

发票分类、识别和处理

2019年10月16日

Vincent Zheng

Presales Technical Consultant

vincent.zheng@uipath.com



前言

发票

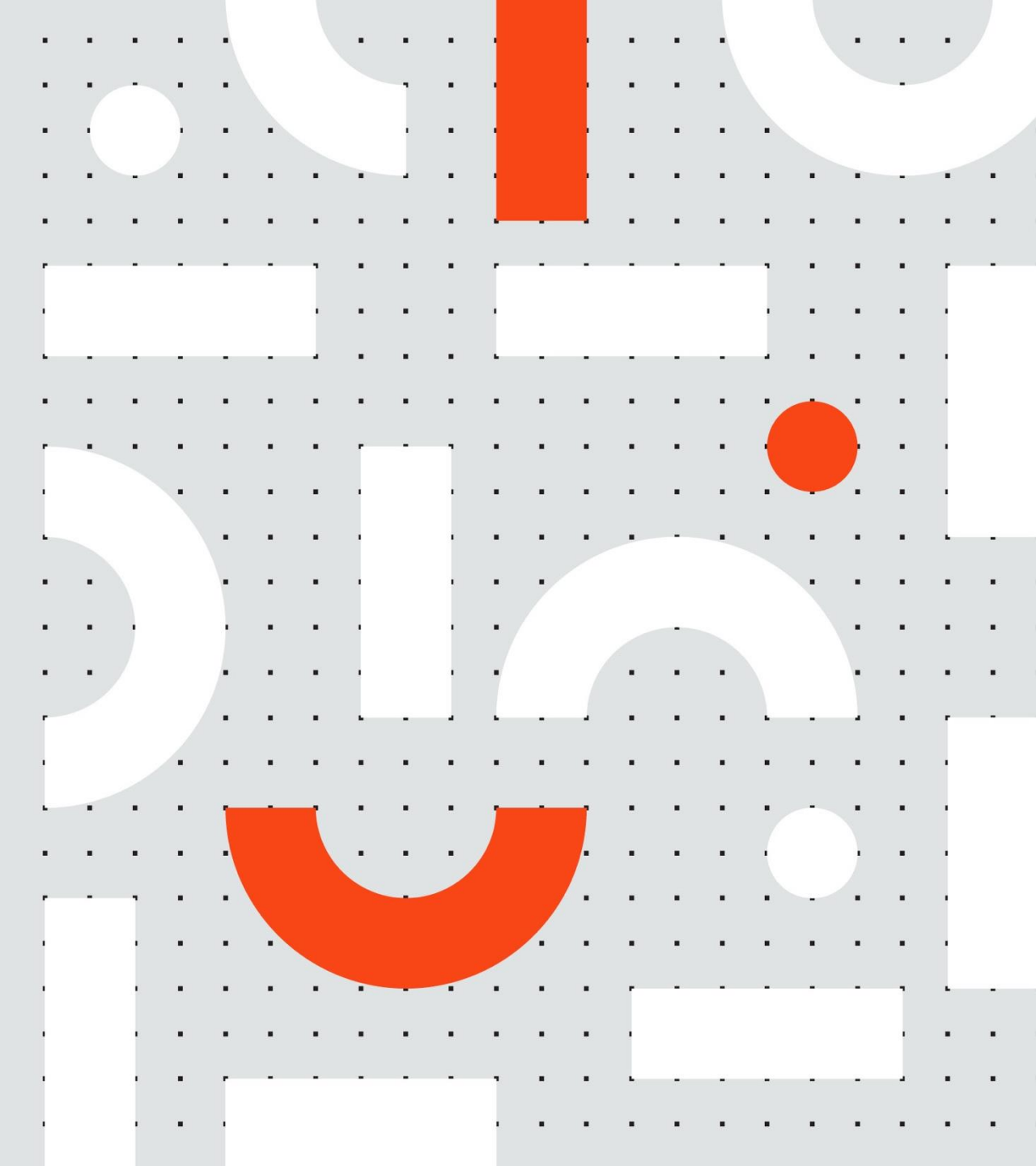
- 分类
- 识别
- 处理

演示

问答



1. 前言



发票数字化白皮书：2018年市场规模达179.8亿元，同比增长18.1%

未来发票数字化将呈现四大趋势

- 发票数字化数据和技术不断向融合方向发展，服务更多层次的用户
- RPA将全面提升财务处理效率
- 未来发票数字化服务将向发票数据服务方向延展
- 标准化统一平台将成为发票数字化平台产品的变革趋势。

2016-2020 中国发票数字化市场规模与增长



Use machine learning to Unlock the knowledge and insights hiding in your documents 通过机器学习解锁隐藏在文档中的知识和见解



