**盈智科技单据智能识别软件V1.0**

**使用说明**

1. 操作系统环境
2. 操作系统：本软件需运行在win10 64位系统或者Linux Centos7中，理论上其它Linux系统也可以支持。
3. 硬件配置：CPU：I5-3337U（或以上）；

内存：4G（或以上）；

硬盘：500G（或以上）；

接口：com口\*6，VGA\*1，HDMI\*1，USB3.0\*4，USB2.0\*4

1. 软件安装

本系统基于java语言开发，是一个基于Web浏览器的软件系统，并需要有python的支持，因此需要java和python的运行环境和一个web容器来运行软件，本软件选择spring boot作为开发框架，打包成jar文件自动内置web容器：tomcat。

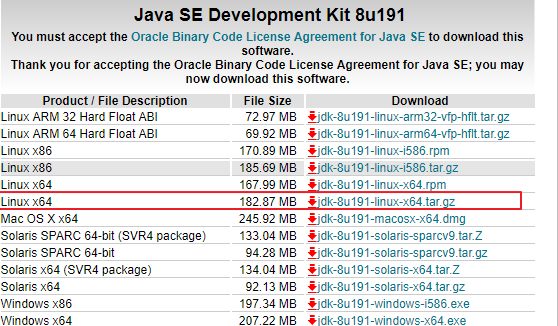
本系统对海运单据的各种类型，采用关键字标记定位的方式，识别并提取文件中的内容，并将内容以格式化的数据格式返回到第三方系统中，便于数据的自动录入和数据分析，节省了人工录入的工作，提高工作效率。

本系统支持常用的各种海运单据格式：word，excel，pdf，各种图片格式等。

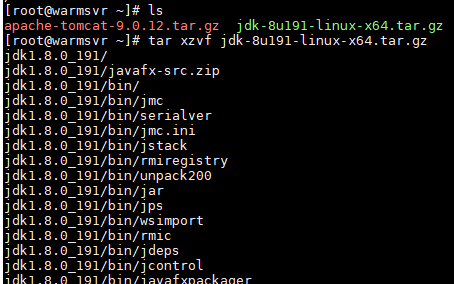
本文的介绍以Linux平台为例：

1. **安装JDK**

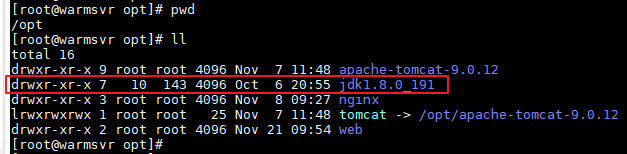
进入Oracle的官方网站下载JDK：<https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>



