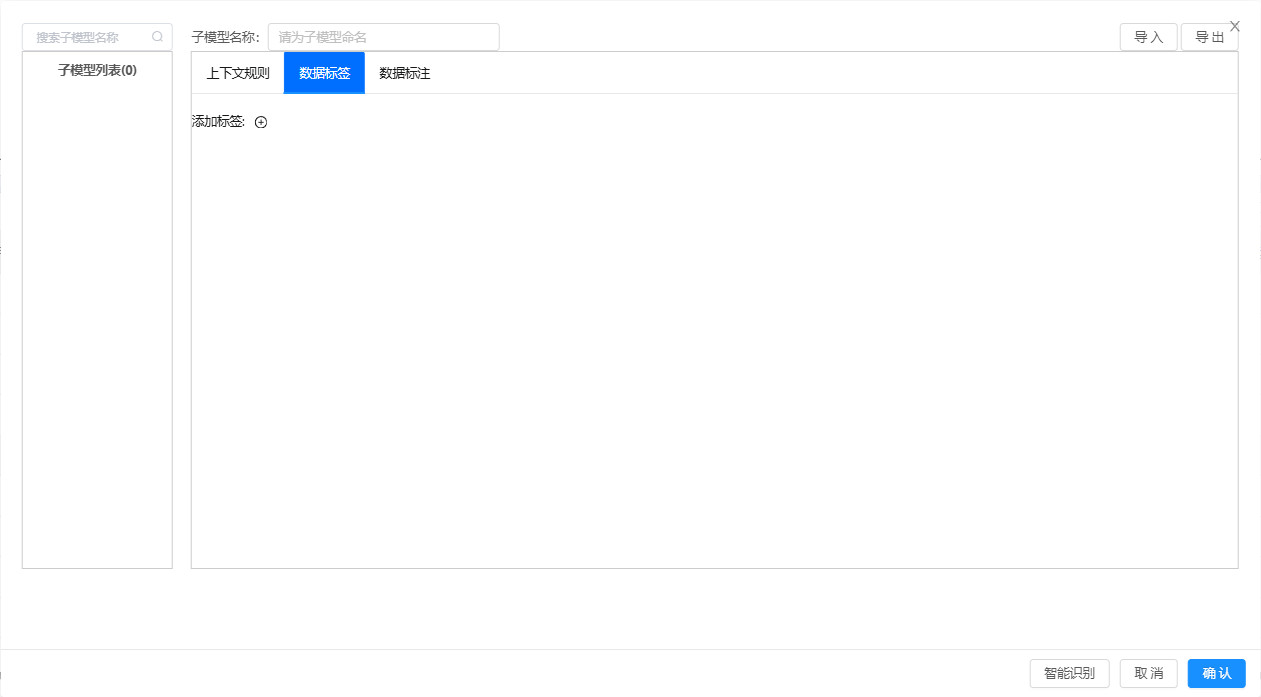
### ①选择标注信息



### ②上下文、标签信息处理

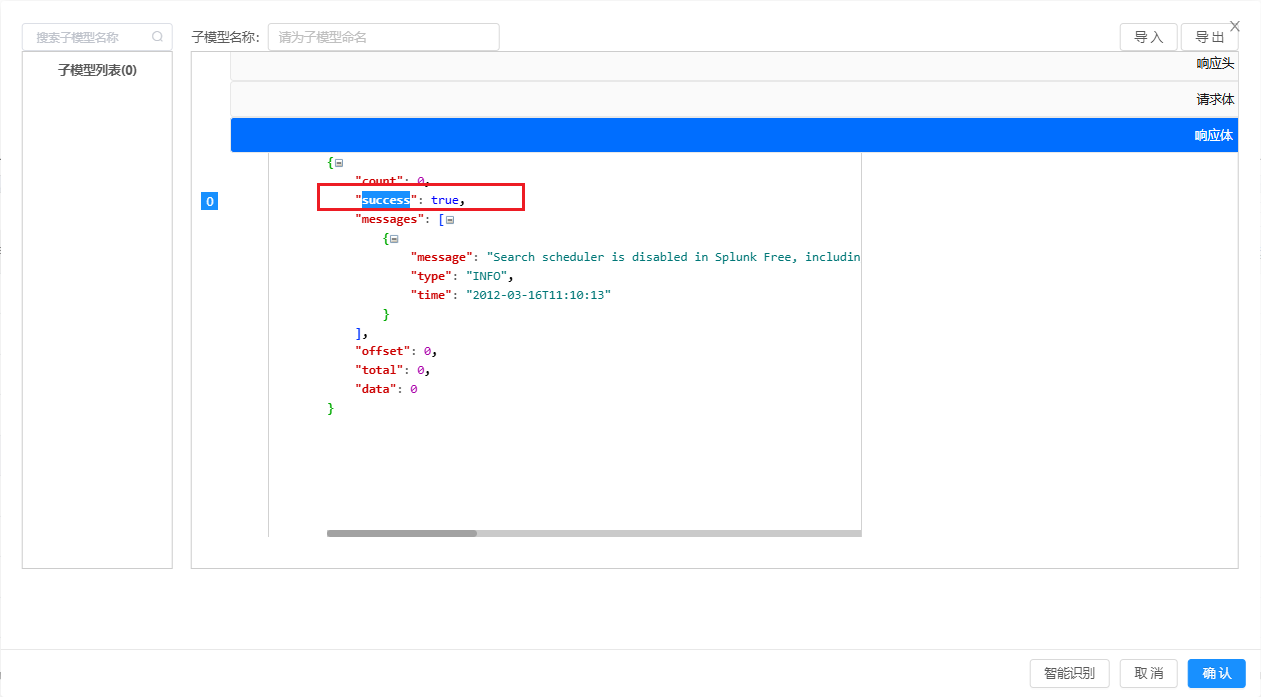


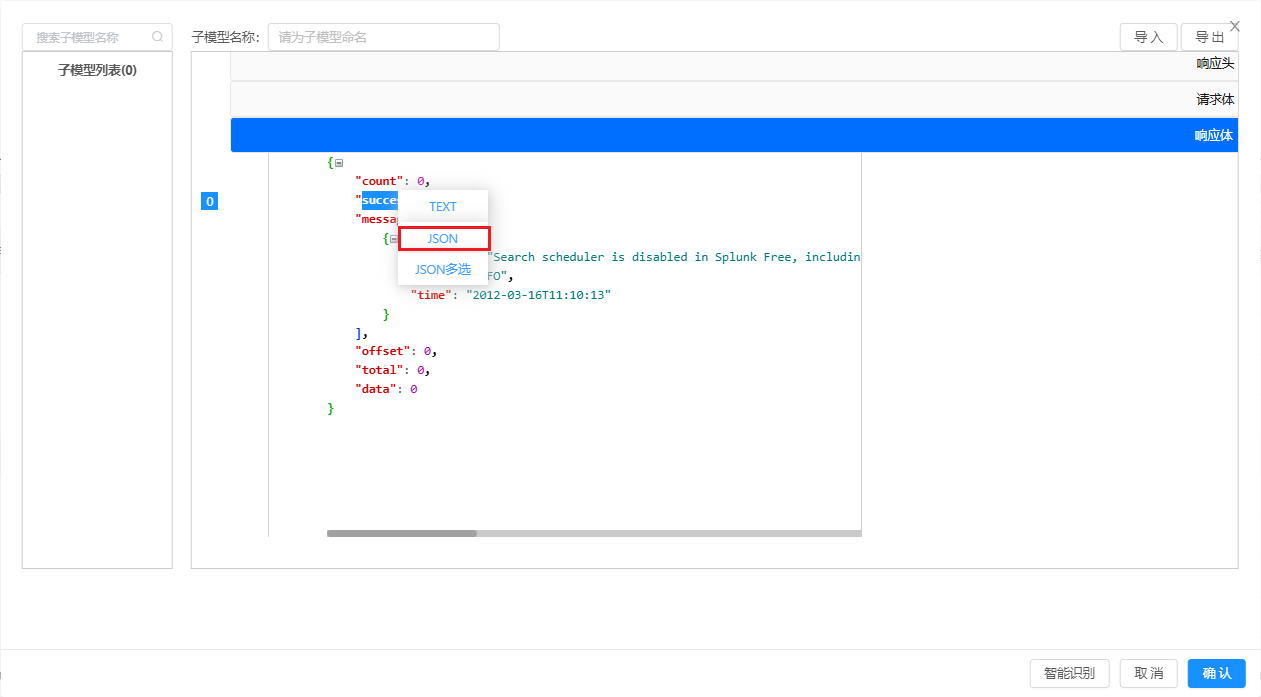