UiPath RPA+AI 结合, 解决发票处理难题

行业痛点

- 手工处理发票过程繁琐单调重复
- 图像质量差, 识别率低, 需要人工介入
- 跨业务系统集成困难



解决方案



方案效果



提升效率
覆盖所有的供应商。

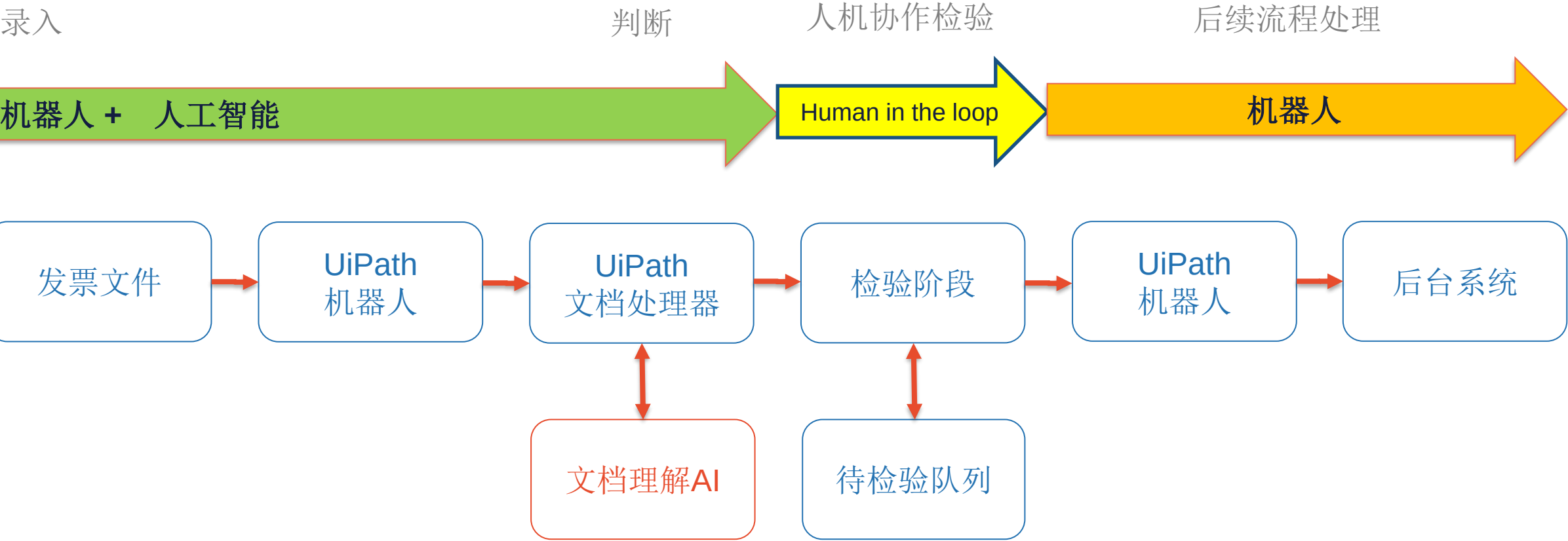


降低处理成本
结合验证站, 降低出错率。

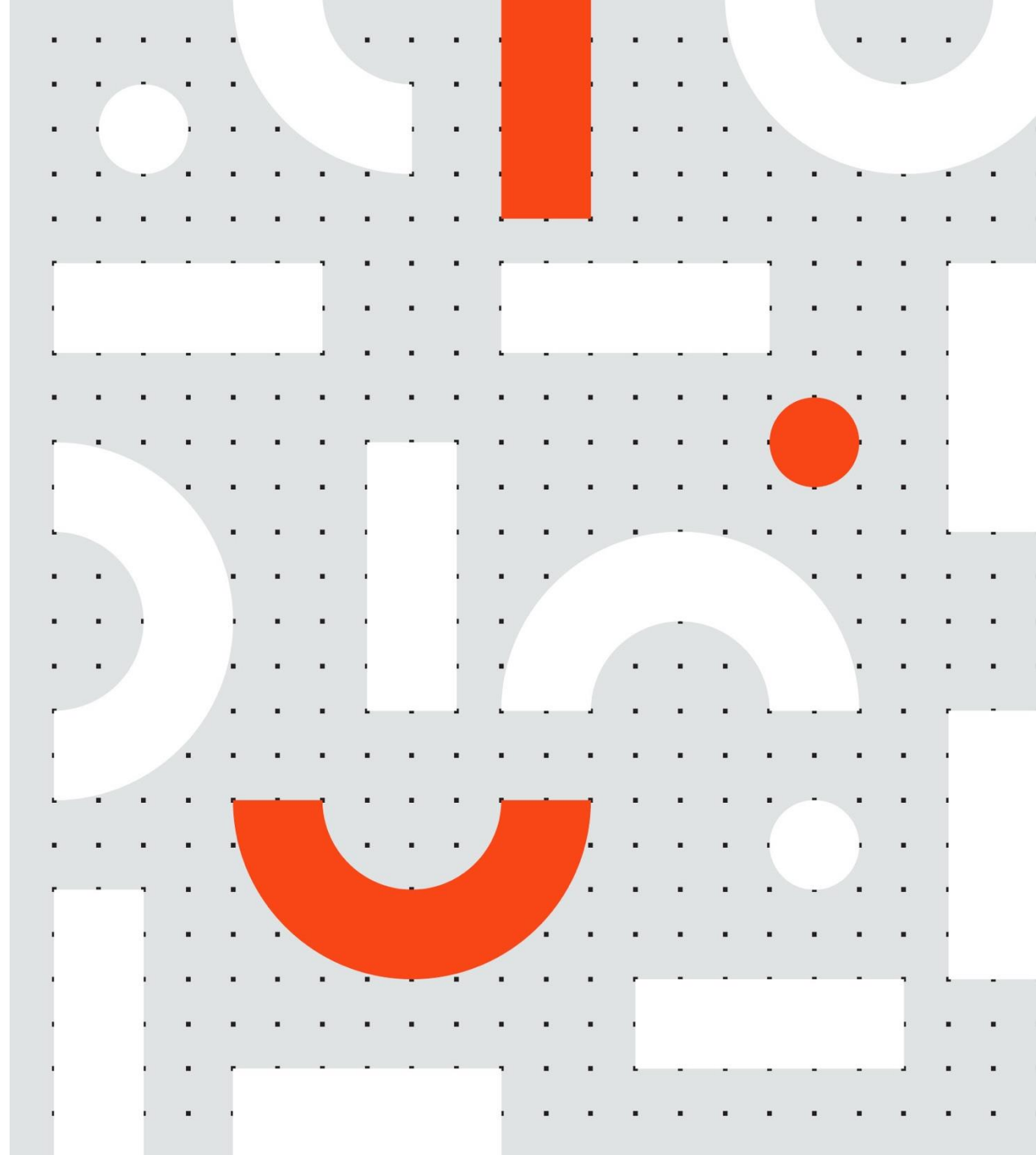


减少操作风险
识别不能接受的业务实践。

UiPath文档处理的最佳实践



2. 发票处理实践

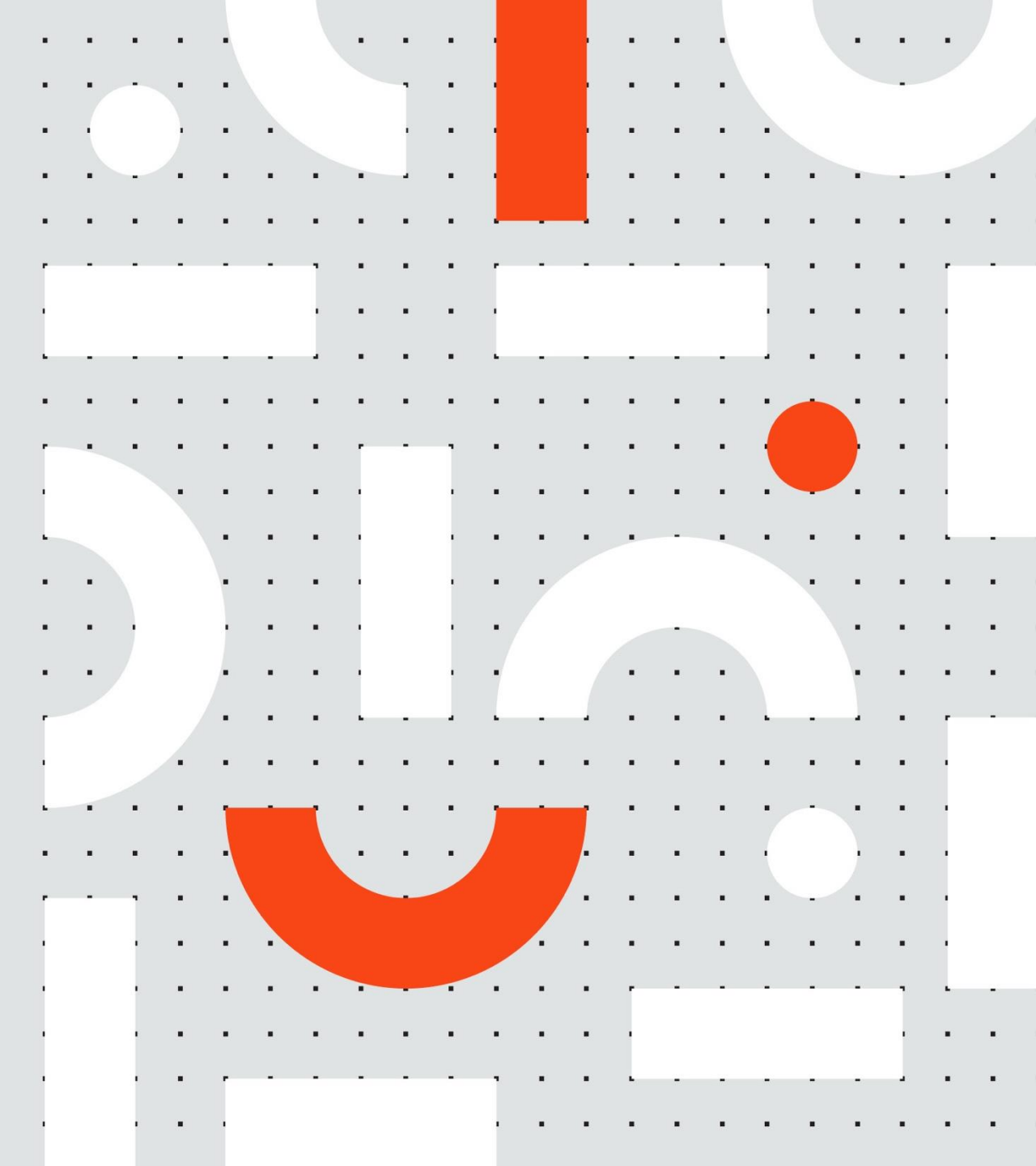


发票样例

字段	值
InvoiceNum	15031
SellerName	上海
CommodityTaxRate	1]
SellerBank	招商
Checker	唐毓
NoteDrawer	张颖
CommodityAmount	1]
InvoiceDate	2019年
CommodityTax	1]
PurchaserName	天津
InvoiceTypeOrg	上海
CommodityNum	1]
PurchaserBank	天津
Remarks	
Password	
SellerAddress	虹口
PurchaserAddress	天津
InvoiceCode	31001
CommodityUnit	[]
Payee	來惠
PurchaserRegisterNum	91120
CommodityPrice	[]
TotalAmount	2264.
AmountInWords	贰仟
AmountInFiguers	0.00
TotalTax	135.8
InvoiceType	专用
SellerRegisterNum	91310
CommodityName	1]
CommodityType	[]
CheckCode	
CommodityTaxRate	