JDK为解压版，解压后需要配置环境变量即可使用。如将jdk安装到/opt路径下：

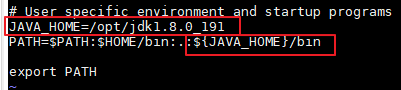


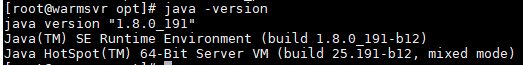
查看解压结果：



修改环境变量，将java的安装目录添加到path中,并通过java命令验证安装是否成功：

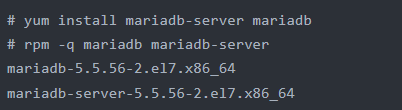






1. **安装数据库**

本软件采用开源的mariadb数据库存储软件需要的数据，通过 yum 方式安装 MariaDB ，执行如下命令即可。



安装完成后，将 MariaDB 设置为开机启动，操作如下：



1. **安装python**

采用Linux系统自带的Python



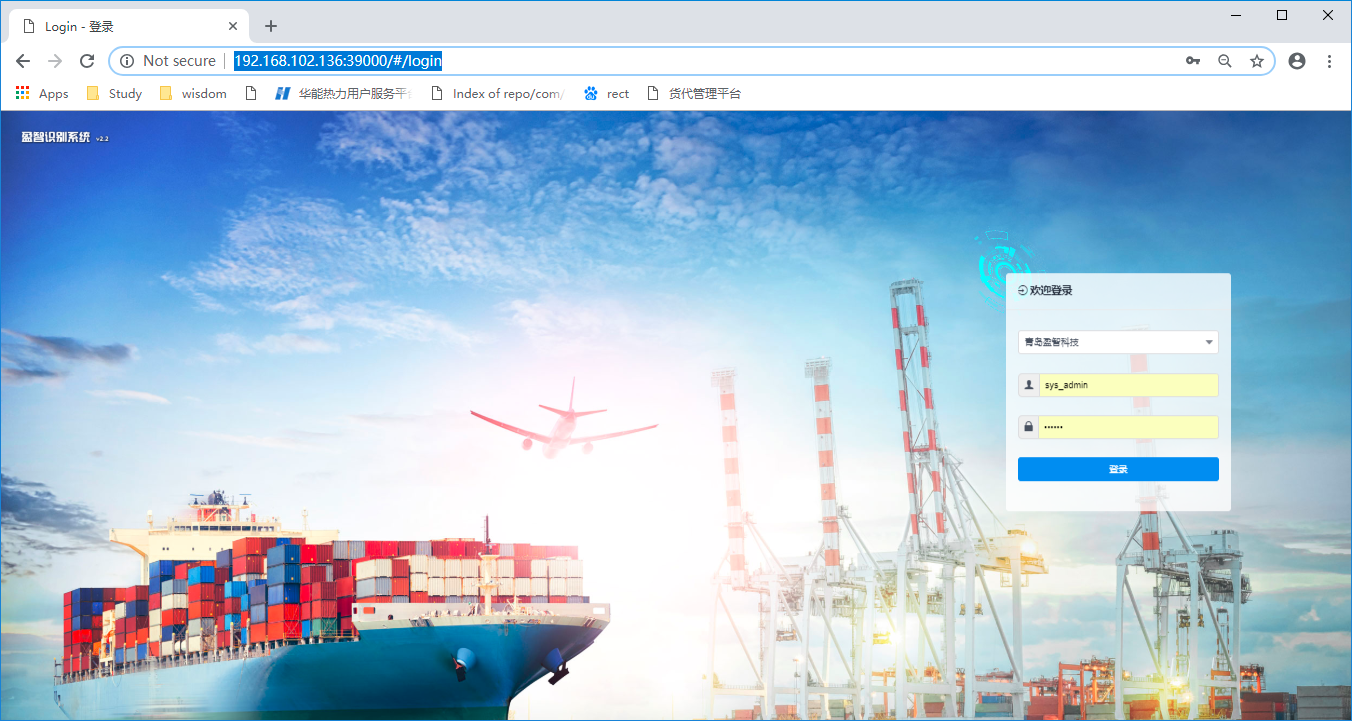
1. **安装软件**

本软件的发布格式为jar文件，直接运行即可，无需额外安装

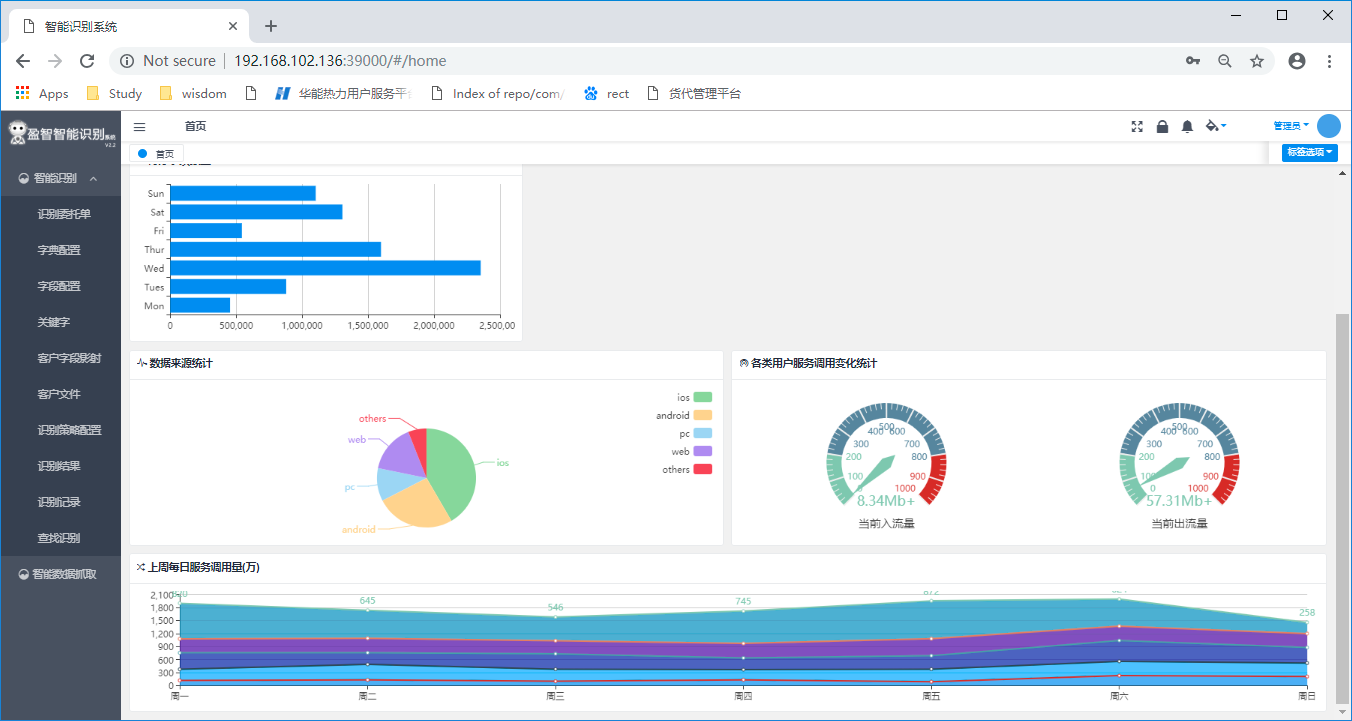


3、安装界面提示，按任意键退出安装程序，完成软件安装过程；

1. **系统设置**
   * + 1. 打开浏览器，输入URL：htt://{IP}/#/login登录单据智能识别系统，如下图：



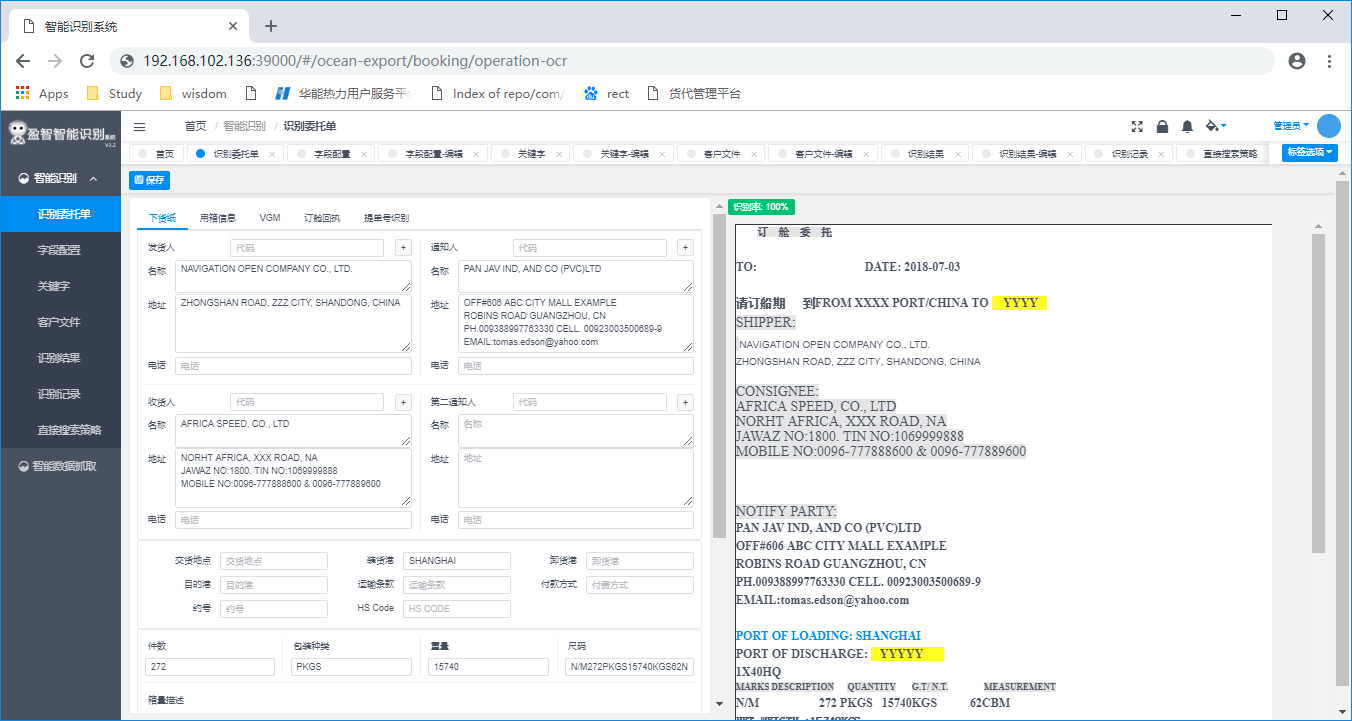
登录后主页显示如下，显示一些数据统计信息，主要功能在左侧下拉菜单中：



* + - 1. **识别委托单**

此处为智能识别系统的主窗口，窗口分为左右两个区域，右侧区域为识别区，将显示需要识别的文件原稿，左侧区域为识别结果显示区。

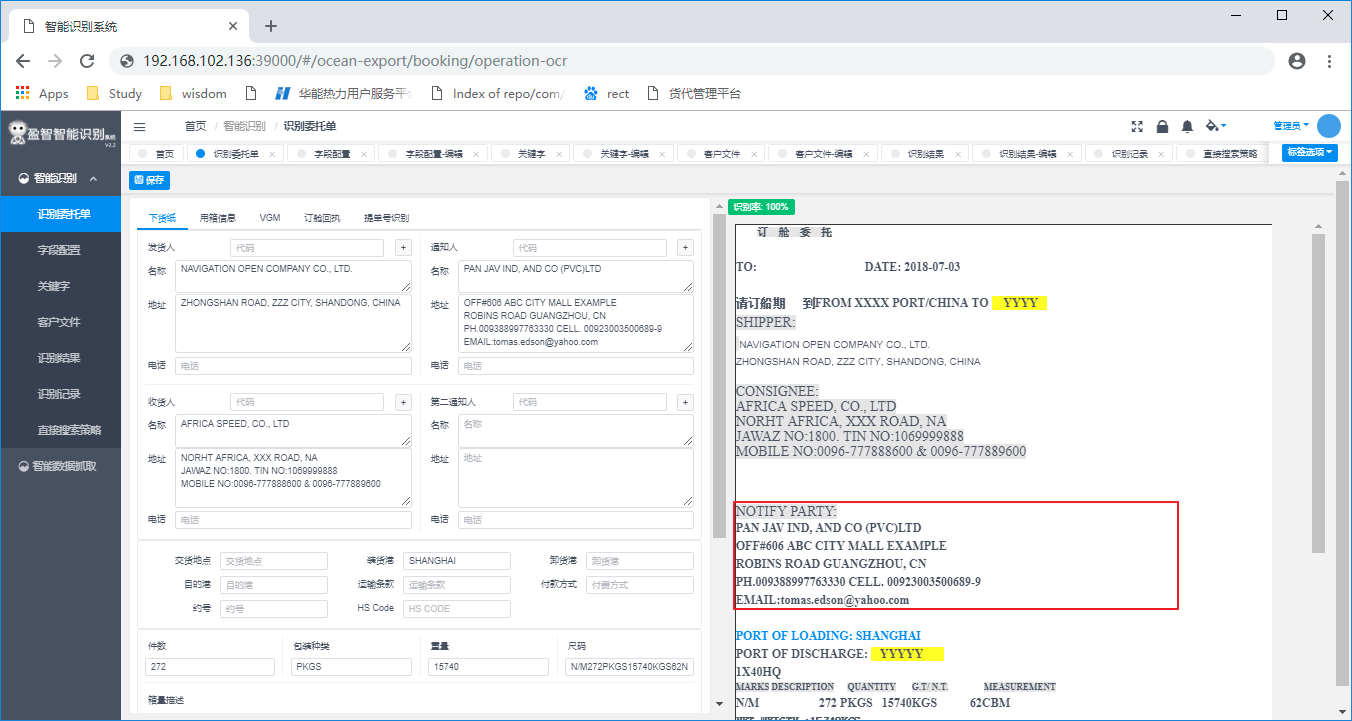
客户可以根据配置的字段，将文件拖拽到右侧识别区域，系统将自动识别文本内容，并将文本内容识别后，回显到左侧的区域，如下图所示：



