



<搜象 OCR 增值税发票结构化识别> 接口开发参考

搜象（厦门）信息技术有限公司

2022 年 02 月 21 日

目 录

1	第一章. 引擎接口开发	3
1.1	接口概述	3
1.2	服务请求报文	3
1.2.1	请求报文示例.....	3
1.2.2	请求数据项说明.....	3
1.3	服务返回报文	3
1.3.1	返回报文示例.....	3
1.3.2	返回数据项说明.....	7
1.4	失败返回码定义	8
1.4.1	失败返回示例.....	8
1.4.2	失败返回码定义.....	8
1.5	附-识别字段.....	9
1.5.1	增值税发票.....	9

1 第一章. 引擎接口开发

1.1 接口概述

请求地址：请参考商品页 API 接口调用地址

请求方式：POST

请求格式：Content-Type: application/json

返回类型：JSON

1.2 服务请求报文

1.2.1 请求报文示例

```
{
  "fileType": 1,
  "filePath": "http://192.168.0.2:8080/test.jpg"
}
```

1.2.2 请求数据项说明

数据项	数据类型	描述
fileType	int	识别文件类型，1-图片，2-PDF，不传则自动判断
filePath	string	识别文件路径（filePath 和 fileBase64 二者择一传参即可）
fileBase64	string	识别文件 Base64 编码（需带 Base64 头，如：图片【data:image/jpeg;base64,】, PDF【data:application/pdf;base64,】）
pageIndex	int	识别 PDF 第几页，不传则默认首页，首页从 0 开始，该参数对 PDF 有效

1.3 服务返回报文

1.3.1 返回报文示例

```
{
  "code": 0,
```

```

"data": {
  "angle": 0.9839911,
  "anglePic": 0,
  "columnResultList": [
    {
      "columnBox": [1030,346,1190,349,1189,381,1029,377],
      "columnKey": "invoice_type",
      "columnName": "发票类型",
      "columnValue": "专用发票"
    },
    {
      "columnBox": [927,430,1089,434,1088,475,927,471],
      "columnKey": "couplet_type",
      "columnName": "抵扣/发票联",
      "columnValue": "发票联"
    },
    {
      "columnBox": [1537,383,1655,385,1654,418,1537,416],
      "columnKey": "invoice_code",
      "columnName": "发票代码",
      "columnValue": "3400161130"
    },
    {
      "columnBox": [1529,416,1652,419,1651,447,1529,444],
      "columnKey": "invoice_number",
      "columnName": "发票号码",
      "columnValue": "00666375"
    },
    {
      "columnBox": [1454,453,1624,457,1623,481,1454,477],
      "columnKey": "invoice_date",
      "columnName": "开票日期",
      "columnValue": "2016-11-11"
    },
    {
      "columnBox": [582,497,889,504,888,527,582,521],
      "columnKey": "buyer_name",
      "columnName": "购买方名称",
      "columnValue": "北京游久新媒信息技术有限公司"
    },
    {
      "columnBox": [604,530,947,537,947,564,603,557],
      "columnKey": "buyer_identifier",

```

```

        "columnName": "购买方纳税人识别号",
        "columnValue": "91310114585239729M"
    },
    {
        "columnBox": [584,564,1109,575,1108,599,583,589],
        "columnKey": "buyer_addr_tell",
        "columnName": "购买方地址电话",
        "columnValue": "北京市海淀区工业汇源路 55 号 H 幢 3 层 A 区 021-63460206"
    },
    {
        "columnBox": [584,598,1069,608,1068,635,583,625],
        "columnKey": "buyer_bank_account",
        "columnName": "购买方开户行及账号",
        "columnValue": "中国工商银行北京市海淀区支行 1001700819300415148"
    },
    {
        "columnBox": [1153,507,1599,516,1597,628,1151,618],
        "columnKey": "password_area",
        "columnName": "密码区",
        "columnValue":
"1-8-208>4>2/*3123<12<8<5224\n+<7<41-845>62<0*3>-94<- /53+\n<5/+9>--1576+020*31
45*322>\n43<***5>41</6572192*-72>/-<"
    },
    {
        "columnBox": [374,664,446,665,445,694,374,692],
        "columnKey": "project_col",
        "columnName": "货物或服务名称列",
        "columnValue": "信息费"
    },
    {
        "columnBox": [1318,679,1400,681,1400,704,1318,702],
        "columnKey": "amount_col",
        "columnName": "金额列",
        "columnValue": "2524.75"
    },
    {
        "columnBox": [1424,681,1455,682,1455,705,1423,704],
        "columnKey": "tax_rate_col",
        "columnName": "税率列",
        "columnValue": "6%"
    },
    {
        "columnBox": [1583,684,1651,686,1651,708,1583,707],