80494810001		31001 脱敏处理		上海增值税专用发票		No. 15031 脱敏处理	
		开票日期: 2019年04月22日					
名称: 天津		纳税人识别号: 9		密码: 5-163+<0-267094/305+327<268		6*9>143411/3*8>/5/976583///	
地址、电话: 天津		开户行及账号: 天津		密 码: 90-806+/3<2/5*6+2/4-4052764		>3+46<->142711/3-+>>1/11>79	
货物或应税劳务、服务名称: *住宿服务*房费		规格型号		单位		数量	
2143						单价	
8888						金 额	
405614						税率	
						税 额	
						2264.15	
						135.85	
合 计						¥ 2264.15	
含税合计(大写)						¥ 135.85	
销 名: 上海		纳税人识别号: 9		备 注: 2019年05月13日-2019年05月15日 CWT		脱敏处理	
地 址、电 话: 上海		开户行及账号: 招商银行上海分行长阳支行214080494810001		注 册 号: 913101096072453703		脱敏处理	
收款人: 朱惠强		复核: 周晓青		开票人: 陈颖		脱敏处理	

1) 分类方法



增值税发票样例

增值税普通发票

011001800304 北京增值税普通发票 No 2909427 脱敏处理

校验码 44201 18958 开票日期: 2019年10月11日

脱敏处理

货物或应税劳务、服务名称	规格型号	单位	数量	单价	金额	税率	税额
*餐饮服务*餐费		份	1	265.09433962	265.09	6%	15.91
合计					¥265.09		¥15.91
价税合计(大写)					¥281.00		

脱敏处理

销方 名称: 北京哈 纳税人识别号: 91 地址、电话: 北京市 开户行及账号: 建行北 收款人: 刘丹 复核: 金岩 开票人: 肖怀敬

增值税专用发票

3700161130 山东增值税专用发票 No 0215 脱敏处理

代开 开票日期: 2016年05月04日

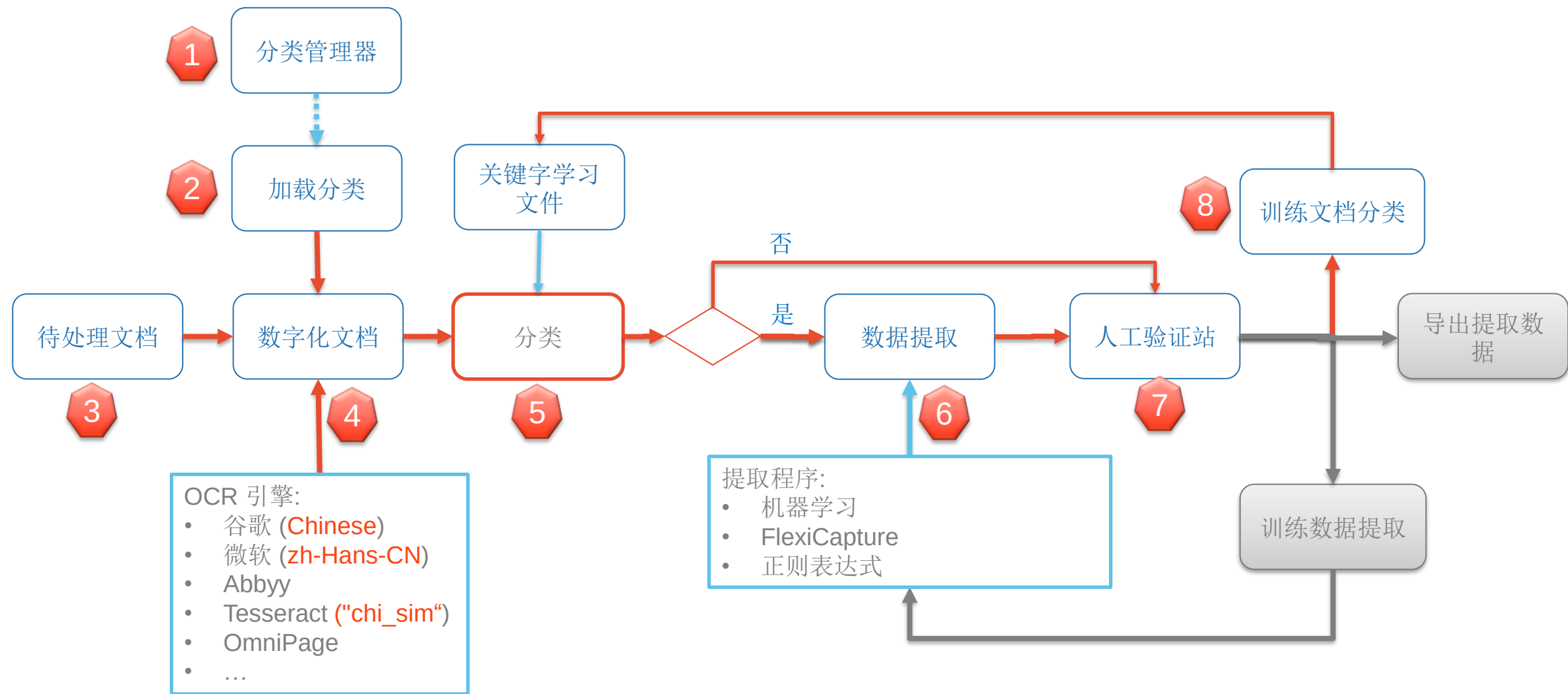
脱敏处理

货物或应税劳务、服务名称	规格型号	单位	数量	单价	金额	税率	税额
设计费		平方米	1	48543.6993	48543.69	3%	1456.31
合计					¥48543.69		¥1456.31
价税合计(大写)					¥50000.00		

脱敏处理

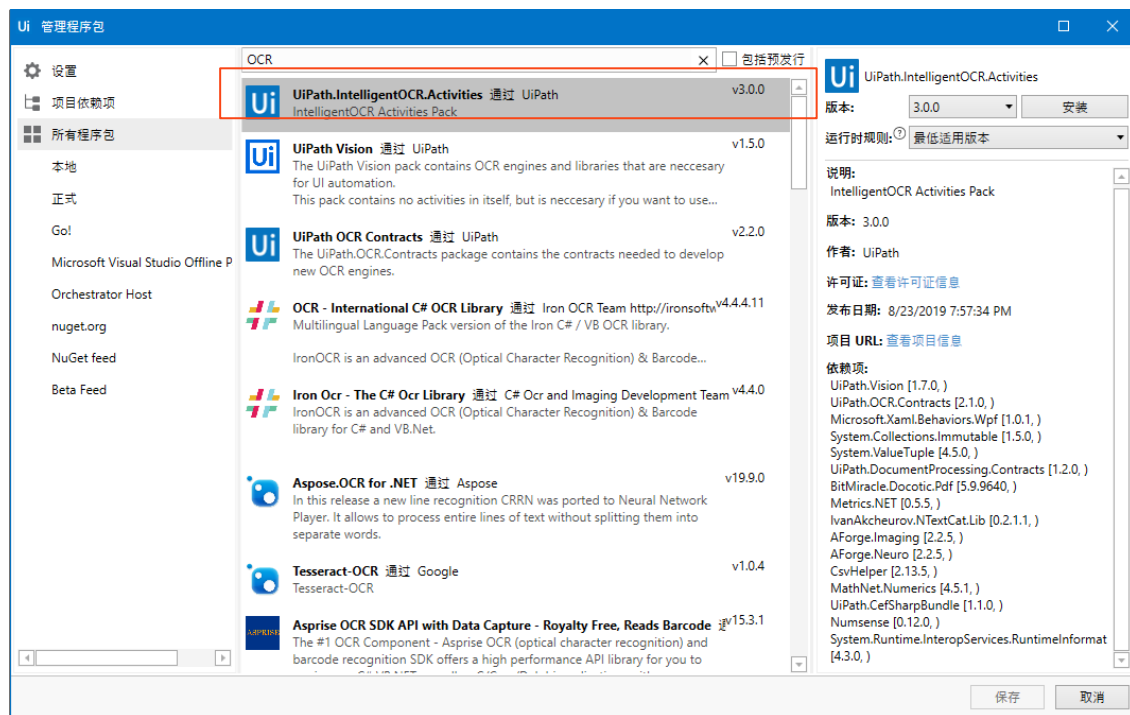
销方 名称: 淄博高新区 纳税人识别号: 3703 地址、电话: 淄博开发区 开户行及账号: 320160504000299936 (完税凭证号) 收款人: 刘海东 复核: 刘海东 开票人: 刘海东 销售方: (章)