可以看到识别后的结果为格式化后的内容，对于第三方的数据对接非常方便，可以将数据直接录入到第三方的业务系统。

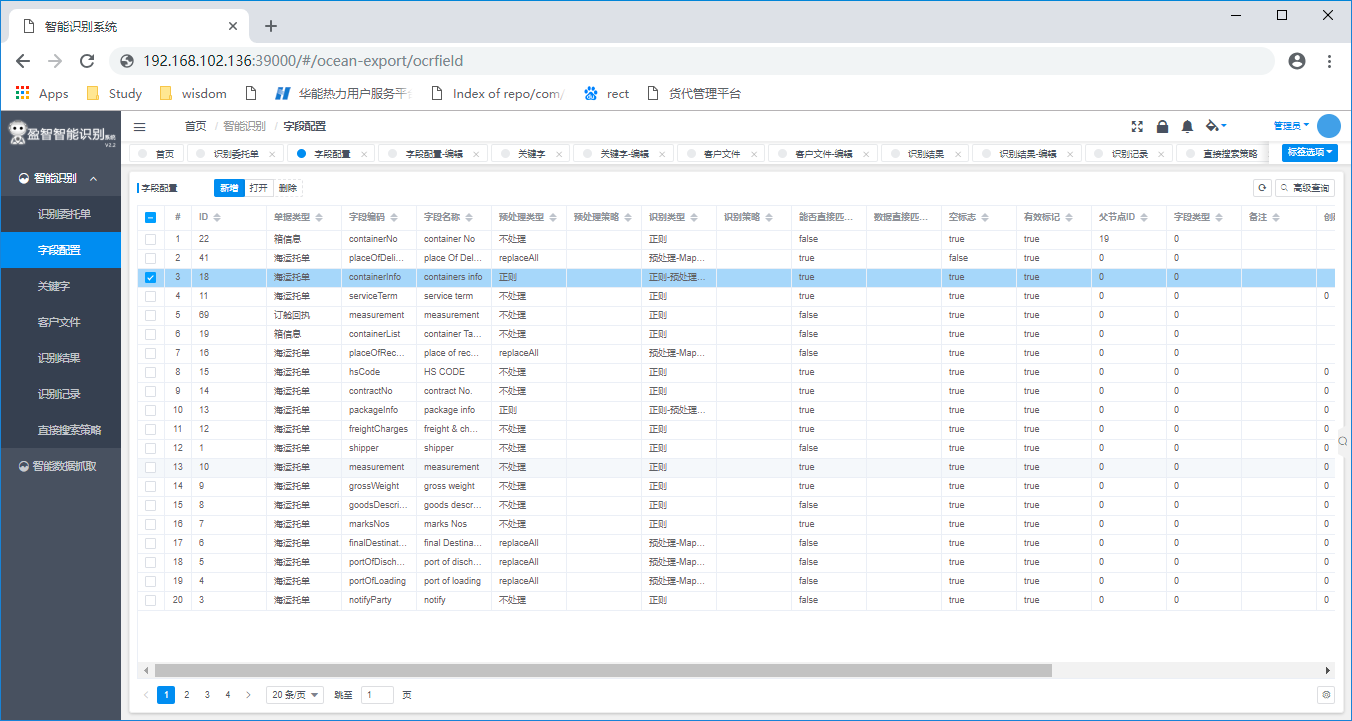
如果此时用户将鼠标点击左侧识别区域的框内，系统将自动显示此内容从原稿识别的源位置，如下图所示，点击左侧的通知人，右边的通知人会被红色框选，标识此通知人是从那个区域识别出来的，便于人工校对。

左侧区域中的内容如果有误，可以人工核对后修改，系统将自动保存修改后的内容以及原识别出来的内容，便于系统进行对内容的学习，后续可以做一些自动校对的工作，提高识别的准确率。

