



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Invoice Processing System along with Chat Bot Using RPA

Ms. Manimozhi R¹, Mr. P. Anbumani²

¹(MCA), ²MCA, M. Phil, NET., Associate Professor, Department of MCA
Krishnasamy College of Engineering and Technology, Cuddalore.

ABSTRACT

The process of Robotic Process Automation (RPA) can be applied to a number of finance and accounting operations, invoice processing. RPA Chat Bot can automate data input, error reconciliation, and some of the decision-making required by finance staff when processing invoices. UiPath's RPA Chat Bot are able to constantly monitor a dedicated storage where invoices are saved in PDF format. Once robots detect the presence of an invoice in the storage, admin begin to extract information from the document. Using intelligent Optical Character Recognition (OCR) and natural language processing (NLP) capabilities, Chat Bot are able to read out the information from invoice. After robots extract the key information, the robots start processing the invoices one-by-one transferring over the relevant invoice information. During the process, the Chat Bot are also running background activities such as monitoring the dedicated invoice storage or its email address, performing basic checks to see if the company's dataset is open, and verifying whether invoice information (e.g. Invoice number) on the invoice matches in the dataset.

Key Terms: Reconciliation, Presence, Optical Character Recognition.

I. Introduction

In today's technologically advanced world, organizations rely on various systems and processes to streamline their operations. The goal of this project is to automate the process of invoice processing extracting data from invoice PDF files and store admin in an Excel sheet using Robotic Process Automation (RPA) techniques [1]. RPA involves the use of software robots to automate repetitive tasks, mimicking human actions to complete them more efficiently and accurately [7]. By automating the extraction and storage of invoice data using RPA to increase efficiency, reduce manual effort, and minimize errors associated with manual data entry [3]. User enables organizations to streamline their invoice processing workflows, improve data accuracy, and free up human resources for more value-added tasks [11]. The manual process of extracting data from invoice PDFs and entering into Excel sheets can be time-consuming by automating process [4]. RPA can significantly reduce the time required to handle each invoice, resulting in faster turnaround times and increased productivity.

II. Proposed Methodology

In this paper further enhance its capabilities and provide additional benefits to organizations. The Invoice processing along with chat bot will be able to handle complex inquiries, provide personalized recommendations. The invoice processing captured, the system should automatically extract relevant data from the invoice such as invoice number, invoice date, billed to and totals. This can be achieved through Utilize OCR (Optical Character Recognition) technology to extract text from the invoice PDF files. OCR tools like image-based PDFs into machine-readable text and data extraction algorithms.

Advantages

- 1. Increased Efficiency:** This method used to OCR technology and data extraction algorithm, the system can automate the extraction of relevant data from invoice PDF files. This eliminates the need for manual data entry, saving time and reducing human error. The system can process invoices more quickly and efficiently, resulting in improved productivity and faster invoice processing cycles.
- 2. Enhanced Accuracy:** This method used to reduces the chances of human errors associated with manual data entry and improves the overall accuracy of the invoice processing system. Organizations can rely on the system to provide accurate and reliable invoice data for further processing and analysis.
- 3. Cost Savings:** This method used for invoice processing and reducing manual effort, the system helps organizations save costs associated with labor-intensive invoice handling processes. The reduction in errors also minimizes the need for rework and dispute resolution, leading to cost savings and improved financial management.
- 4. Scalability and Adaptability:** This method can handle a large volume of invoices efficiently, admin can easily scale to accommodate increasing invoice volumes, making this suitable for organizations experiencing growth or seasonal variations in invoice processing.