文档分类的处理方法

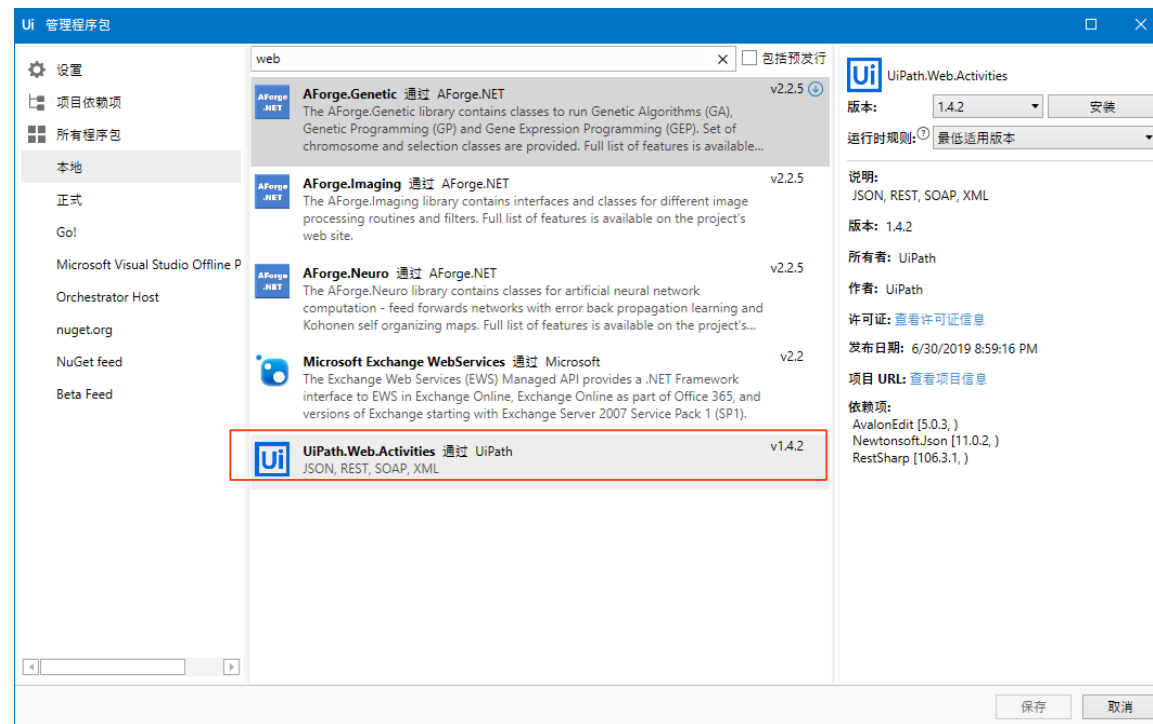


安装活动包

UiPath.IntelligentOCR.Activities



UiPath.Web.Activities



分类管理器

UI 分类管理器

增值税

任何类别

+

按名称搜索

添加新文档类型

单击文档类型进行编辑。

• 增值税

◦ 专用

▪ 增值税专用发票

◦ 普通

▪ 增值税普通发票

文档类型详情

文档类型 ID: 增值税.普通.增值税普通发票

名称

增值税普通发票

组

增值税

类别

普通

文档类型代码

VAT_N_01

新字段

- 点击管理程序包, 安装 IntelligentOCR Activities Pack;
- 运行分类管理器, 定义文档的组和类别;
- 添加文档类型;
- 在DocumentProcessing文件夹中, 会产生分类文件: taxonomy.json。

分类定义 – Taxonomy.json

```
{
  "DataContractVersion": "1.0",
  "DocumentTypes": [
    {
      "DocumentTypeId": "增值税.专用.增值税专用发票",
      "Group": "增值税",
      "Category": "专用",
      "Name": "增值税专用发票",
      "OptionalUniqueIdentifier": "VAT_S_01",
      "TypeField": {
        "FieldId": "增值税.专用.增值税专用发票.DocumentType",
        "FieldName": "Document Type"
      },
      "Fields": []
    },
    {
      "DocumentTypeId": "增值税.普通.增值税普通发票",
      "Group": "增值税",
      "Category": "普通",
      "Name": "增值税普通发票",
      "OptionalUniqueIdentifier": "VAT_N_01",
      "TypeField": {
        "FieldId": "增值税.普通.增值税普通发票.DocumentType",
        "FieldName": "Document Type"
      },
      "Fields": []
    }
  ]
}
```


分类管理-人工验证-培训分类关键字

- 对每一张发票进行分类处理;
- 打开验证站点, 查看是否已经匹配?
- 如果没有分类, 或分类不正确, 则手动分类并保存;
- 验证站点关闭后, 培训 KeywordLearning.json 文件。



培训结果 – KeywordLearning.json

```
[
  {
    "DocumentTypeId": "增值税.专用.增值税专用发票",
    "Keywords": [
      {
        "Values": ["增 住 枕 专 用 发 票"],
        "NumberOfTimesConfirmed": 2,
        "LastUpdated": "2019-10-14T07:59:32.7257043+08:00"
      }
    ]
  },
  {
    "DocumentTypeId": "增值税.普通.增值税普通发票",
    "Keywords": [
      {
        "Values": ["首 通 发 票"],
        "NumberOfTimesConfirmed": 2,
        "LastUpdated": "2019-10-14T07:58:34.0097333+08:00"
      },
      {
        "Values": ["电 普 通 发 票"],
        "NumberOfTimesConfirmed": 2,
        "LastUpdated": "2019-10-14T07:58:50.4212667+08:00"
      },
      {
        "Values": ["增值稅 手 通 发 票"],
        "NumberOfTimesConfirmed": 2,
        "LastUpdated": "2019-10-14T07:59:05.9182711+08:00"
      },
      {
        "Values": ["增 信 花 普 通 发 票"],
        "NumberOfTimesConfirmed": 2,
        "LastUpdated": "2019-10-14T07:59:19.9413187+08:00"
      },
      {
        "Values": ["增值 花 普 通 发 票"],
        "NumberOfTimesConfirmed": 2,
        "LastUpdated": "2019-10-14T07:59:43.4137333+08:00"
      }
    ]
  }
]
```