标签信息为空也可以

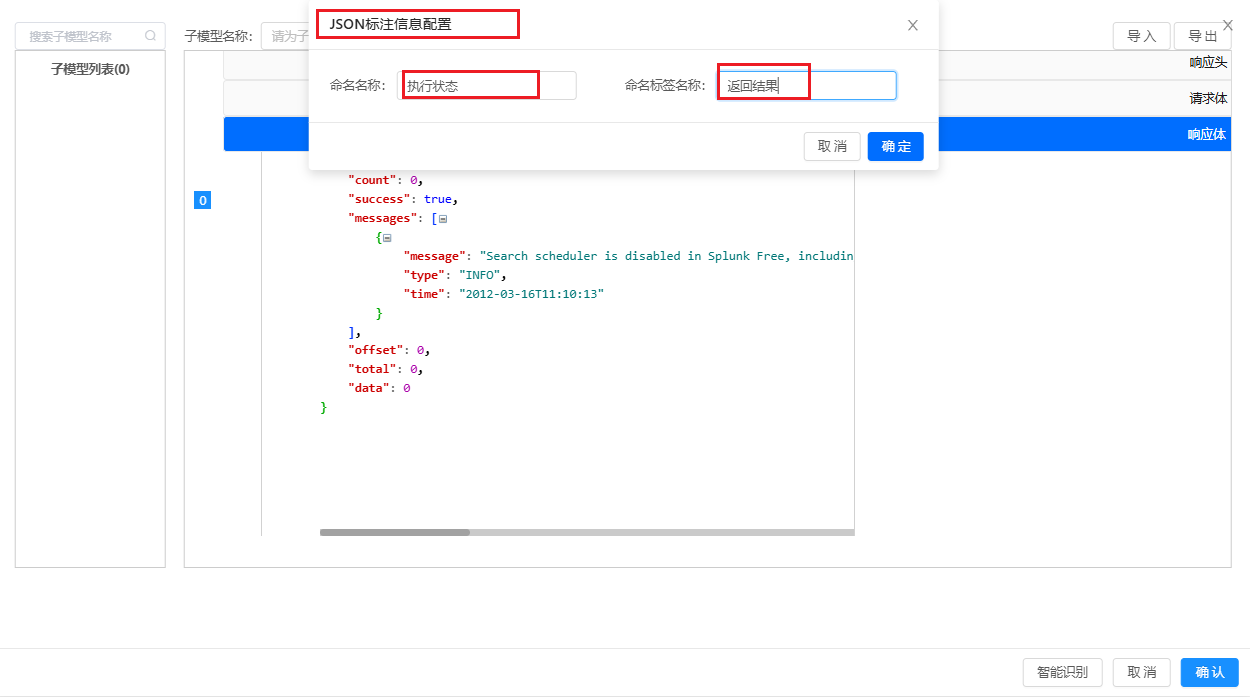


### ③数据标注

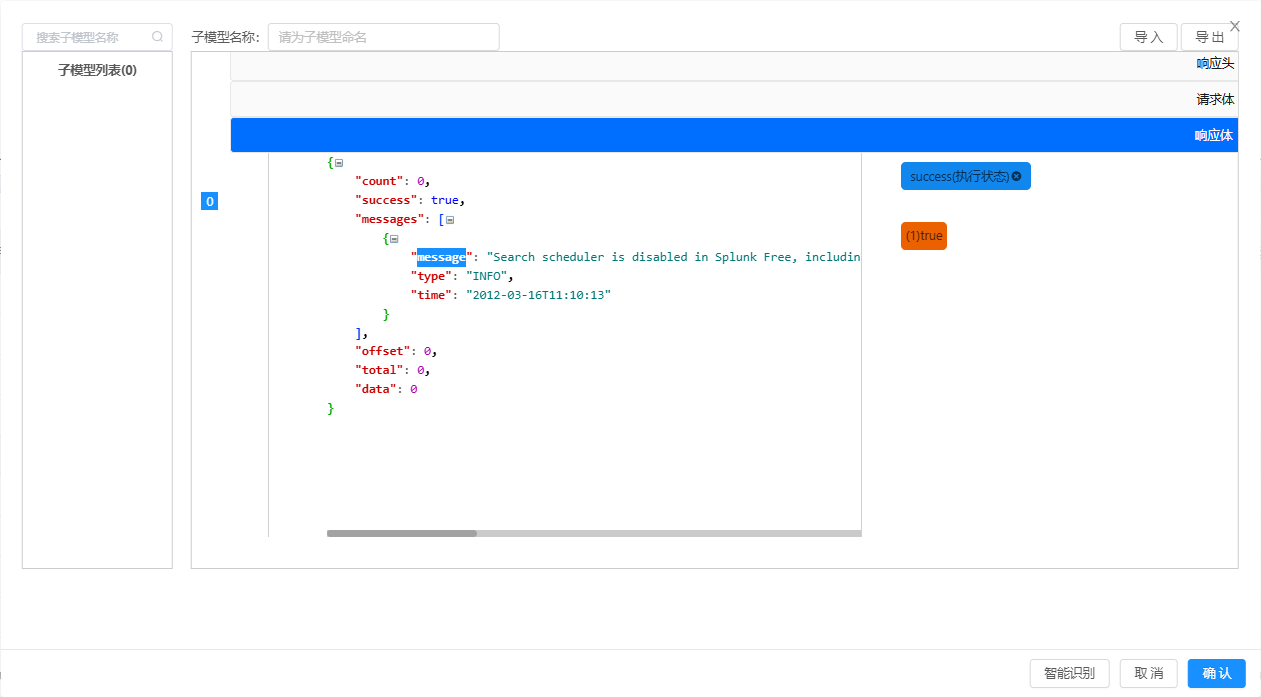
例如我们想要获取该HTTP请求的执行状态，标注“success”,选择JSON模式；