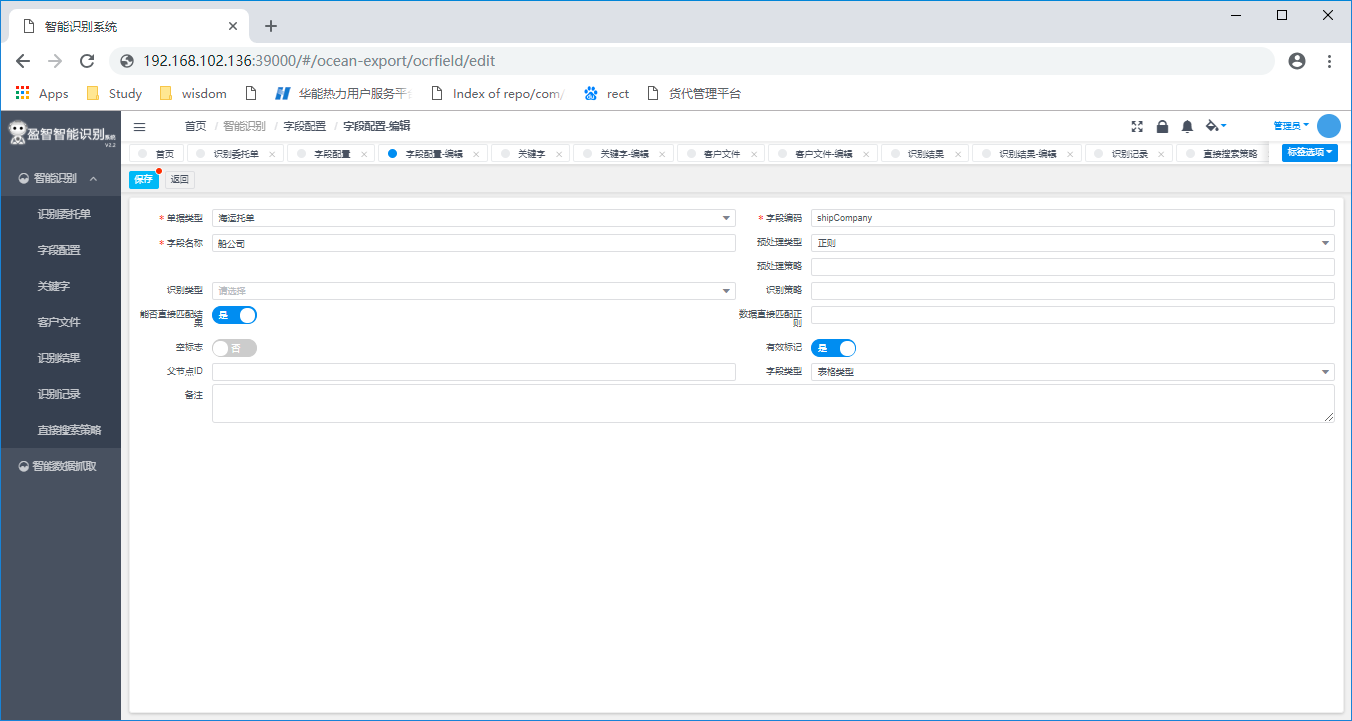


* + - 1. **字段配置**

此处配置客户需要从单据中识别出哪些字段，以及一些获取该字段的策略和采用的识别方法。



点击新增按钮，添加一条需要识别的字段：



输入需要的内容，关键输入内容如下：

单据类型：系统会根据不同的单据类型采用不同的识别策略，目前系统支持海运托单，各个船公司的订舱回执单，扫描件图片，订舱截单等单据类型。通过下拉选择不同的单据类型。

字段编码：返回到客户端时的数据的key，如：grossweight：16909KGS ，字段编码为grossweight

字段名称，为该字段取的别名。

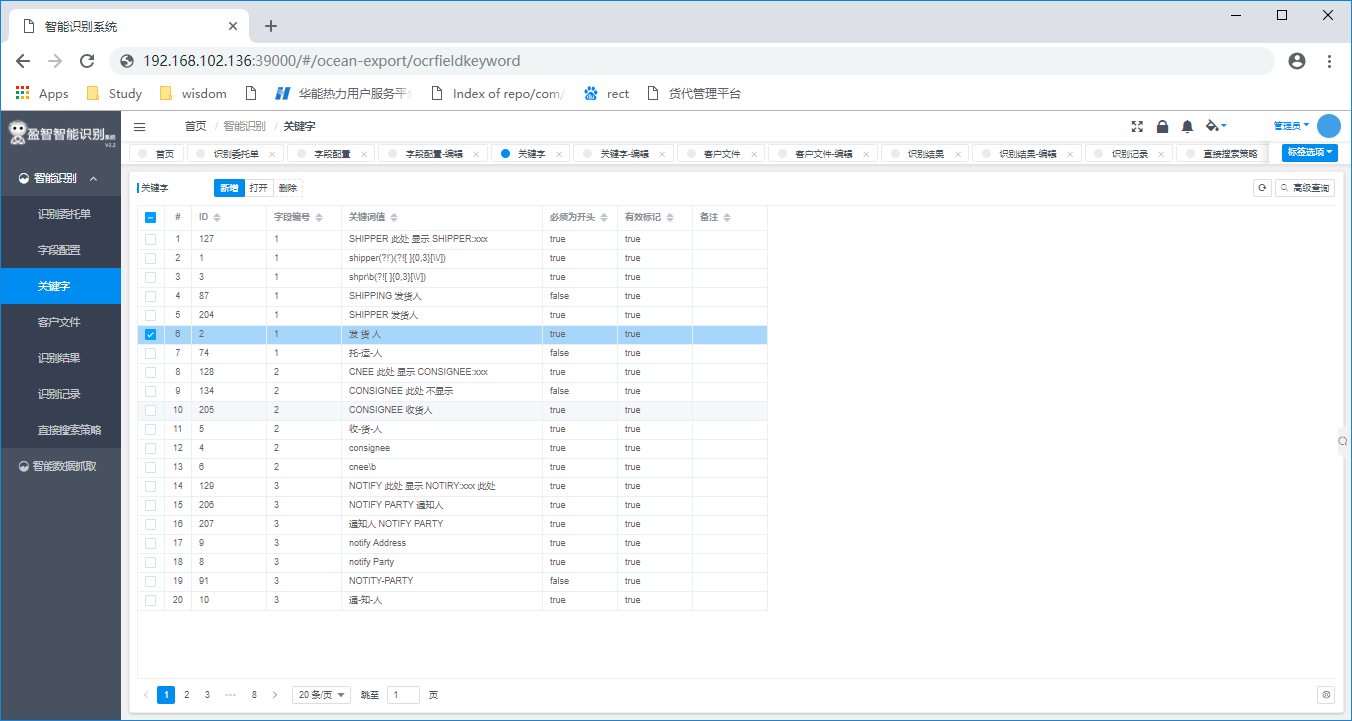
预处理类型：对数据进行预处理的一些规则配置，可选项为：0-不处理 1-正则 2-split 3-replaceAll