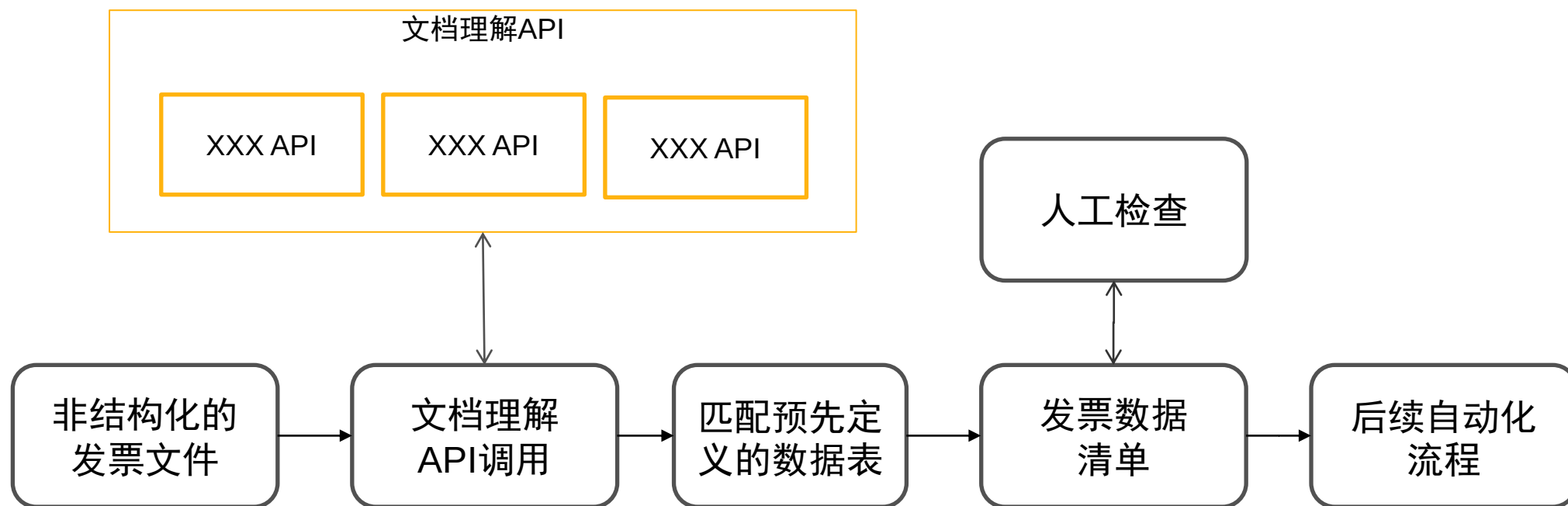
2) 增值税发票识别

几个智能票据识别 API介绍



http://www.hanwang.com.cn/	https://www.ccint.com	https://cloud.baidu.com/
文档: https://pan.baidu.com/s/1KGNjEOwk1xoXqBjOWN1QXQ	文档: https://ai.ccint.com/doc/api/vat_invoice https://ai.ccint.com/user/sign-in 用户名: Uipathtest 口令: Uipath123	文档: http://ai.baidu.com/docs#/OCR-API-VatInvoice/top
API调用地址: (增值税发票为例) http://223.223.177.134:3004/hwocr/v1.0/invoice	API调用地址: (增值税发票为例) https://ocr-api.ccint.com/ocr_service?app_key=	API调用地址: (增值税发票为例) https://aip.baidubce.com/rest/2.0/ocr/v1/vat_invoice
Key: 652023C877DCD922A4487296FEDCA6B0	App_Key: 3e5df5cf63b993ca36a0ff4b534b7ea4 App_secret: 51f6f57890334013b81bc5acaacd33f6	Access_Token: 自行登陆申请获得

UiPath RPA 调用文档理解API 实现发票处理的思路

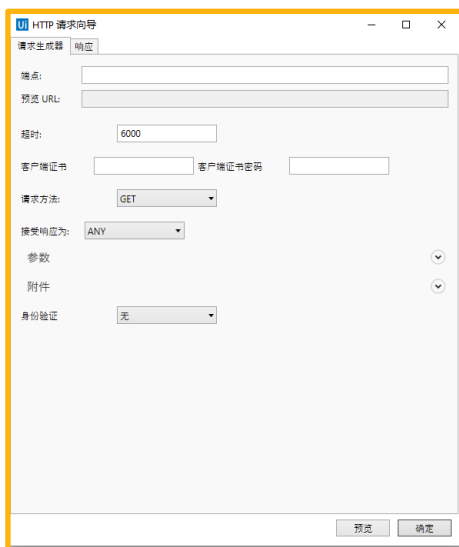


几个常用的活动或方法

通过HTTP请求调用API接口

UiPath.Web.Activities.HttpClient

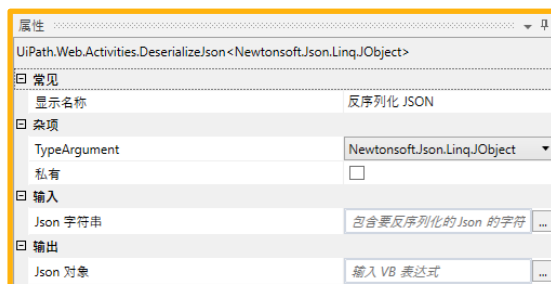
- 执行对指定Web API的HTTP请求
- 发送或获取信息



Json 反序列化

UiPath.Web.Activities.DeserializeJson

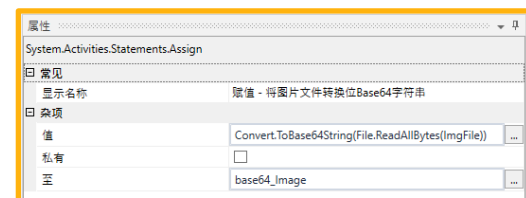
- 输入: JsonString - 包含要反序列化的JSON的字符串
- 输出: **JsonObject**-反序列化为 Newtonsoft.Json.Linq.JObject 变量。



图像文件转换为base64格式

UiPath.Web.Activities.Assign

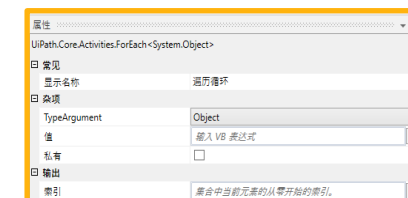
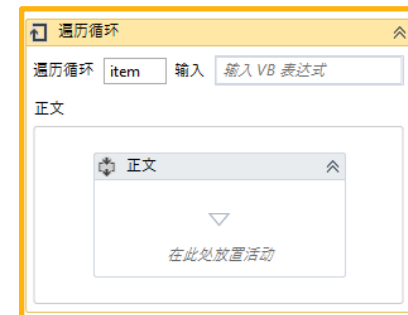
- 值: ImgFile是需要转换格式的图片文件的存放路径。
- 至: 名为 base64_Image 的变量



循环遍历 (代码片段)

UiPath.Core.Activities.ForEach

- 值: ImgFile是需要转换格式的图片文件的存放路径。
- 至: 名为 base64_Image 的变量



增值税发票API调用方法 - 汉王科技

实现步骤

1. 先将图片转换为Base64字符串:
`Convert.ToBase64String(File.ReadAllBytes(图片文件))`
2. 调用HTTP请求, 输入必要的参数:
 - 设置端点: `"http://223.223.177.134:3004/hwocr/v1.0/invoice"`
 - 方法: `Post`
 - 接受格式: `Json格式`
 - 选项参数: `key(从汉王获得的Key)`, `image_base64(第1步生成的变量)`, `probability("true")`
 - 输出: 设置接收"标题","状态码","结果"的变量
3. 如果调用成功, 将返回结果反序列化,写入文本文件中。[可选]
4. 读取该文本文件, 进行数据整理, 存入Excel表中。

增值税发票



- URL: <http://223.223.177.134:3004/hwocr/v1.0/invoice>
- 请求类型: POST
- 请求参数

是否必选	参数	类型	说明
是	key	String	申请的key字符串
image_file image_base64 image_url 三选一	image_file	file	一个图片, 二进制文件, 需要用 post multipart/form-data 的方式上传。
	image_base64	String	base64编码的二进制图片数据 如果同时传入了image_url、image_file和image_base64参数, 本API使用顺序为image_file优先, image_url最低。
	image_url	String	图片的URL 注: 在下载图片时可能由于网络等原因导致下载图片时间过长, 建议使用image_file 或image_base64 参数直接上传图片。 在此版本中, 暂时还没有支持此类型, 不可用!

返回值说明

字段	类型	说明
code	int	状态值
message	String	状态值对应的说明
items	JSONArray	增值税发票识别结果(支持一张纸上多张发票) 注: 若没有检测出发票, 则为空数组

返回结果



数据获取的方法

- 定义JCode为返回结果的Json对象
- 遍历 JCode.Item("items")(0) 可以获得发票的表头信息信息;

```
{
  "code": 200,
  "message": "ok",
  "items": [
    {
      "agency": false,
      "checkCode": "",
      "confidence": "0.66",
      "details": [
        {
          "amount": "135.85",
          "debitName": "**住宿服务*房费",
          "money": "2264.15",
          "quantity": "01",
          "specification": "",
          "taxRate": "6%",
          "unit": "甲位",
          "unitPrice": "6115.85"
        }
      ],
      "moneyAmount": "¥ 2264.15",
      "moneyTaxAmount": "¥ 135.85",
      "moneyTaxAmountChn": "人民币",
      "receiptCode": "310018",
      "receiptDate": "2019404",
      "receiptHead": "上海增信",
      "receiptNo": "15031873",
      "seal": true,
      "taxAmount": "¥ 135.85",
      "vendeeAddrTel": "天津 15号022-28405614",
      "vendeeBkAcct": "天津 10620120109022143区",
      "vendeeId": "91120000",
      "vendeeName": "天津银",
      "vendorAddrTel": "虹口 号021-6545888",
      "vendorBkAcct": "招商 支行",
      "vendorId": "913101096",
      "vendorName": "上海远"
    }
  ]
}
```