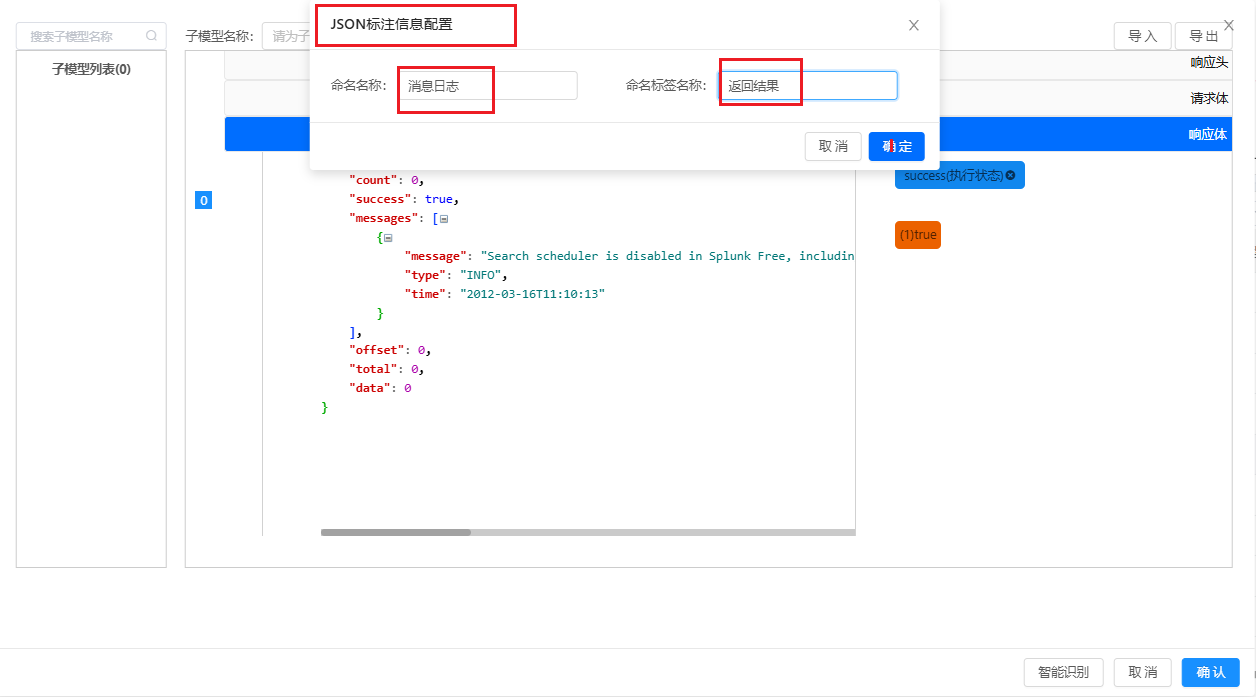




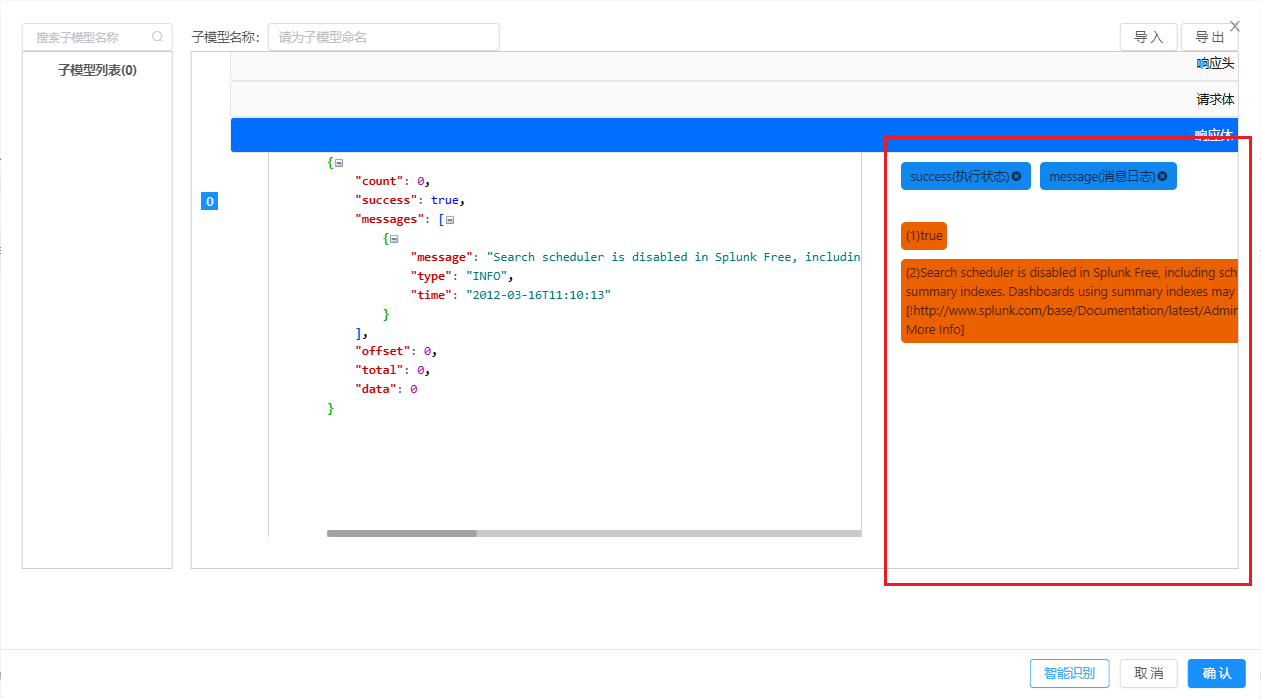


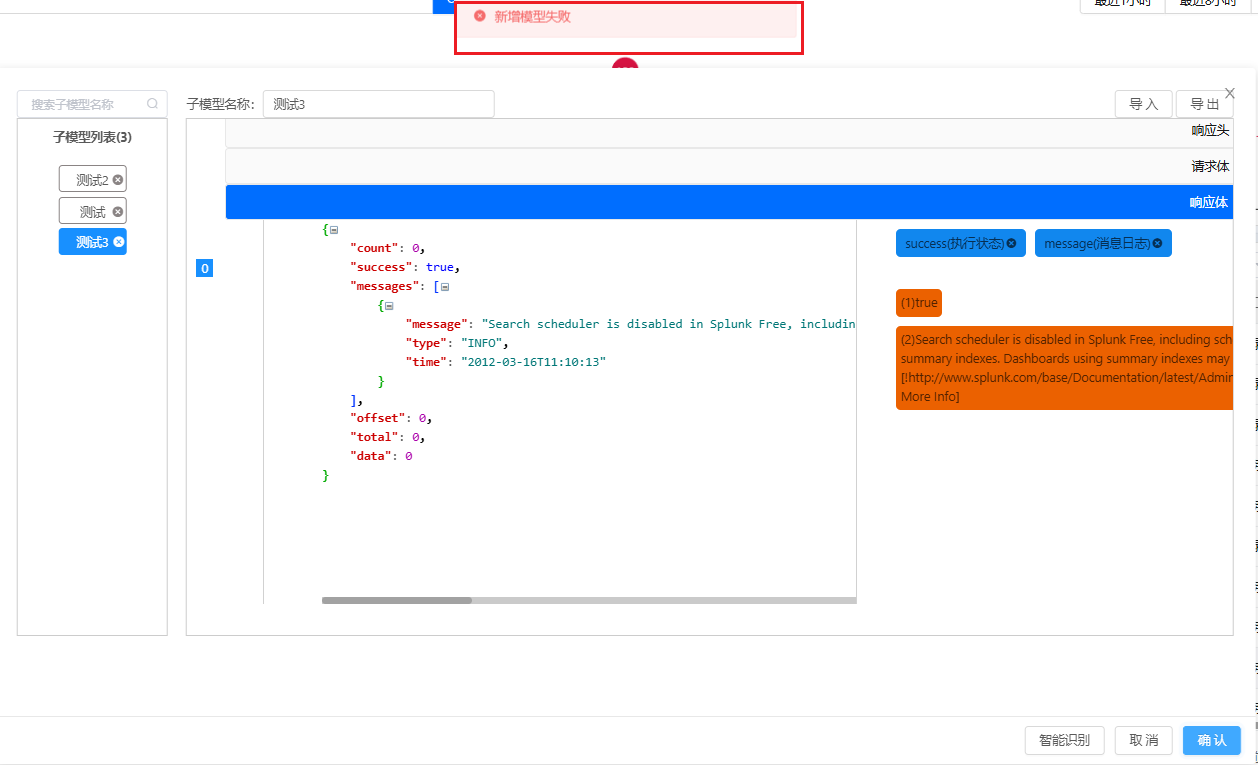
想要获取message信息，我们就可以标注message