```

```

        "columnKey": "tax_amount_col",
        "columnName": "税额列",
        "columnValue": "151.49"
    },
    {
        "columnBox": [1264,895,1395,897,1394,925,1263,922],
        "columnKey": "total_amount",
        "columnName": "合计金额(元)",
        "columnValue": "2524.75"
    },
    {
        "columnBox": [1533,898,1649,900,1648,927,1533,925],
        "columnKey": "total_tax_amount",
        "columnName": "合计税额(元)",
        "columnValue": "151.49"
    },
    {
        "columnBox": [742,930,1044,936,1044,966,742,960],
        "columnKey": "total_amount_tax_cap",
        "columnName": "价税合计(大写)",
        "columnValue": "贰仟陆佰集拾陆圆贰角肆分"
    },
    {
        "columnBox": [1394,936,1539,939,1538,974,1393,971],
        "columnKey": "total_amount_tax_low",
        "columnName": "价税合计(小写/元)",
        "columnValue": "2676.24"
    },
    {
        "columnBox": [581,978,886,984,885,1010,580,1004],
        "columnKey": "seller_name",
        "columnName": "销售方名称",
        "columnValue": "合肥乐堂动漫信息技术有限公司"
    },
    {
        "columnBox": [608,1009,947,1016,946,1045,607,1038],
        "columnKey": "seller_identifier",
        "columnName": "销售方纳税人识别号",
        "columnValue": "91340100686877076E"
    },
    {
        "columnBox": [578,1037,1055,1047,1054,1081,577,1071],
        "columnKey": "seller_addr_tell",

```

```

        "columnName": "销售方地址电话",
        "columnValue": "合肥市金泰路 71 号科茂大厦 5 层 0551-65411799"
    },
    {
        "columnBox": [579,1072,985,1080,984,1107,578,1099],
        "columnKey": "seller_bank_account",
        "columnName": "销售方开户行及账号",
        "columnValue": "招商银行合肥南七支行 551903169110102"
    },
    {
        "columnBox": [454,1099,521,1100,520,1129,453,1128],
        "columnKey": "cashier",
        "columnName": "收款人",
        "columnValue": "李平"
    },
    {
        "columnBox": [792,1102,866,1103,865,1138,791,1137],
        "columnKey": "reviewer",
        "columnName": "复核",
        "columnValue": "李平"
    },
    {
        "columnBox": [1124,1109,1196,1110,1196,1141,1123,1139],
        "columnKey": "drawer",
        "columnName": "开票人",
        "columnValue": "秦丽萍"
    }
],
"templateCatId": "825454941819633664",
"templateCateName": "增值税发票",
"templateId": "855283192611995648",
"templateName": "增值税发票"
},
"message": "success"
}

```

1.3.2 返回数据项说明

数据项	数据类型	描述
code	int	返回编码，0-代表成功，其余失败
data	object	返回具体数据项
+angle	float	文字顺时针偏转方向

+anglePic	int	图片四分类方向, 0-正向, 90-右向, 180-反向, 270-左向
+templateCateId	string	模板分类 ID
+templateCateName	string	模板分类名称
+templateId	string	模板 ID
+templateName	string	模板名称
+columnResultList	objectArray	文字识别数据项列表
++columnName	string	字段中文名称
++columnKey	string	字段英文名称
++columnValue	string	字段值 (\n 字符表示字段值中的换行或较大间隔)
++columnBox	intArray	字段坐标->[x1,y1,x2,y2,x3,y3,x4,y4] (非固定项, 坐标方式提取的字段有返回, 规则方式提取的字段无返回)
message	string	返回信息, 包括成功失败信息

1.4 失败返回码定义

1.4.1 失败返回示例

```
{
  "code": 1011,
  "data": {},
  "message": "无识别文件"
}
```

1.4.2 失败返回码定义

返回编码	返回信息	描述
1011	无识别文件	filePath 和 fileBase64 同时为空
1012	不支持的文件类型	支持图片和 PDF, 图片包括但不限于 jpg、jpeg、png、bmp、gif、tif、tiff、jfif、heic、jpe、dib 等格式
1014	文件 URL 错误	文件 URL 有误, 读取不到资源
1015	文件 BASE64 串错误	文件 BASE64 串有误, 解析文件异常
1099	服务器端异常	返回具体异常信息

1.5 附-识别字段

1.5.1 增值税发票

字段中文名称	字段英文名称	字段中文名称	字段英文名称
发票类型	invoice_type	单价列	unit_price_col
抵扣/发票联	couplet_type	金额列	amount_col
校验码	check_code	税率列	tax_rate_col
发票代码	invoice_code	税额列	tax_amount_col
发票号码	invoice_number	合计金额(元)	total_amount
开票日期	invoice_date	合计税额(元)	total_tax_amount
购买方名称	buyer_name	价税合计(大写)	total_amount_tax_cap
购买方纳税人识别号	buyer_identifier	价税合计(小写/元)	total_amount_tax_low
购买方地址电话	buyer_addr_tell	销售方名称	seller_name
购买方开户行及账号	buyer_bank_account	销售方纳税人识别号	seller_identifier
密码区	password_area	销售方地址电话	seller_addr_tell
货物或服务名称列	project_col	销售方开户行及账号	seller_bank_account
规格型号列	model_col	收款人	cashier
单位列	unit_col	复核	reviewer
数量列	number_col	开票人	drawer