预处理策略：对一些固定内容的值做一些策略，或者是一些关键字范围的定义

识别类型：识别的规则策略配置，如可选择： 0-正则 1-map 2-预处理-Map比对 3-正则-预处理-MAP比对

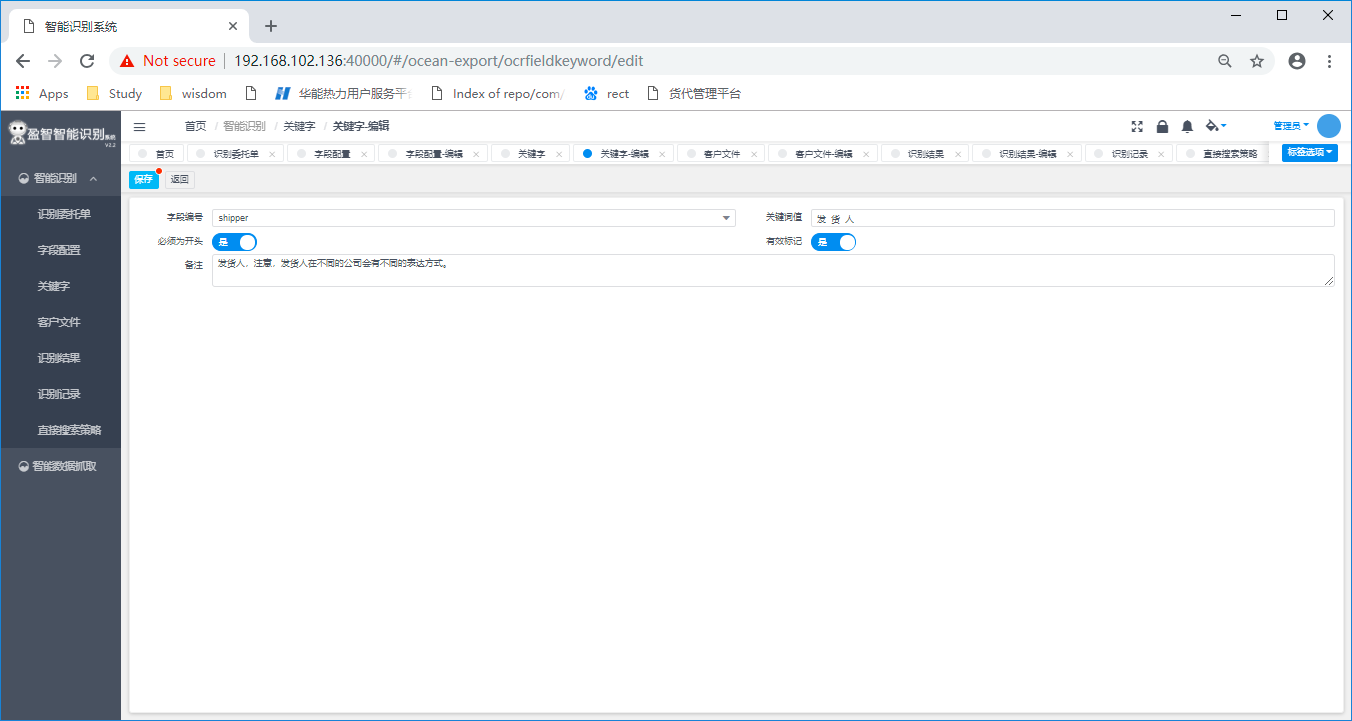
按照界面提示逐个输入字段的信息，为系统的文字识别设置规则。

* + - 1. **关键字配置**



如前所讲，本系统是根据关键字来划分区域，区别各个字段，因此关键字配置对系统来说很重要。

点击新增关键字：



字段编号：标识此关键字是为了识别哪个字段而配置的，通过下拉框选择字段

关键词值：标识关键字，比如对发货人的关键字可能有：shipper， shpr，发件人，发 件 人等各种格式

必须为开头：标识是否此关键字要求以某个值开头。

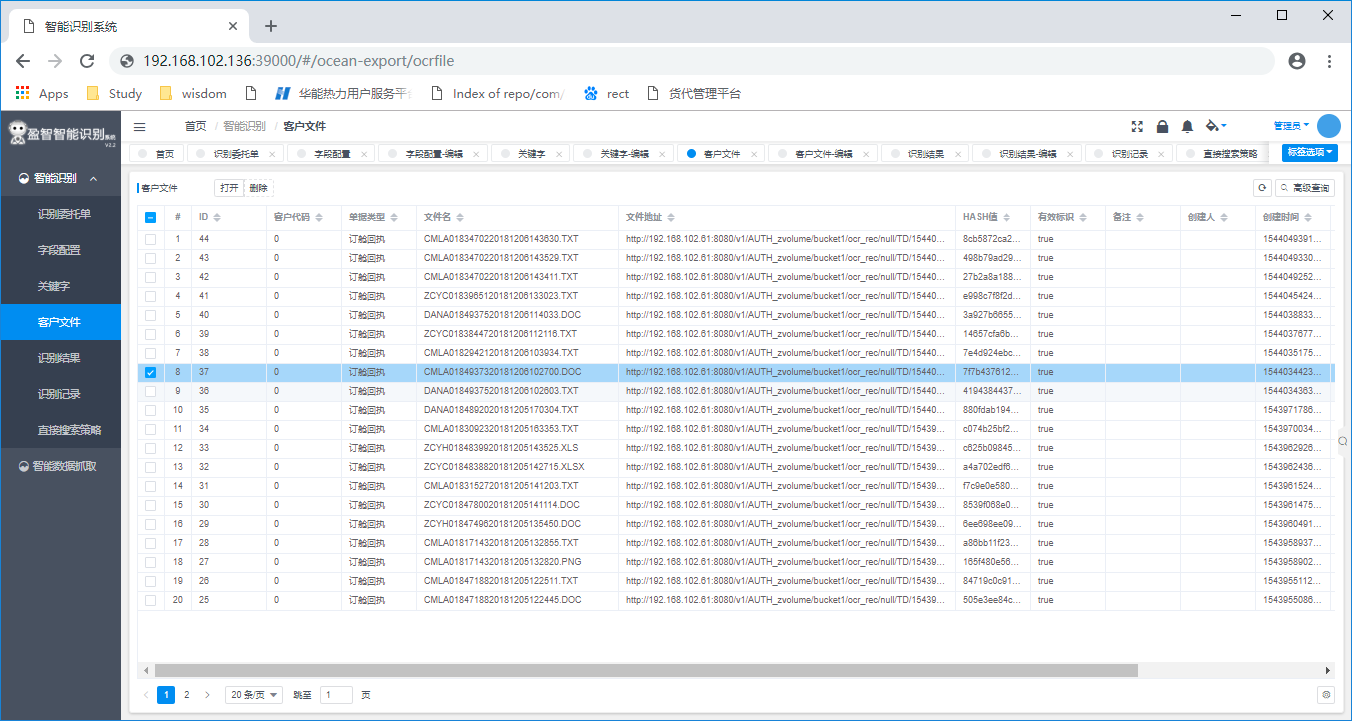
依次为每个字段配置可能的关键字。

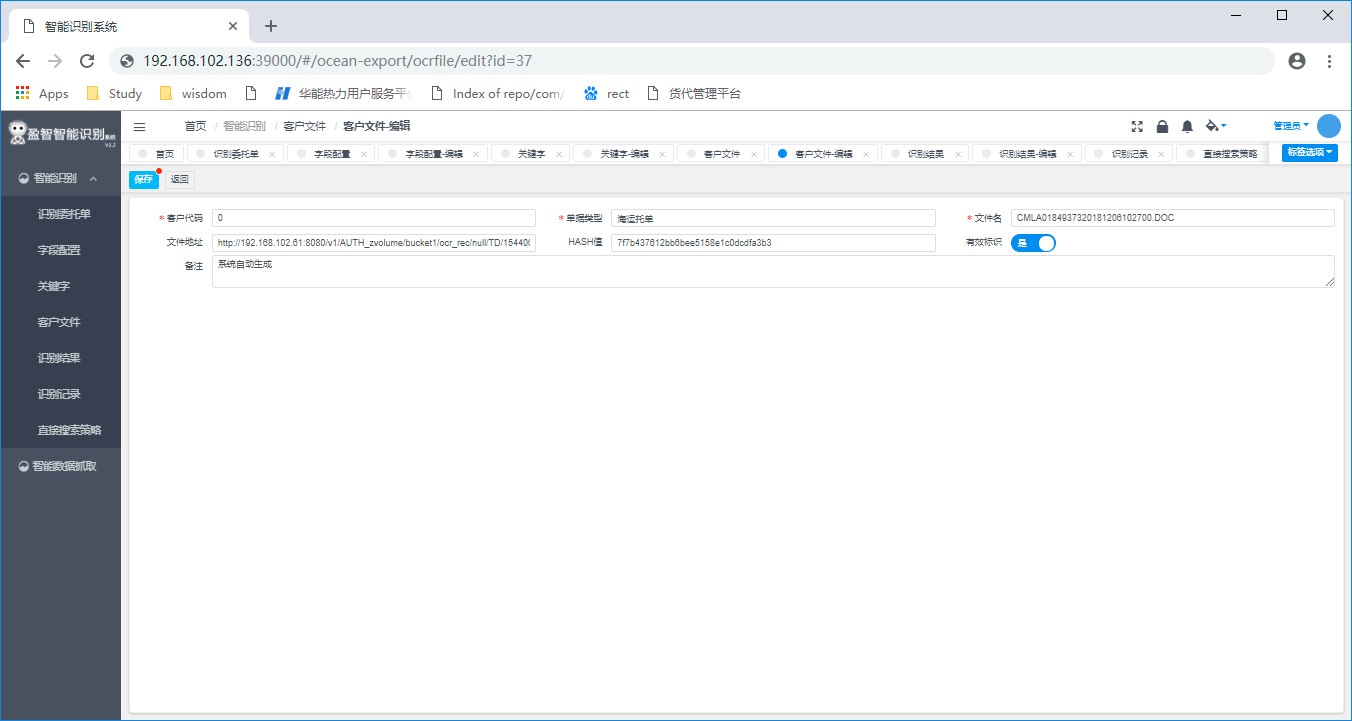
* + - 1. **客户文件**

系统会对不同客户上传的文件，以及文件的识别结果做存档处理，主要目的是：

1. 去重，如果新上传文件的hash值匹配到了某个以及存在的文件的hash，则此文件被认为以前识别过，直接取出结果返回到界面，无需重复解析
2. 对文件和识别结果做存档，便于后期对识别结果做进一步的分析，提高识别的准确度。