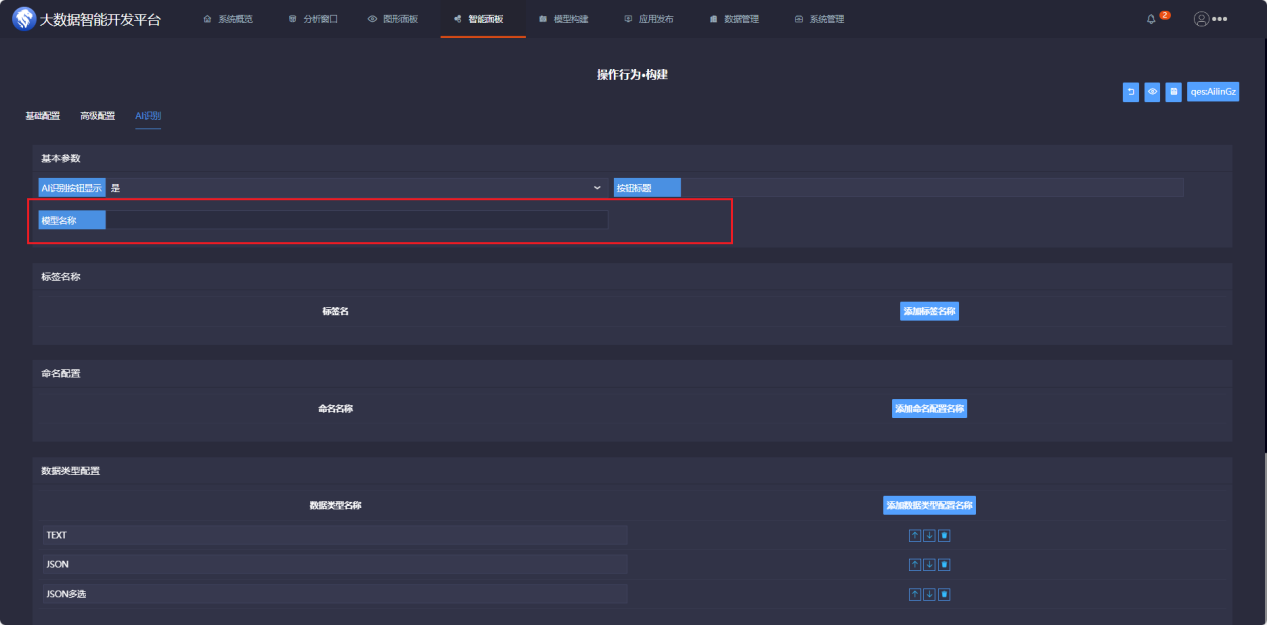


### ④保存数据结果

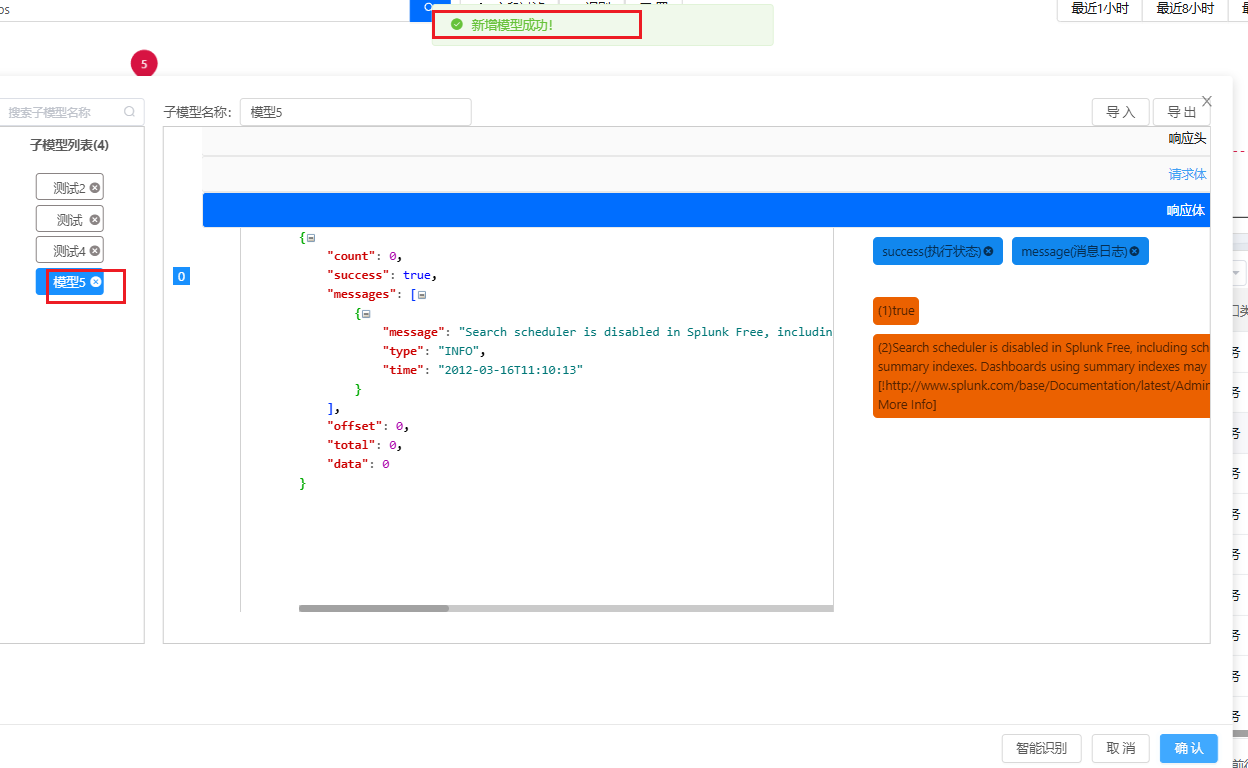




新增模型失败的原因是因为 我们后台没有配置模型名称