脱敏处理

增值税发票API调用方法 – 合合信息

实现步骤

1. 先将图片转换成为Base64字符串:

```
imageData = Convert.ToBase64String(File.ReadAllBytes(图片文件))
```

2. 按API调用要求, 构建Json文本

```
appSecret = 从合合申请的"app_secret"的安全码串
```

```
jsonTemplate="{\"image_data\": \"{0}\", \"app_secret\": \"{1}\"}"
```

```
jsonInput = jsonTemplate.Replace("{0}",imageBase64).Replace("{1}",appSecret)
```

3. 调用HTTP请求, 输入必要的参数:

- 设置端点: “https://ocr-api.ccint.com/ocr_service?app_key=从合合申请的”app_key”的令牌串
- 方法: Post
- 接受格式: Json
- 设置正文: jsonInput
- 输出: 设置接收”标题”, ”状态码”, ”结果”的变量

4. 如果调用成功, 将返回结果反序列化,写入文本文件中。[可选]

5. 读取该文本文件, 进行数据整理, 存入Excel表中。



https://ai.ccint.com/doc/api/vat_invoice

增值税发票识别API调用文档

调用地址: `https://ocr-api.ccint.com/ocr_service?app_key=`

说明: url中参数app_key为个人中心实例的app_key

请求方式: POST

返回类型: JSON



POST BODY字段描述

字段	说明
image_data	必填, 图像的base64编码字符串
app_secret	必填, 个人中心实例的app_secret

POST BODY

```
{  "image_data": "", // 必填, 图像的base64编码字符串
  "app_secret": "" // 必填, 个人中心实例的app_secret
}
```

提示: POST BODY 为 **JSON**字符串。

返回结果



数据获取的方法

- 定义JCode为返回结果的Json对象
- 遍历JCode.Item("result")可以获得发票的辅助信息;
- 遍历 JCode.Item("result").Item("ocr_data_list")可以获得表头信息 - Master

```
{
  "cost_time": 3415,
  "result": {
    "image_angle": 0,
    "type_description": "增值税普通发票",
    "rotated_image_width": 1803,
    "rotated_image_height": 1080,
    "kind": "service",
    "ocr_data_list": [
      {
        "value": "非电子发票",
        "position": "",
        "key": "vat_invoice_type",
        "description": "发票类型"
      },
      {
        "value": "普票",
        "position": "",
        "key": "vat_invoice_zh",
        "description": "专票/普"
      },
      {
        "value": "01100180030",
        "position": "",
        "key": "vat_invoice_da",
        "description": "发票代码"
      },
      {
        "value": "29094279",
        "position": "",
        "key": "vat_invoice_ha",
        "description": "发票号码"
      },
      {
        "value": "44201189581",
        "position": "",
        "key": "vat_invoice_co",
        "description": "校验码"
      },
      {
        "value": "2019年10月1",
        "position": "",
        "key": "vat_invoice_iss",
        "description": "开票日期"
      },
      {
        "value": "44201189581169106738",
        "position": "431,181,710,181,710,206,431,206",
        "key": "vat_invoice_corrent_code_print",
        "description": "打印校验码"
      },
      {
        "value": "发票联",
        "position": "804,146,10",
        "key": "vat_invoice_he",
        "description": "标题发"
      },
      {
        "value": "29094279",
        "position": "1270,72,14",
        "key": "vat_invoice_ha",
        "description": "增值税"
      },
      {
        "value": "01100180030",
        "position": "1493,80,16",
        "key": "vat_invoice_da",
        "description": "右侧打"
      },
      {
        "value": "29094279",
        "position": "1484,111,1635,111,1635,1",
        "key": "vat_invoice_ha",
        "description": "右侧打"
      },
      {
        "value": "2019年10月1",
        "position": "1332,155,1582,155,1582,1",
        "key": "vat_invoice_iss",
        "description": "右侧打"
      },
      {
        "value": "上海才赋人力",
        "position": "428,224,8",
        "key": "vat_invoice_pa",
        "description": "购买方"
      },
      {
        "value": "91310107MA",
        "position": "415,261,88",
        "key": "vat_invoice_rat",
        "description": "购买方"
      }
    ]
  }
}
```

增值税发票API调用方法 – 百度智能云

实现步骤

1. 先将图片转换为Base64字符串:
`base64_Image = Convert.ToBase64String(File.ReadAllBytes(图片文件))`
2. 调用HTTP请求, 输入必要的参数:
 - 设置端点: "https://aip.baidubce.com/rest/2.0/ocr/v1/vat_invoice"
 - 方法: **Post**
 - 接受格式: **Json**
 - 设置参数: access_token = “注册用户并付费后获得的令牌”
image = **base64_Image**
 - 输出: 设置接收”标题”, ”状态码”, ”结果”的变量
3. 如果调用成功, 将返回结果反序列化,写入文本文件中。[可选]
4. 读取该文本文件, 进行数据整理, 存入Excel表中。

<https://cloud.baidu.com/doc/OCR/OCR-API.html#.E8.AF.B7.E6.B1.82.E8.AF.B4.E6.98.8E>

请求示例

HTTP 方法: **POST**

请求URL: `https://aip.baidubce.com/rest/2.0/ocr/v1/vat_invoice`

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考“ Access Token获取 ”

Header如下:

参数	值
Content-Type	application/x-www-form-urlencoded

Body中放置请求参数, 参数详情如下:

参数	是否必选	类型	可选值范围	说明
image	是	string	-	图像数据, base64编码后进行urlencode, 要求base64编码和urlencode后大小不超过4M, 最短边至少15px, 最长边最大4096px,支持jpg/jpeg/png/bmp格式
accuracy	否	string	normal、high	normal (默认配置) 对应普通精度模型, 识别速度较快, 在四要素的准确率上和high模型保持一致, high对应高精度识别模型, 相应的时延会增加, 因为超时导致失败的情况也会增加 (错误码282000)

返回结果



返回结果分析

- 定义JCode为返回结果的Json对象
- 遍历JCode.Item("words_result")可以获得发票的表头信息

```
{
  "log_id": 118894384487804206,
  "direction": 0,
  "words_result_num": 21,
  "words_result": {
    "InvoiceNum": "1",
    "SellerName": "上海",
    "CommodityTax": {
      "word": "66%",
      "row": "1"
    }
  },
  "SellerBank": "招",
  "Checker": "唐毓",
  "NoteDrawer": "5",
  "CommodityAmount": {
    "word": "2264.",
    "row": "1"
  },
  "InvoiceDate": "2",
  "CommodityTax": {
    "word": "4135.",
    "row": "1"
  },
  "PurchaserName": "天津银行股份有限公司",
  "InvoiceTypeOrg": "上海增值税专用发票",
  "CommodityNum": [
    {
      "word": "",
      "row": "1"
    }
  ],
  "PurchaserBank": "部",
  "Remarks": "",
  "Password": "",
  "SellerAddress": "路1171号021-",
  "PurchaserAddress": "西區友谊路15",
  "InvoiceCode": "3",
  "CommodityUnit": "套",
  "Payee": "來惠琴",
  "PurchaserRegister": "9112000010307029",
  "CommodityPrice": "套",
  "TotalAmount": "2",
  "AmountInWords": "套",
  "AmountInFigures": "套",
  "TotalTax": "135.8",
  "InvoiceType": "专",
  "SellerRegisterNu": "6072453703",
  "CommodityName": {
    "word": "**住宿服务*房费",
    "row": "1"
  },
  "CommodityType": [],
  "CheckCode": ""
}
```


3) 发票数据整理

数据整理的思路 (API > Json > Datatable > Excel)