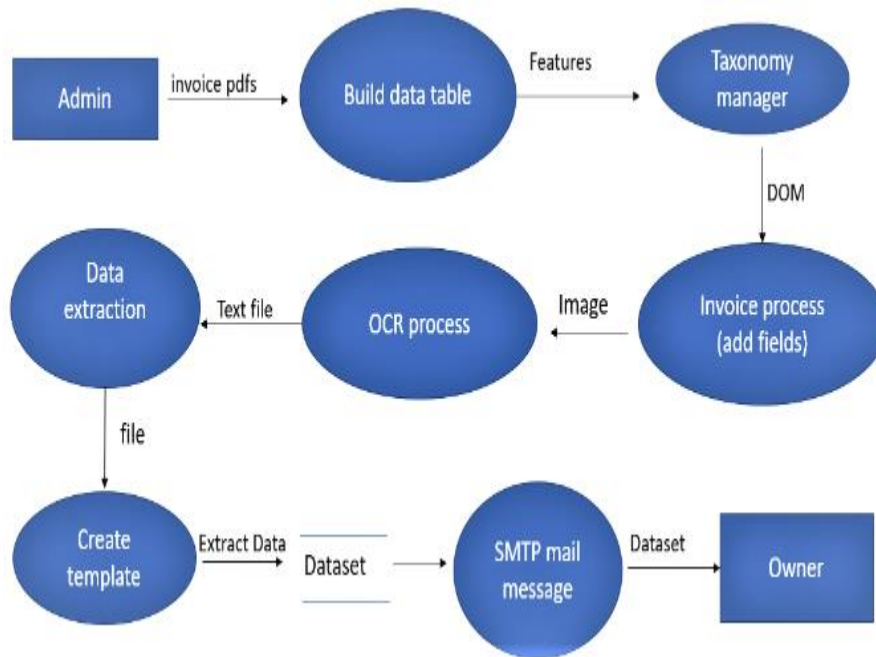


Fig. 1: Data Flow Diagram

III. Functions

Identify the PDF files: Determine the source and location of the PDF files that contain the data want to extract. To open UiPath Studio and create a new project to work on. Add necessary activities of Drag and drop the necessary activities from the Activities panel onto the workflow canvas to perform the required tasks. Use the "Assign" activity to define a variable that stores the path to the folder where to search for PDF files and the "Directory.GetFiles" activity to retrieve a list of file paths within the specified folder. Set the filter to "*.pdf" to get only PDF files then "For Each" loop activity to iterate over each file path obtained from the previous step. Within the loop perform further operations on each PDF file, such as extracting data or performing specific actions [5]. Implement PDF-related operations work with PDF files, may need to install and use third-party activities or custom libraries in UiPath. One popular library for PDF-related operations in UiPath is the "UiPath.PDF.Activities" package, which provides activities like "Read PDF Text" to extract text from PDF files. The important to handle any exceptions that may occur during the process, such as if a folder doesn't exist or there are no PDF files found. Use "Try Catch" activities to catch exceptions and handle them appropriately [7]. Once implemented the necessary activities, run the automation by clicking the "Run" button in UiPath Studio or by executing the workflow from UiPath Orchestrator are working with a published process.

Taxonomy module: The process of taxonomy defining and organizing a structured classification system for data or entities. Organization helps in standardizing and categorizing information, making the easier to manage and analyse. The taxonomy structure and categories that best fit data or entities. For example, working with customer data, the taxonomy could include categories like name, address, phone number, email, etc. Create a taxonomy file in a spreadsheet format (e.g., Excel file) or a text file. Each row in the file represents a category or a subcategory in the taxonomy, and the columns represent different attributes or properties associated with the category. Use the "Import Taxonomy" activity in UiPath to import the taxonomy file [1]. This activity allows to specify the file path and format of the taxonomy file. Once the taxonomy is imported to use to automation workflows to classify or categorize data. For example, extract information from a document or web page and assign to the relevant taxonomy category using the "Assign" activity or custom code. This changes directly in the imported taxonomy file or use the "Manage Taxonomy" activity in UiPath to add, delete, or modify categories and attributes. Test automation workflows with the taxonomy module to ensure that the classification and categorization are working correctly. Verify that the data is being assigned to the appropriate categories as per the defined taxonomy [3]. The best practices of taxonomy design, such as keeping the taxonomy simple and intuitive, using consistent naming conventions, and involving relevant stakeholders to ensure that the taxonomy meets business requirements by improve data analysis capabilities, and create more structured and standardized automation workflows.