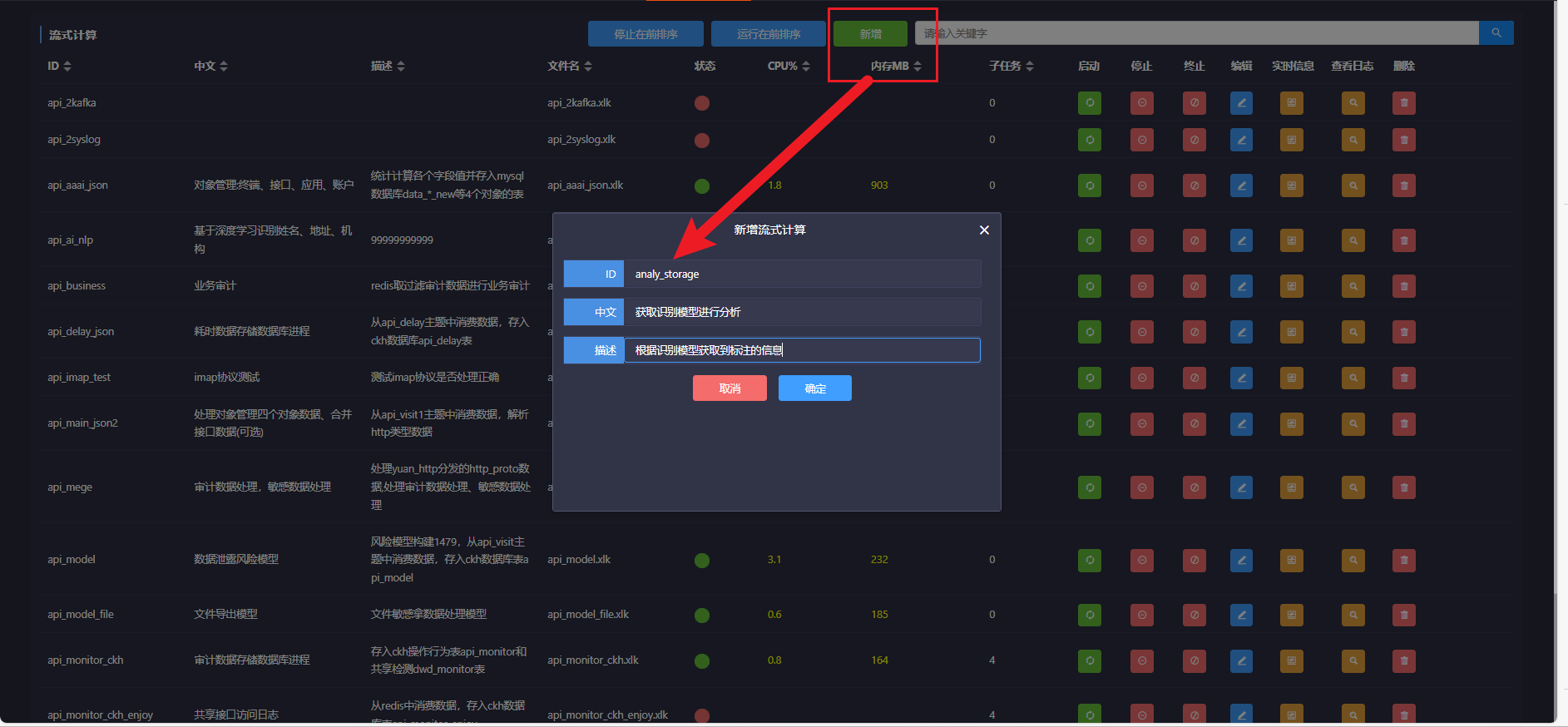


加上之后就会新增成功



# 开发人员使用模型

## 新增流式计算



## 导入识别模型的模块

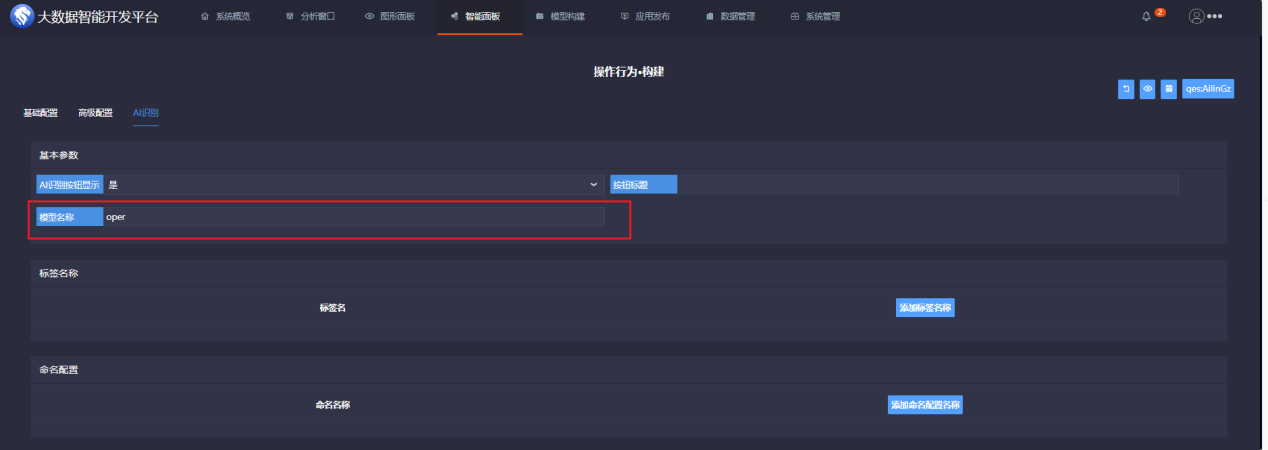
from intell\_analy\_new import \*



