



<搜象 OCR 税务登记证结构化识别> 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.....	5
1.4	失败返回码定义	6
1.4.1	失败返回示例.....	6
1.4.2	失败返回码定义.....	6
1.5	附-识别字段.....	7
1.5.1	税务登记证.....	7

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": 359.83392,
  "anglePic": 0,
  "columnResultList": [
    {
      "columnBox": [832,225,975,225,974,251,831,251],
      "columnKey": "tax_number",
      "columnName": "税字",
      "columnValue": "440300561523899"
    },
    {
      "columnBox": [837,262,916,262,915,282,836,283],
      "columnKey": "national_tax_code",
      "columnName": "国税纳税编码",
      "columnValue": "83023258"
    },
    {
      "columnBox": [836,293,916,292,915,314,836,314],
      "columnKey": "local_tax_code",
      "columnName": "地税纳税编码",
      "columnValue": "07370847"
    },
    {
      "columnBox": [316,267,580,266,580,294,316,295],
      "columnKey": "name",
      "columnName": "纳税人名称",
      "columnValue": "深圳市兆丰年酒店用品有限公司"
    },
    {
      "columnBox": [314,317,383,316,383,347,314,347],
      "columnKey": "legal_person",
      "columnName": "法定代表人",
      "columnValue": "王连志"
    },
    {
      "columnBox": [314,369,655,368,655,396,314,397],
      "columnKey": "address",
      "columnName": "地址",
      "columnValue": "深圳市龙岗区布吉街道美杜兰华庭 C 座 13E"
    },
    {
      "columnBox": [316,422,473,421,472,445,316,446],
      "columnKey": "regist_type",

```

```

        "columnName": "登记注册类型",
        "columnValue": "私营有限责任公司"
    },
    {
        "columnBox": [316,458,1020,454,1017,531,317,535],
        "columnKey": "operation_scope",
        "columnName": "经营范围",
        "columnValue": "酒店用品、装饰材料、装修材料的购销国内贸易（法律、行政法规、国务院决定规定在登记前须经批准的项目除外）货物及技术进出口（法律、行政法规禁止的项目除外法律、行政法规限制的项目须取得许可后方可经营）"
    },
    {
        "columnBox": [315,554,512,553,512,576,315,578],
        "columnKey": "approval_org",
        "columnName": "批准设立机关",
        "columnValue": "深圳市市场监督管理局"
    },
    {
        "columnBox": [315,605,479,604,479,632,315,633],
        "columnKey": "withhold_obligation",
        "columnName": "扣缴义务",
        "columnValue": "依法确定"
    }
],
"templateCatelId": "825455953208934400",
"templateCateName": "税务登记证",
"templateId": "852666410357227520",
"templateName": "税务登记证(带纳税编码)"
},
"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 税务登记证

字段中文名称	字段英文名称	字段中文名称	字段英文名称
税字	tax_number	地址	address
国税纳税编码	national_tax_code	登记注册类型	regist_type
地税纳税编码	local_tax_code	经营范围	operation_scope
纳税人名称	name	批准设立机关	approval_org
法定代表人	legal_person	扣缴义务	withhold_obligation