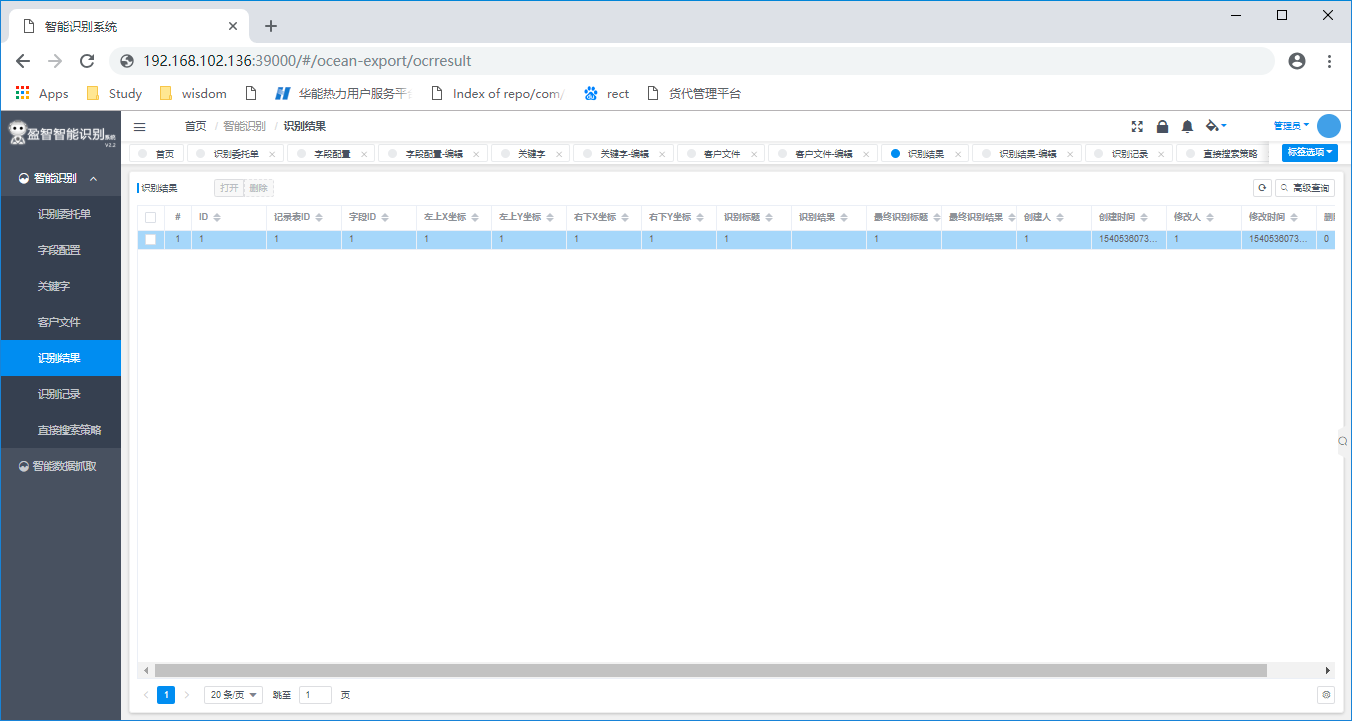
客户文件为系统识别过程中自动存档，无添加按钮，但可以查看文件的相关信息，如文件名，Hash等。





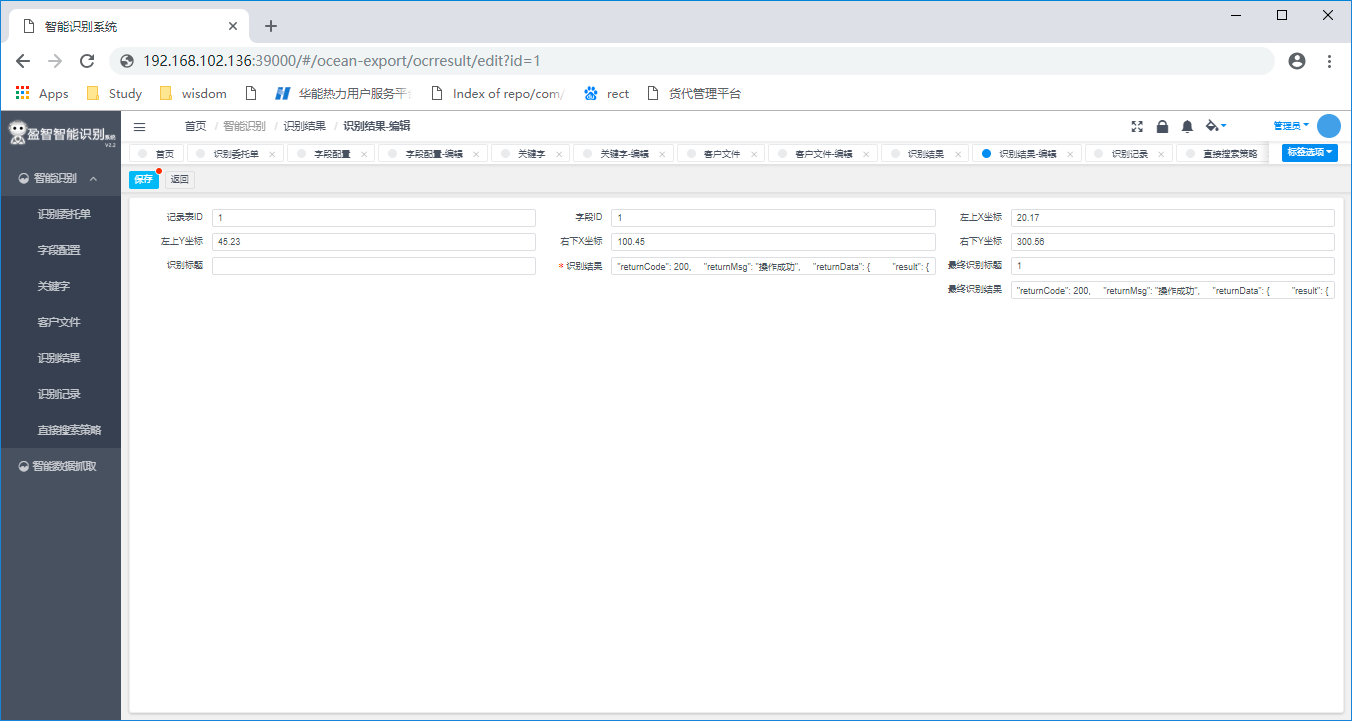
* + - 1. **识别结果**

对于每个文件的识别结果，系统会自动存档，并对校对结果一并做保存，这样便于在修改算法后，对先前识别的文件做自动化测试，检查是否新的算法修改影响到了先前的某些文件的识别。



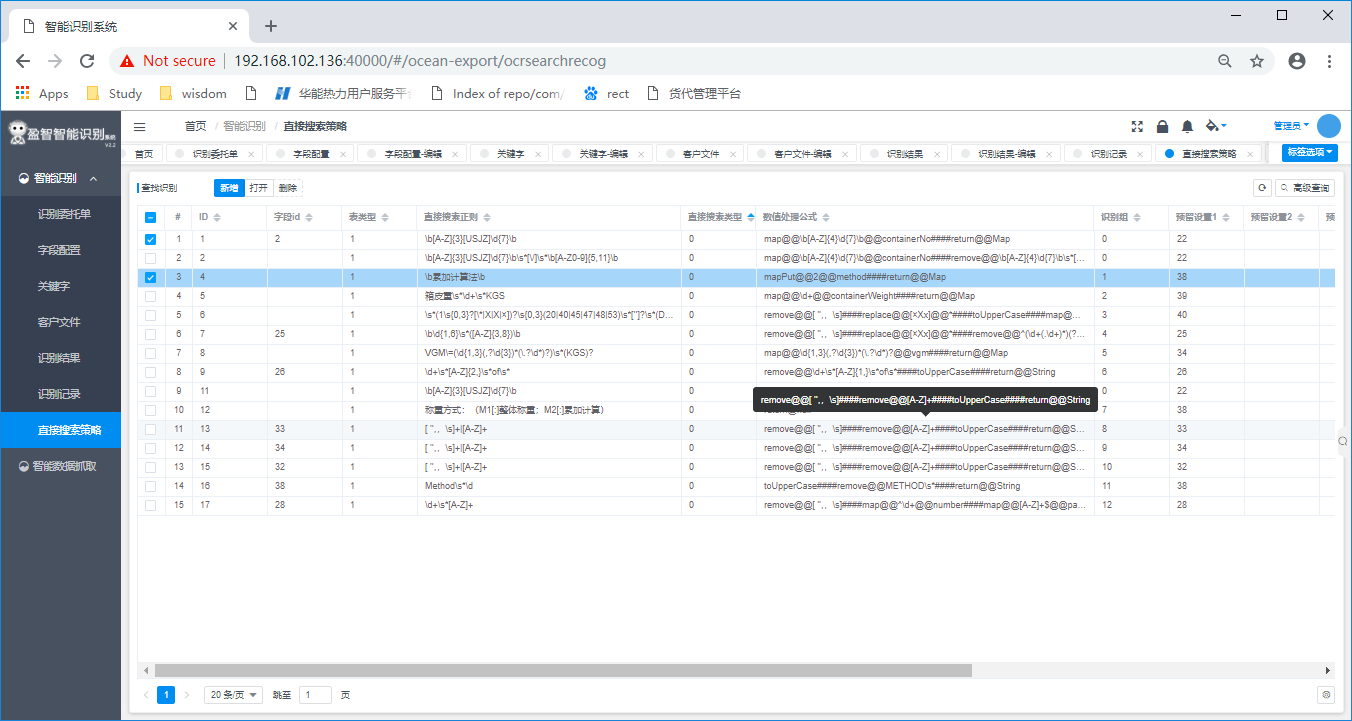
可以打开记录查看具体识别结果的具体内容，对于结果字段因为内容较多，无法在页面全部展示，建议全选+拷贝到其它的工具中查看具体内容。