读取调用发票识别API返回的Json 文本

执行反序列化, 将文本转化为Json对象

对Json对象节点进行遍历读取

读取出的字段信息, 与配置表中的字段进行匹配

将不同格式的API返回信息, 填入统一格式的Excel表格中

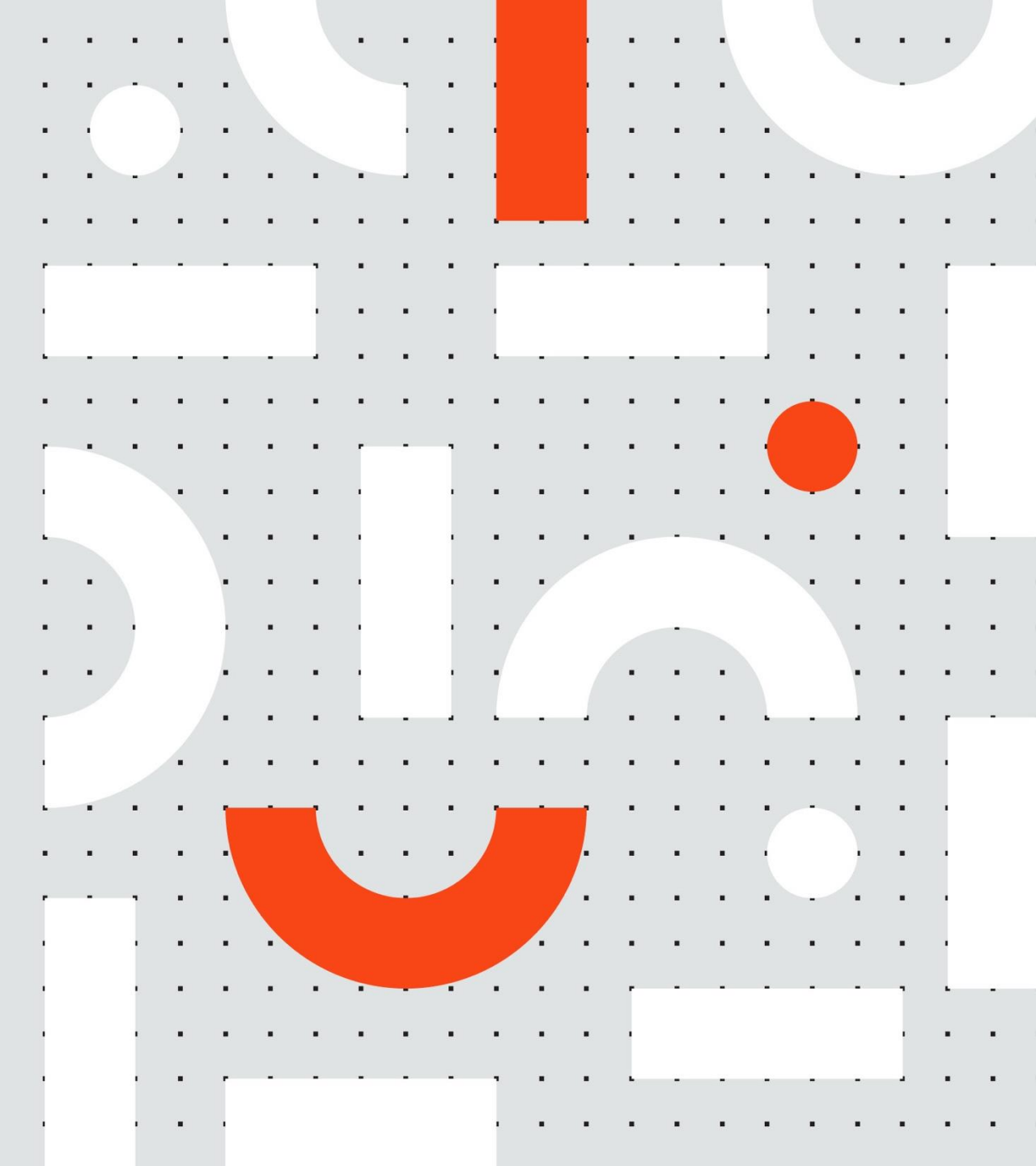
各API返回结果字段匹配表

统一格式		 百度智能云		 汉王 Hanvon		 合合信息		
字段名	字段说明	字段名-BD	字段说明-BD	字段名-HW	字段说明-HW	字段名-CC	字段说明-CC	备注
confidence	可信度			confidence	可信度		可信度	
InvoiceType	发票种类	InvoiceType	发票种类			type_description		
InvoiceTypeOrg	发票名称	InvoiceTypeOrg	发票名称	receiptHead	发票抬头	vat_invoice_type	TBD	
InvoiceCode	发票代码	InvoiceCode	发票代码	receiptCode	发票代码	vat_invoice_daima	发票代码	
InvoiceNum	发票号码	InvoiceNum	发票号码	receiptNo	发票号码	vat_invoice_haoma	发票编号	
CheckCode	校验码	CheckCode	校验码	checkCode	校验码	vat_invoice_correct_code	发票校验码	
InvoiceDate	开票日期	InvoiceDate	开票日期	receiptDate	开票日期	vat_invoice_issue_date	开具发票的日期	
PurchaserName	购方名称	PurchaserName	购方名称	vendeeName	购买方名称	vat_invoice_payer_name	购买方的名称	
PurchaserRegisterNum	购方纳税人识别号	PurchaserRegisterNum	购方纳税人识别号	vendeeld	购买方纳税人识别号	vat_invoice_rate_payer_id	购买方的纳税人识别号	
PurchaserAddress	购方地址及电话	PurchaserAddress	购方地址及电话	vendeeAddrTel	购买方地址电话	vat_invoice_payer_addr_tell		
PurchaserBank	购方开户行及账号	PurchaserBank	购方开户行及账号	vendeeBkAcct	购买方开户行及账号	vat_invoice_payer_bank_account		
SellerName	销售方名称	SellerName	销售方名称	vendorName	销售方名称	vat_invoice_seller_name	销售方的名称	
SellerRegisterNum	销售方纳税人识别号	SellerRegisterNum	销售方纳税人识别号	vendorId	销售方纳税人识别号	vat_invoice_seller_id	销售方的纳税人识别号	
SellerAddress	销售方地址及电话	SellerAddress	销售方地址及电话	vendorAddrTel	销售方地址电话	vat_invoice_seller_addr_tell		
SellerBank	销售方开户行及账号	SellerBank	销售方开户行及账号	vendorBkAcct	销售方开户行账号	vat_invoice_seller_bank_account		
TotalAmount	合计金额	TotalAmount	合计金额	moneyAmount	金额合计	vat_invoice_total	合计金额	
TotalTax	合计税额	TotalTax	合计税额	taxAmount	税额合计	vat_invoice_tax_total	纳税金额	
AmountInWords	价税合计(大写)	AmountInWords	价税合计(大写)	moneyTaxAmountChn	价税合计大写	vat_invoice_total_cover_tax	含税总金额(大写)	
AmountInFiguers	价税合计(小写)	AmountInFiguers	价税合计(小写)	moneyTaxAmount	价税合计小写	vat_invoice_total_cover_tax_digits	含税总金额(小写)	
Payee	收款人	Payee	收款人			vat_invoice_payee		
Checker	复核	Checker	复核			vat_invoice_review		
NoteDrawer	开票人	NoteDrawer	开票人			vat_invoice_drawer		
Remarks	备注	Remarks	备注			vat_invoice_total_note		
seal	是否有印章			seal	是否有印章	exist_stample		
agency	是否代开			agency	是否代开	vat_invoice_dai_kai_flag		
detials	货物详情	CommodityName		detials	货物详情	vat_invoice_goods_list	发票中的商品列表	

统一输出 格式

API	编号	发票文件	日期时间	可信度	发票种类	发票名称	发票代码	发票号码	校验码	开票日期	购方名称	购方纳税人识别号	购方地址及电话	购方开户行及账号	销售方名称	销售方纳税人识别号	销售方地址及电话	销售方开户行及账号	合计金额	合计税额	价税合计(大写)	价税合计(小写)	收款人	复核	开票人	备注	是否有印章	是否代开	货物详情

3. 演示



功能实现

- 将图片文件提前做一个分类处理, 获得待处理文件的类型;
- 将图片转换成为文档理解AI的API接口要求的格式;
- 通过调用文档感知AI的API接口, 提取图片的关键信息, 通常以JSON格式返回;
- 将不同的API返回数据统一整理,成为统一格式;
- 判断返回的结果是否成功, 检查可信度值(Confidence Level);
 - 可信度不高, 则交由人工查验和处理, 比如在单独的Excel表格, 或者应用UiPath Validation Station验证站点;
 - 可信度满足要求, 则将Json格式数据, 整理到Excel表格中。
- 将整理好数据用于后续自动化流程中。

小结

1. UiPath RPA发票分类、识别和处理的思路
2. 以机器学习 API调用为例发票处理的Demo演示
3. 涉及到一些方法
 - 如何对发票分类;
 - 如何将图片转换成Base64字符串;
 - 通过HTTP Request 活动, 调用API接口的几种方法;
 - 如何将Json数据转换成数据表, 并将其存入Excel表中。

UiPath 机器人发展方向, 自带认知和AI能力!