## 读取模型文件及识别操作



模型名称：oper 与上述读取的模型信息相同



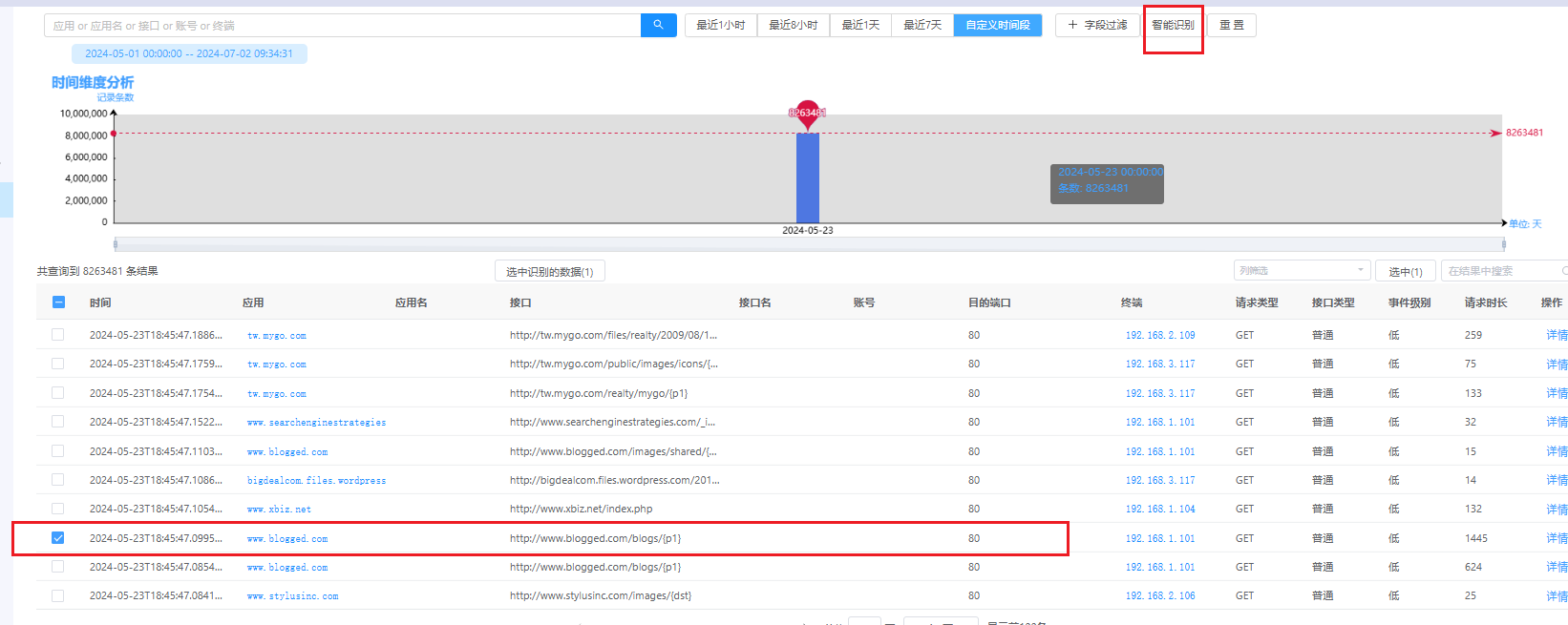
注：O是对应查询分析存储的数据字段，获取源数据的时候也应该跟识别模型的查询分析表相同；