目前返回的数据格式为JSON格式，因此建议拷贝到可以做JSON格式化的工具中查看，这里建议直接在网站json.cn 在线查看。

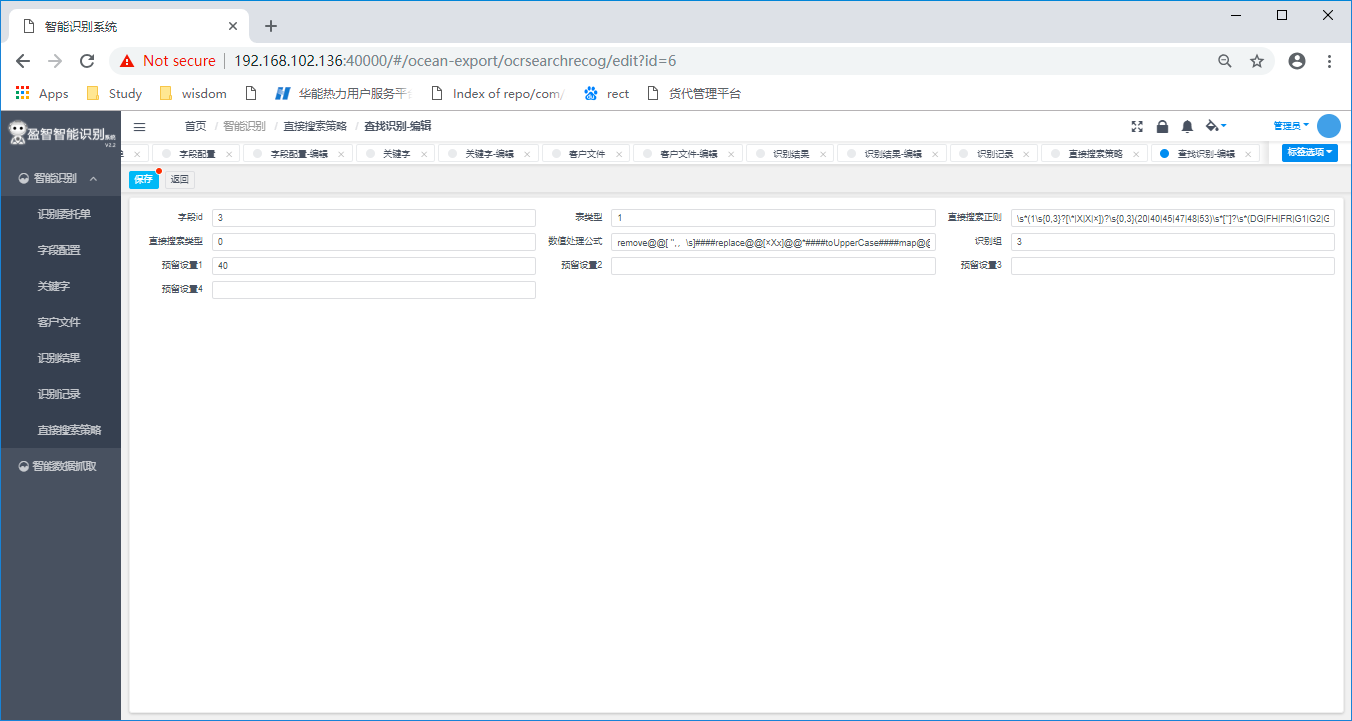


* + - 1. 直接搜索策略

这里对一些需要全文检索的搜索字段配置一些搜索规则。



比如以下规则将在全文中匹配跟Container weight相关的关键字，并抓取其重量的值：



1. API对接

系统支持第三方系统通过RestfulAPI的方式调用系统接口，并返回格式化的数据，以便第三方系统对数据做自动化录入等操作。

目前系统支持以下接口：

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 接口 | 输入参数 | 返回值 |
| 1 | http://ip:port/base/api/reco/sta | File | {  "returnCode": 200,  "returnMsg": "操作成功",  "returnData": {  "result": {  "bookingNumber": "HASLNR60B800CH00",  "receiptPlace": null,  "loadingPort": "QINGDAO",  "transshipmentPorts": [],  "dischargePort": "BANGKOK",  "deliveryPlace": "BANGKOK",  "destinationPlace": "BANGKOK",  "commodity": null,  "vesselVoyages": [  {  "vessel": "NORDLILY",  "voyage": "1810S"  }  ],  "containerTypeNumbers": [  {  "cntType": "40RH",  "cntNumber": "6",  "cntRemark": null  }  ],  "info": {  "packages": "16800",  "grossWeight": "168000.000",  "measurement": "300.000",  "temperature": null,  "ventilation": null,  "humidity": null  },  "containerReleaseTime": null,  "yard": null,  "payMethod": "PREPAID",  "agreementNumber": null,  "carriageCondition": null,  "specialHandling": null,  "onBehalfOf": null  },  "custNo": "TBN",  }  } |
| 2 | http://ip:port/base/api//ocrOperation | File | {  "returnCode": 200,  "returnMsg": "操作成功",  "returnData": {  "orderNo": null,  "fieldData": {  "shipper": {  "top": 117,  "bottom": 143,  "left": 0,  "right": -1000,  "width": 0,  "height": 1,  "text": "\n\nSHIPPER:\n\n NAVIGATION OPEN COMPANY CO., LTD. \n\nZHONGSHAN ROAD, ZZZ CITY, SHANDONG, CHINA",  "keyStr": "\n\nSHIPPER:\n\n",  "valueStr": "NAVIGATION OPEN COMPANY CO., LTD. ",  "kvMap": {},  "y": null,  "x": null,  "type": "KV",  "fieldId": 1,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": true,  "validKey": true,  "pointer": false  },  "consignee": {  "top": 195,  "bottom": 221,  "left": 0,  "right": -1000,  "width": 0,  "height": 1,  "text": "\n\n\n\nCONSIGNEE:\n\nAFRICA SPEED, CO., LTD\n\nNORHT AFRICA, XXX ROAD, NA\n\nJAWAZ NO:1800. TIN NO:1069999888\n\nMOBILE NO:0096-777888600 & 0096-777889600",  "keyStr": "\n\n\n\nCONSIGNEE:\n\n",  "valueStr": "AFRICA SPEED, CO., LTD",  "kvMap": {},  "y": null,  "x": null,  "type": "KV",  "fieldId": 2,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": true,  "validKey": true,  "pointer": false  },  "contractNo": {  "top": null,  "bottom": null,  "left": null,  "right": null,  "width": null,  "height": null,  "text": null,  "keyStr": "",  "valueStr": "",  "kvMap": {},  "y": null,  "x": null,  "type": null,  "fieldId": null,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": false,  "validKey": true,  "pointer": false  },  "shipAddr": {  "top": 117,  "bottom": 143,  "left": 0,  "right": -1000,  "width": 0,  "height": 1,  "text": "\n\nSHIPPER:\n\n NAVIGATION OPEN COMPANY CO., LTD. \n\nZHONGSHAN ROAD, ZZZ CITY, SHANDONG, CHINA",  "keyStr": "\n\nSHIPPER:\n\n",  "valueStr": "ZHONGSHAN ROAD, ZZZ CITY, SHANDONG, CHINA",  "kvMap": {},  "y": null,  "x": null,  "type": "KV",  "fieldId": 1,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": true,  "validKey": true,  "pointer": false  },  "portOfDischarge": {  "top": 636,  "bottom": 662,  "left": 0,  "right": -1000,  "width": 0,  "height": 1,  "text": " \n\nPORT OF DISCHARGE: YYYYY \n\n1X40HQ",  "keyStr": " \n\nPORT OF DISCHARGE:",  "valueStr": "",  "kvMap": {},  "y": null,  "x": null,  "type": "KV",  "fieldId": 5,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": true,  "validKey": true,  "pointer": false  },  "measurement": {  "top": 714,  "bottom": 740,  "left": -500,  "right": -1000,  "width": 1,  "height": 1,  "text": " MEASUREMENT\n\nN/M 272 PKGS 15740KGS 62CBM\n\nNET WEIGTH :15740KGS \n\n\n\nDESCRIPTION :\n\nTYRES \n\nCARGO INTRANSIT TO KANDHAR AFGHANISTAN VIA KARACHI\n\nPLEASE TAKE 21DAY FREE DETENTION ON CONTAINER\n\nPLZ ENTRY THIS CONTAINER IGM VIA QCHM\n\nGOODS CAN NOT BE UNLOADED ON QASIM PORT\n\n\n\n",  "keyStr": " MEASUREMENT\n\n",  "valueStr": "N/M272PKGS15740KGS62NETWEIGTH:15740KGSDESCRIPTION:TYRESCARGOINTRANSITTOKANDHARAFGHANISTANVIAKARACHIPLEASETAKE21DAYFREEDETENTIONONCONTAINERPLZENTRYTHISCONTAINERIGMVIAQCHMGOODSCANNOTBEUNLOADEDONQASIMPORT",  "kvMap": {},  "y": null,  "x": null,  "type": "KV",  "fieldId": 10,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": true,  "validKey": true,  "pointer": false  },  "serviceTerm": {  "top": null,  "bottom": null,  "left": null,  "right": null,  "width": null,  "height": null,  "text": null,  "keyStr": "",  "valueStr": "",  "kvMap": {},  "y": null,  "x": null,  "type": null,  "fieldId": null,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": false,  "validKey": true,  "pointer": false  },  "marksNos": {  "top": null,  "bottom": null,  "left": null,  "right": null,  "width": null,  "height": null,  "text": null,  "keyStr": "",  "valueStr": "N/M",  "kvMap": {},  "y": null,  "x": null,  "type": null,  "fieldId": null,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": false,  "validKey": true,  "pointer": false  },  "grossWeight": {  "top": null,  "bottom": null,  "left": null,  "right": null,  "width": null,  "height": null,  "text": null,  "keyStr": "",  "valueStr": "15740",  "kvMap": {},  "y": null,  "x": null,  "type": null,  "fieldId": null,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": false,  "validKey": true,  "pointer": false  },  "consAddr": {  "top": 195,  "bottom": 221,  "left": 0,  "right": -1000,  "width": 0,  "height": 1,  "text": "\n\n\n\nCONSIGNEE:\n\nAFRICA SPEED, CO., LTD\n\nNORHT AFRICA, XXX ROAD, NA\n\nJAWAZ NO:1800. TIN NO:1069999888\n\nMOBILE NO:0096-777888600 & 0096-777889600",  "keyStr": "\n\n\n\nCONSIGNEE:\n\n",  "valueStr": "NORHT AFRICA, XXX ROAD, NA\nJAWAZ NO:1800. TIN NO:1069999888\nMOBILE NO:0096-777888600 & 0096-777889600",  "kvMap": {},  "y": null,  "x": null,  "type": "KV",  "fieldId": 2,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": true,  "validKey": true,  "pointer": false  },  "hsCode": {  "top": null,  "bottom": null,  "left": null,  "right": null,  "width": null,  "height": null,  "text": null,  "keyStr": "",  "valueStr": "",  "kvMap": {},  "y": null,  "x": null,  "type": null,  "fieldId": null,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": false,  "validKey": true,  "pointer": false  },  "notifyPartyAddr": {  "top": 351,  "bottom": 377,  "left": 0,  "right": -1000,  "width": 0,  "height": 1,  "text": "\n\n\n\n\n\n\n\nNOTIFY PARTY:\n\nPAN JAV IND, AND CO (PVC)LTD \n\nOFF#606 ABC CITY MALL EXAMPLE\n\nROBINS ROAD GUANGZHOU, CN\n\nPH.009388997763330 CELL. 00923003500689-9\n\nEMAIL:tomas.edson@yahoo.com",  "keyStr": "\n\n\n\n\n\n\n\nNOTIFY PARTY:\n\n",  "valueStr": "OFF#606 ABC CITY MALL EXAMPLE\nROBINS ROAD GUANGZHOU, CN\nPH.009388997763330 CELL. 00923003500689-9\nEMAIL:tomas.edson@yahoo.com",  "kvMap": {},  "y": null,  "x": null,  "type": "KV",  "fieldId": 3,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": true,  "validKey": true,  "pointer": false  },  "notifyParty": {  "top": 351,  "bottom": 377,  "left": 0,  "right": -1000,  "width": 0,  "height": 1,  "text": "\n\n\n\n\n\n\n\nNOTIFY PARTY:\n\nPAN JAV IND, AND CO (PVC)LTD \n\nOFF#606 ABC CITY MALL EXAMPLE\n\nROBINS ROAD GUANGZHOU, CN\n\nPH.009388997763330 CELL. 00923003500689-9\n\nEMAIL:tomas.edson@yahoo.com",  "keyStr": "\n\n\n\n\n\n\n\nNOTIFY PARTY:\n\n",  "valueStr": "PAN JAV IND, AND CO (PVC)LTD ",  "kvMap": {},  "y": null,  "x": null,  "type": "KV",  "fieldId": 3,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": true,  "validKey": true,  "pointer": false  },  "containerInfo": {  "top": null,  "bottom": null,  "left": null,  "right": null,  "width": null,  "height": null,  "text": null,  "keyStr": "",  "valueStr": "",  "kvMap": {},  "y": null,  "x": null,  "type": null,  "fieldId": null,  "businessCode": null,  "businessType": null,  "dataList": [  {  "containerCount": "1",  "containerCode": "40HQ"  }  ],  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": false,  "validKey": true,  "pointer": false  },  "portOfLoading": {  "top": 584,  "bottom": 610,  "left": 0,  "right": -1000,  "width": 0,  "height": 1,  "text": " \n\n \n\nPORT OF LOADING: SHANGHAI",  "keyStr": " \n\n \n\nPORT OF LOADING:",  "valueStr": "SHANGHAI",  "kvMap": {},  "y": null,  "x": null,  "type": "KV",  "fieldId": 4,  "businessCode": "CNSHA",  "businessType": "PORT\_MAP",  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": true,  "validKey": true,  "pointer": false  },  "packageInfo": {  "top": null,  "bottom": null,  "left": null,  "right": null,  "width": null,  "height": null,  "text": null,  "keyStr": "",  "valueStr": "",  "kvMap": {},  "y": null,  "x": null,  "type": null,  "fieldId": null,  "businessCode": null,  "businessType": null,  "dataList": [  {  "packageCount": "1",  "packageType": "PKGS",  "packageSum": "272"  }  ],  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": false,  "validKey": true,  "pointer": false  },  "freightCharges": {  "top": null,  "bottom": null,  "left": null,  "right": null,  "width": null,  "height": null,  "text": null,  "keyStr": "",  "valueStr": "",  "kvMap": {},  "y": null,  "x": null,  "type": null,  "fieldId": null,  "businessCode": null,  "businessType": null,  "dataList": null,  "dataMap": {},  "level": null,  "fieldSplitList": null,  "cellFormula": null,  "finish": false,  "validKey": true,  "pointer": false  }  },  "data": null  }  } |
| 3 | http://ip:port/base/api//containerInfo | File | {  "returnCode": 200,  "returnMsg": "操作成功",  "returnData": {  "id": null,  "reviewUrl": "http://192.168.102.61:8080/v1/AUTH\_zvolume/bucket1/ocr\_rec/null/1544106453939.html",  "orderNo": null,  "fieldData": {},  "data": {  "containerList": [  {  "containerNo": "MOFU1407006"  },  {  "containerNo": "MOFU6827110"  },  {  "containerNo": "NYKU8255666"  },  {  "containerNo": "NYKU8369191"  }  ]  }  }  } |