通过UiPath产品，打造全球最大的RPA开放生态系统

采用战略前瞻技术
内置最前沿的技术以加速创新

全面深入Rest API
真正动态集成到您当前的基础架构

数百个组件
嵌入、可定制、可分享的模块化组件

分享与延伸
可重用的资料库，直接与UiPath Go!集成

100万
学生，2020年前

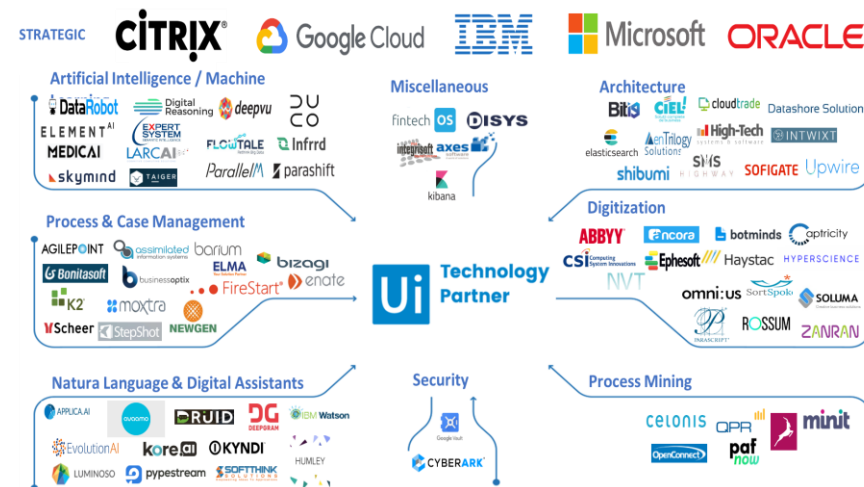
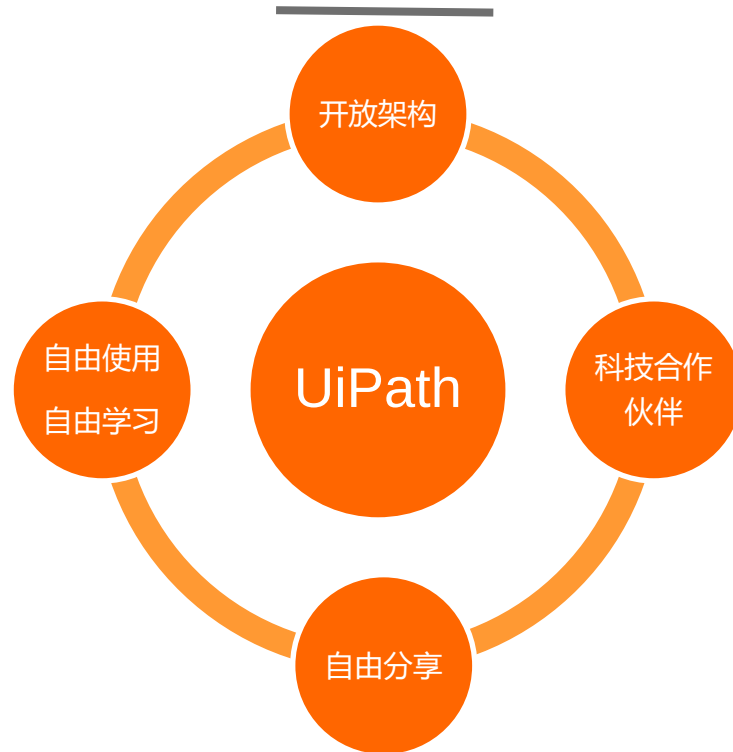
30万
社区版本下载

30万
名在线大学学员

12万
认证开发者

学术联盟

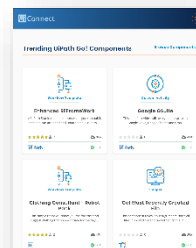
与千家大学、非盈利机构
以及政府联盟，UiPath致
力于为一个自动化优先的
世界提供教育资源



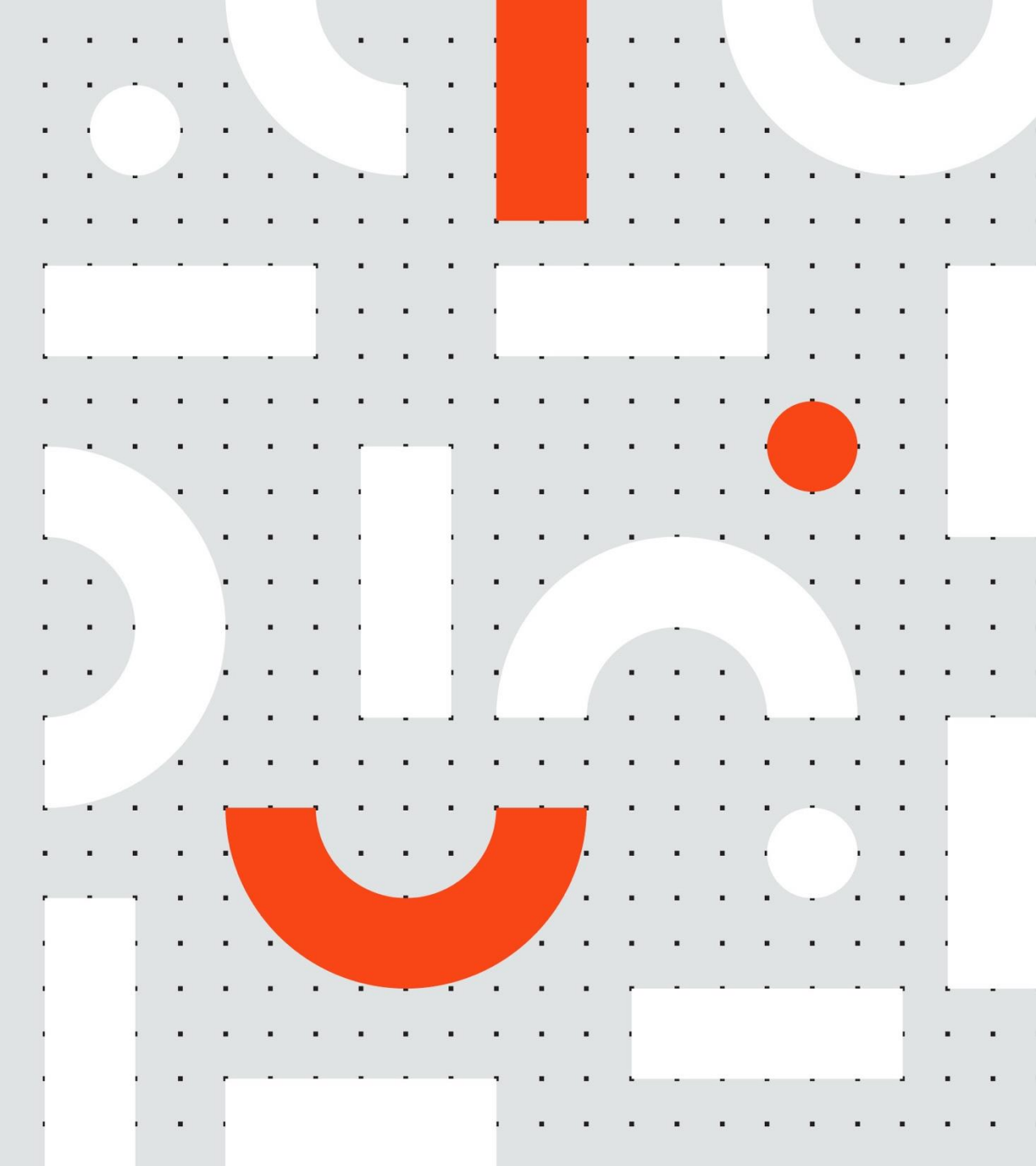
对最前沿的科技生态开放



UiPath Connect
开发者交流合作
UiPath Go
流程市场



3. Q&A 现场问答



感谢您的聆听, 期待您的反馈!

UiPath 官方微信公众号



在线问卷调查表