Optical Character Recognition: OCR (Optical Character Recognition) is a technology that allows the extraction of text from images or scanned documents [11]. In the context of RPA with UiPath, admin plays a crucial role in automating tasks that involve working with unstructured data or data embedded in images [6]. UiPath provides a module that enables developers to incorporate capabilities into their automation workflows. UiPath OCR module offers a set of activities specifically designed for related tasks. These activities can be found in the UiPath.UIAutomation.Activities package and can be accessed from the Activities panel in UiPath Studio. Image Processing of Prior to performing OCR, that module allows to apply various image processing techniques to enhance the quality of the input image. These techniques include resizing, cropping, and applying filters to remove noise or improve contrast [12]. The OCR activities in UiPath provide a range of functionalities, such as extracting text from images, retrieving the location of specific text elements, and performing text recognition with OCR engines. The activities can be used to capture text from various sources like scanned documents, images, PDF files, or even screen captures. admin is essential to choose the appropriate language to achieve accurate results when working with multi documents. To ensure accuracy of module provides validation and correction capabilities [9]. This leverage activities to validate the extracted

text against specific patterns or formats, apply custom logic to correct recognized text, or use intelligent algorithms to reconcile discrepancies. The OCR module seamlessly integrates with other UiPath capabilities, such as data scraping, document understanding, and intelligent automation. This can be combined with UiPath activities from different domains to automate end-to-end processes involving document processing, data extraction, and data validation.

Extract the desired data: The text from of PDF extracted uses string manipulation, regular expression, or other text processing techniques to extract this specific data need. This could involve identifying patterns, keywords or specific sessions within the PDF. Admin can be a storage, a desktop application, a database, an Excel file, a PDF document, or any other structured or semi-structured data source. Use the "Data Scraping" wizard or the "Extract Structured Data" activity to extract data from websites or applications with tabular structures [4]. This technique allows to select the elements and define the extraction patterns. The "Screen Scraping" wizard or the "Get Text" activity to extract data from desktop applications or websites with non-tabular structures. This technique allows to indicate the target elements on the screen and extract their text or other attributes. The "Document Understanding" capabilities in UiPath to extract structured data from PDFs or other document formats [12]. This involves training and using machine learning models to identify and extract relevant information. Such activities like "Read Range" or "Read Cell" to extract data from Excel files convert text regular expressions, or custom logic to extract specific data patterns from text or string variables implement the chosen extraction technique by using the relevant activities or methods [6]. Configure the activities to specify the input source, extraction patterns, target elements, and other parameters based on requirements. Validate the extracted data to ensure accuracy and reliability techniques like data validation, comparison with expected results, or rule-based validations to verify the extracted data.

Document Object Model: The Document Object Model (DOM) is indeed utilized in RPA for scraping purposes. This structured content of a text document as a hierarchical tree of objects, where each element is represented as a node in the tree. This activities like "Find Element" or "Find Image" to locate specific elements on the document using selectors or other identification methods [11]. Selectors are expressions that uniquely identify elements based on their attributes like invoice number, due date or name. Once the desired elements are located, extract data from them using activities such as "Get Text", "Get Attribute", or "Get Full Text". These activities retrieve the text, attribute values, or other relevant information from the targeted elements. RPA bots can navigate through the DOM tree by interacting with elements. Activities like "Click", "Type Into", or "Select Item" can be used to interact with buttons, input fields, dropdowns, or other interactive elements on the source [8]. When scraping multiple similar elements, loops or iteration constructs, such as "For Each" or "While", to iterate through a collection of elements and extract data from each one. Once the data is extracted to stored in the variables, data tables, or external files for further processing, analysis, or integration with other systems.

Excel application:The Excel application in UiPath involves using activities from the UiPath.Excel.Activities package to read, write, and manipulate data in Excel files. The Excel Application Scope Activity used to Drag and drop the "Excel Application Scope" activity from the Activities panel onto Workflow [4]. This activity allows to work with Excel files and contains all the subsequent Excel-related activities. Specify the path of the Excel file want to work with by providing the file's location in the Workbook Path property of the Excel Application Scope activity [7]. To read data from Excel, use the "Read Range" activity within the Excel Application Scope. Specify the range of cells or the sheet name want to read in the Range property and provide the output variable (e.g., DataTable) to store the data. UiPath provides various activities to manipulate data in Excel files. This activities like "Filter Data Table," "Sort Data Table," or "Invoke VBA" to perform operations such as filtering, sorting, or executing [2]. Once finished working with Excel, make sure to include the necessary steps to save and close the Excel application within a proper control flow structure "Save Workbook" activity to save the changes made to the Excel file and the "Close Workbook" activity to close the Excel application.