# 关联模型

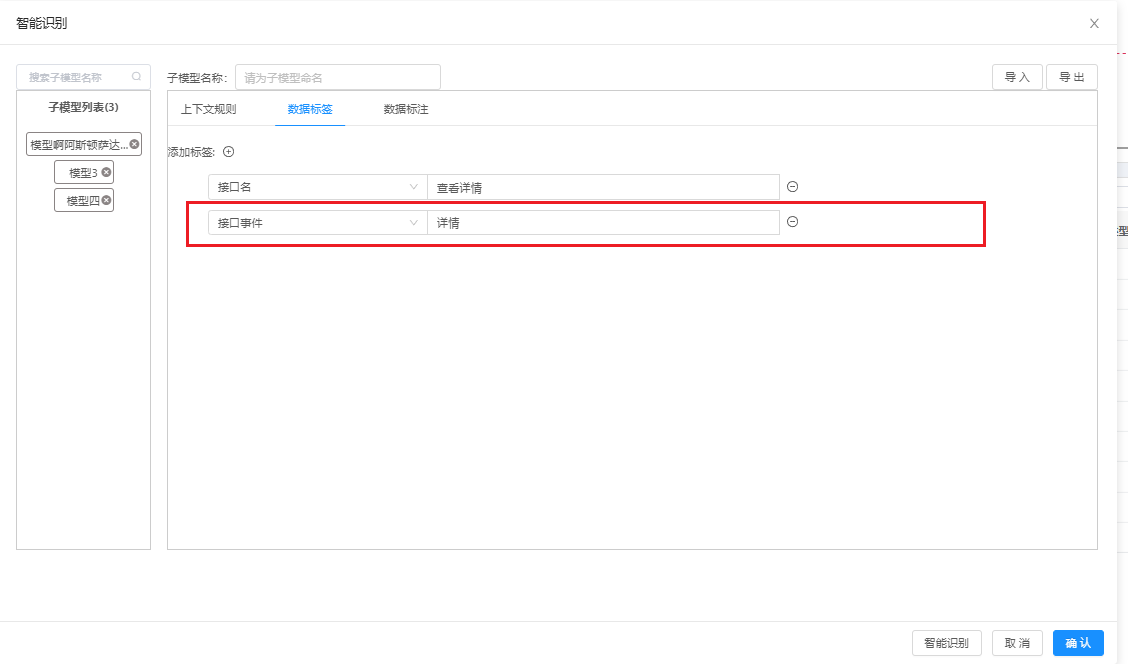
例如：拱墅区查看详情，会返回多数请求，但点击详情面板无法再次获取请求，因为系统站在快速情况下，不会一一查询返回。所以我们在点击详情时，就将多数请求进行存储，通过系统埋点，通过我们存储的请求数据，进行关联分析。

