

作业 1 基于流程自动化的数据处理 RPA 构建

目的:通过基于 RPA 技术实现数据处理流程的自动化,了解多来源 API 的 Mashup 集成方式,熟悉大量简单任务的流程化自动化实现技术。

工具: TagUI

提交时间: 11 月 11 日晚 12: 00 前提交 Canvas 平台 (oc.sjtu.edu.cn)

业务描述及要求

在企业账务管理中,经常需要对发票进行分析和归档,繁琐的发票处理流程常令财务人员感到头疼。为了减轻财务统计负担,现在需要设计并实现一套自动化发票处理程序,它能够从发票中提取账目信息、分析数据并归档存储、简单计算后形成报表并发送。其大致流程如下:

[1]. 数据采集:

从文件夹中批量读取数据,批量大小自行设定。通过 OCR(Optical Character Recognition)技术提取发票中的文字,获取发票中关于交易的信息(包括但不限于交易主体、发票类型、明细、金额等)。

[2]. 数据治理:

(1) 发票审批: 对于数据集 a, 付款方为“浙江大学”、时间在 2015 年内、审批金额在 1600 元以内的发票视为合规; 对于数据集 b, 付款方为“深圳市购机汇网络科技有限公司”、时间在 2016 年 6 月 12 日、审批金额在 2700 元以内的发票视为合规; 对于算法无法识别的发票, 打包通过邮件转发至人工审批。

记录每张发票的审批状态(“通过”、“不通过”、“转人工”三种)。

(2) 财务关系治理: 基于发票信息, 提取出交易主体(买方和卖方)间的交易关系, 并按照金额、交易频度等信息将交易主体按照买方和卖方分类表述(例如大客户、客户、一般客户; 或者重要供应商、供应商、一般供应商)。

[3]. 数据存储:

发票及治理数据归档持久化: 将处理的原始数据和数据结果(包括发票图片)存入数据库中(具体数据库自选)。

[4]. 数据计算:

自选简单算法计算处理发票的统计信息, 包括但不限于发票数量、审批状态比例、当前交易量最大的 Top K 个交易方等。

[5]. 展示应用:

展示数据处理结果。

并汇总发票信息和统计信息并制成 Excel 报表, 通过邮箱发送到主管部门。邮件中应简要说明此批数据处理的情况。

发票数据地址: <https://jbox.sjtu.edu.cn/1/I1Nfgt>

作业要求:

[1]. 2-3 人一组完成作业, 实现上述场景功能并录制自动化运行效果。

[2]. 需要按照 FBS 建模方法给出该 RPA 工具的 FBS 图和表格, 并结合该 RPA 工具的结构(流程的 5 环节), 阐述采用实现技术, 以及带来的结构(性能指标)的特点。

[3]. 需要在流程中体现 RPA 工具的作用, 例如用 TagUI 实现图像识别 API、邮件系统调用等, 体现出多种来源 API 的结合, 而不是用单一方式实现(例如 Python)。

- [4]. 助教会从两个数据集中分别保留一个小的测试集用作现场验收，答辩时能够现场运行的将给予额外分数。
- [5]. 撰写简要文档以说明流程自动化实现的主要技术、数据结构和创新点等，并在文档中标明各组员的姓名、学号、分工情况、贡献比例。将源代码、演示视频、说明文档打包提交到 Canvas 平台。提交文件命名：**MIIS2024HW1-组长姓名-组长学号**。

开发工具参考：

RPA 开源工具：TagUI

- 采用 Apache 2.0, AI Singapore 开发维护。
- Github: <https://github.com/aisingapore/TagUI>
- 中文使用文档: <http://www.tagui.com.cn/index.html>
- 兼容 Linux/Win/MacOS 多操作系统，支持 Web、桌面、云的脚本自动化，支持多种 Office 插件。
- 部分兼容 JavaScript 语法，上手简单，文档和 Q&A 齐全。
- 可以使用 Python 等其它脚本语言进行开发。
- 可以使用第三方库，如 OCR、NLP 的实现库等。

RPA 设计与构建过程参考

主要包括四个部分：

- [1]. 明确流程及细节逻辑：RPA 设计的第一步，首先要明确业务流程（人工），将业务过程拆分为若干个环节，每个环节具有明确的基础功能、能够模块化，如数据检索存储、数据分析加工、图像识别处理、文件写入、邮件通信等。
- [2]. 考虑流程各环节的业务对接系统：逐一思考业务流程的各环节如何运用自动化技术进行替代，有哪些能够应用的库（如 OCR）、模块（如图像识别）或软件系统（如 ERP）。系统无需完全对应人工业务流程，可根据系统功能进行环节的合并或拆分。
- [3]. 设计 RPA 自动化模式的流程和接口：确定业务流程每个环节的 RPA 替换逻辑后，需对业务环节连点成线，设计基于 RPA 自动化模式的业务流程，打通环节之间的数据流和信息流。
- [4]. RPA 自动化流程开发与构建

为大家注册了一个公用的测试邮箱：

Email: sjtuse3301@163.com

Password: SJTUse1896

若出现邮箱服务器限制流量的问题，也可自行注册测试邮箱。