## 对查看详情请求进行标记

通过智能识别，找出详情接口，点击智能识别，生成当前模型，可以通过数据标签来标记该接口为详情；



这里设置一个“接口事件：详情”，用来标记详情接口；需要的话在进行数据标注；



## 详情接口存储

不同用户存在不同sessionID，获取到用户的唯一标识，通过模型识别的结果，获取数据标签的接口事件，去分析是否为详情接口；

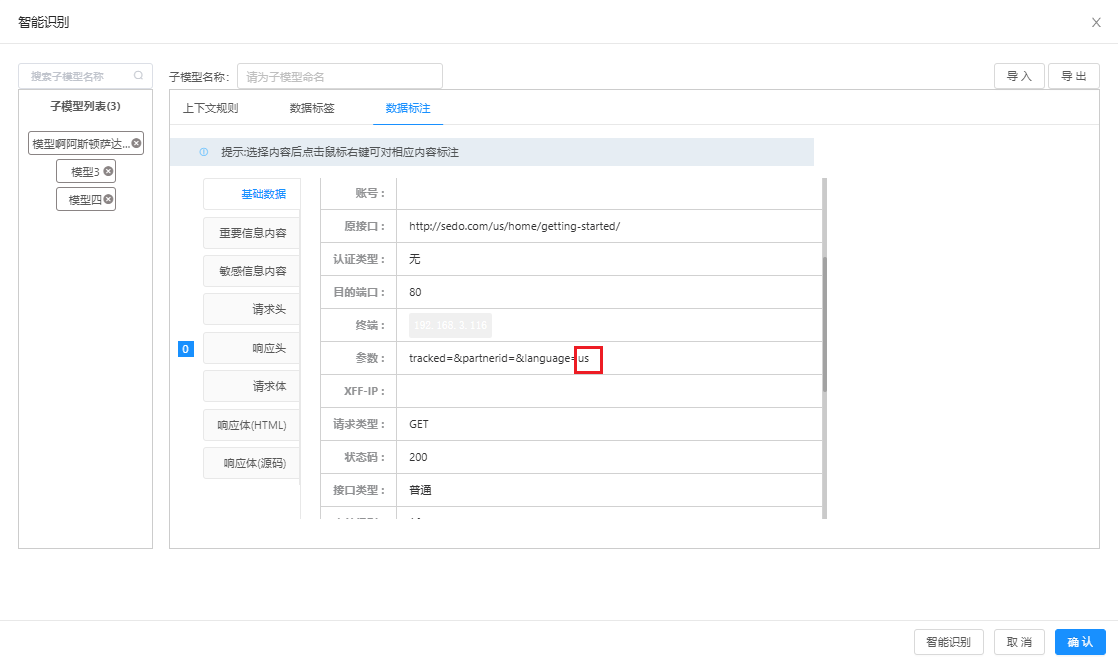
## 埋点关联详情接口

通过埋点的请求，获取到对应详情接口的标识，关联出用户操作行为；

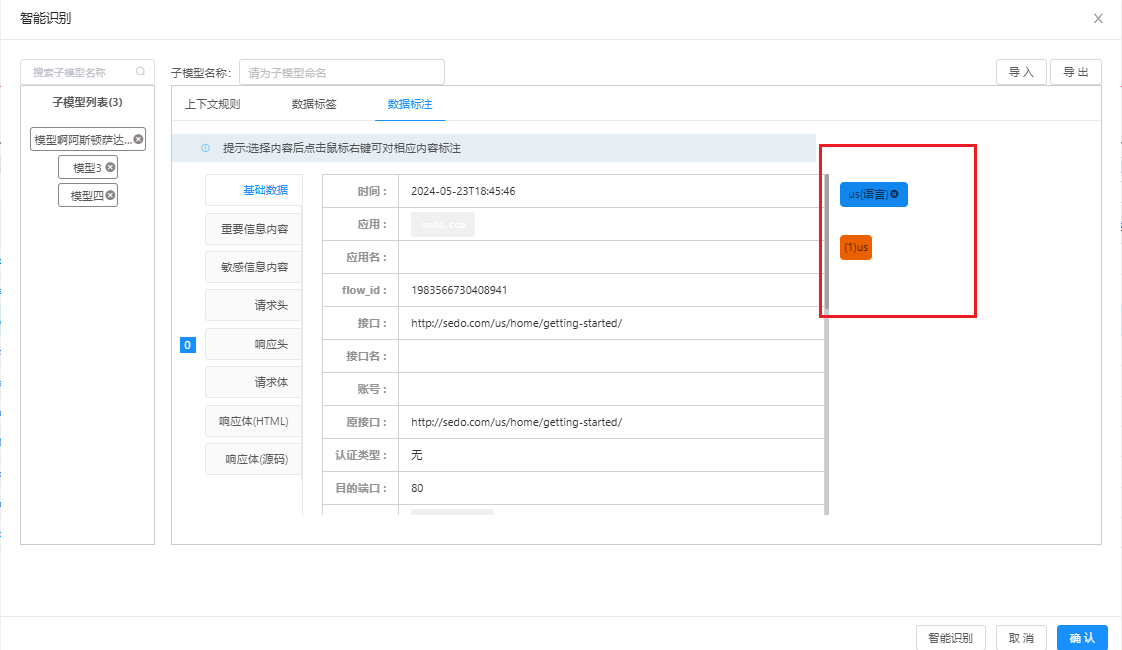
# 标识中出现的情况

## 1.中文转化问题

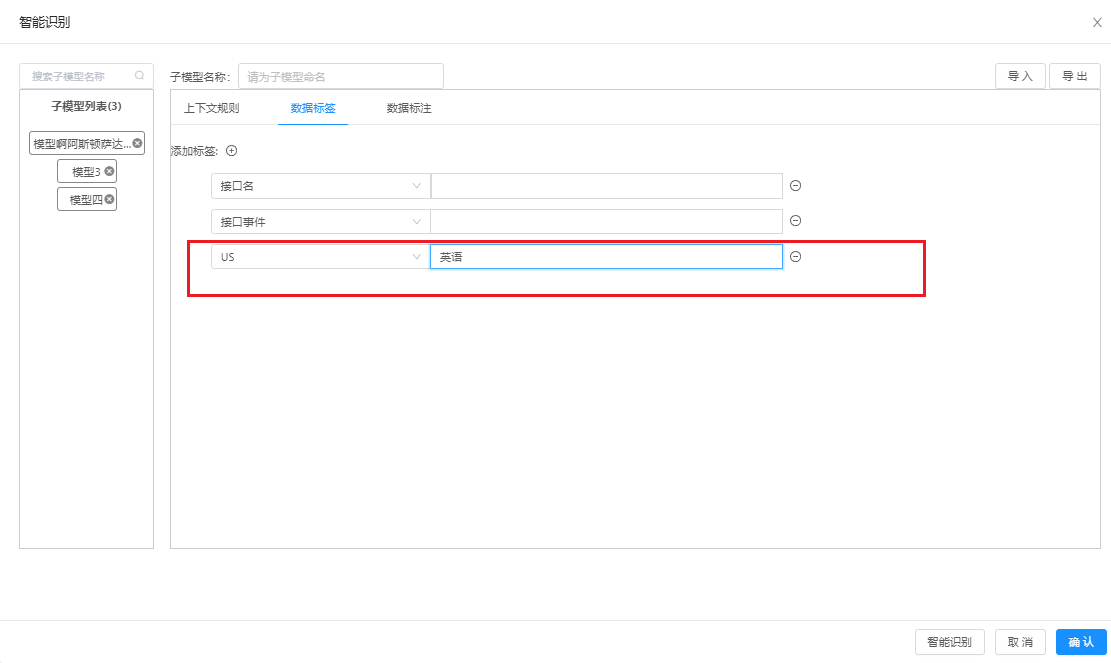
例:在标识中结果为 0 、1、2、true、false 等，识别的结果往往也会是这些，但是用户想要看的是显而易见的中文，因此返回0、1就不满足需求，下面是一个例子，想要知道us是什么语言；



标记出来的结果仍是 “us”

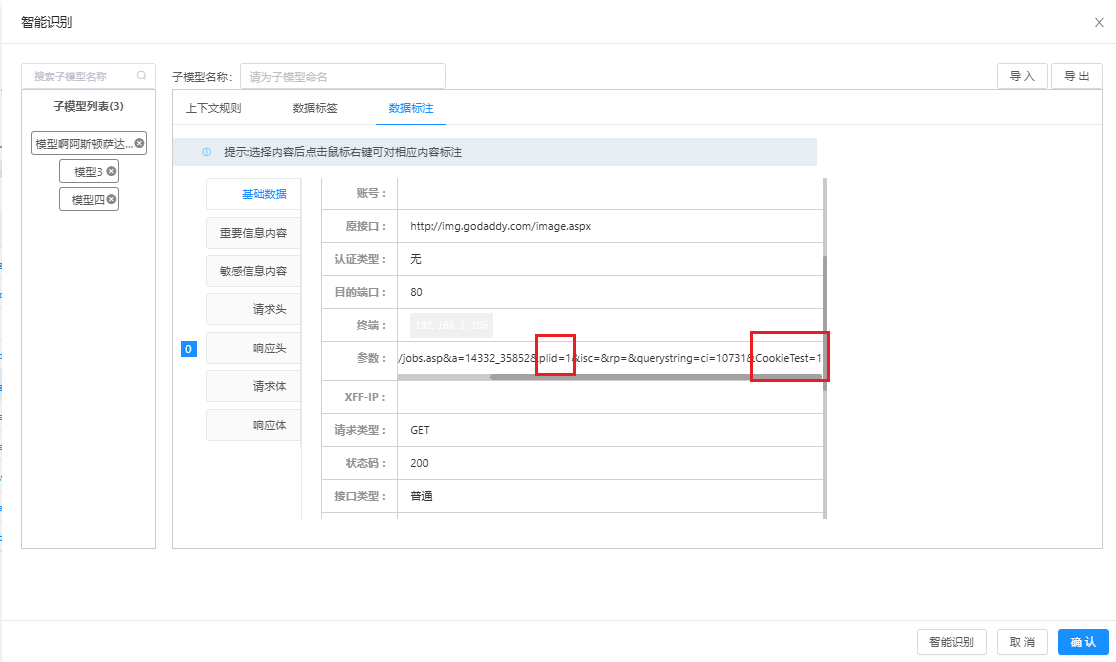


现在处理方法是在数据标签哪里添加结果并给出中文，但是这个代码中还需要做判断，类似这种情况未做自动化处理；



## 2.重复值问题

例如：想要获取下面图片中的plid=1 和 CookieTest=1



那我们标记出1这个值 我们获取到的是 a = 14332\_35852 中的1。所以在识别过程中我们就获取不到plid = 1 这个数据