Write the data to excel: The Excel activities to write the extracted data into the selected Excel sheet. The "Write Cell activity" allows to write data to a specific cell, while the "Write Range activity" allows to write a range of data to Excel, until all the data is extracted and stored in the Excel sheets [10]. Add the necessary activities or actions to read and extract the data want to write to Excel. This could involve interacting with applications, databases, or APIs to retrieve the data [1]. Once the data extracted and stored in variables or data structures proceed with writing admin into to an Excel file. Use the appropriate activity or action provided by RPA tool to write data to Excel [12]. Drag and drop an "Excel Application Scope" activity onto the workflow. In the properties panel of the "Excel Application Scope" activity, specify the path of the Excel file want to write the data Within the "Excel Application Scope," use either the "Write Cell" or "Write Range" activity based on requirements.

Email Automation: SMTP (Simple Mail Transfer Protocol) is the standard protocol for sending email messages over the internet. The "Send SMTP Mail Message" activity in UiPath allows to send emails Sending emails programmatically automate communication and streamline business processes within RPA workflows [8]. The SMTP mail message activity to send RPA workflows can generate reports or extract data from various sources. In the properties panel of the "Send SMTP Mail Message" activity, need to provide the necessary details. This Server Specify the SMTP server address and port number. For example, "smtp.gmail.com" for Gmail with port 587. Authentication provide the username and password of the sender's email account for authentication [3]. Sender Set the sender's email address and assign a constant value or retrieve owner from a variable. Once provided all the required details, save and run automation workflow. During execution, the "Send SMTP Mail Message" activity will establish a connection with the SMTP server and send the email with the specified content and attachments [8]. Ensure that RPA environment has network connectivity to access the SMTP server and the necessary permissions to send email Additionally, handle any potential exceptions or errors that may occur during the email sending process to ensure the reliability of RPA workflows.

IV. Conclusion

In conclusion, Invoice processing system along with chat bot using RPA offers numerous benefits, such as increased efficiency, accuracy, and scalability by leveraging the power of RPA and utilizing UiPath's automation capabilities, organizations can streamline the invoice process and improve overall productivity. The use of the RPA workflows provides a structured approach, ensuring robustness, excel automate and traceability.

V. Future work

As a future scope, RPA chatbots to extract invoice PDFs with higher accuracy, even from complex layouts or handwritten documents. This would enable to extract relevant information more intelligently, categorize data, and perform advanced analysis. Integrating advanced document understanding capabilities, such as machine learning, can enhance the system's ability to extract and process invoice data accurately. This includes automating data extraction, validation, and reconciliation from various invoice formats, reducing manual effort and increasing efficiency. This advanced technology used to custom extraction algorithm focus on several areas to improve accuracy, scalability, and usability. Admin work techniques can improve the algorithm's ability to extract data accurately from complex and unstructured PDF documents. Training the algorithm on a diverse dataset can help it learn patterns, identify key data fields, and adapt to different document layouts.

VI. References

1. Devanshi Desai , Ansh Jain , Dhaivat Naik , Nishita Panchal , Dattatray Sawant "INVOICE PROCESSING USING RPA & AI", International Conference on Smart Data Intelligence (ICSMDI2021) , Available at SSRN: <https://ssrn.com/abstract=3852575>.
2. Larsson, A. and Segeras, T., 2016. Automated invoice handling with machine learning and OCR. 1st ed. Stockholm, Sweden: Degree Projects in Computer Engineering.
3. Mullakara, N., &Asokan, A. K. (2020). Robotic process automation projects: Build real-world RPA solutions using UiPath and Automation Anywhere. Birmingham: Packt Publishing.
4. RPI consultants, "End to End Accounts Payable Automation for Invoice Processing" June 2018.
5. Oracle, "Process Automation for Accounts Payable," November 2015. Eduardo B. Fernandez and Xiaohong Yuan, "An Analysis Pattern for Invoice Processing" in 16th PLoP/ACM, Chicago, IL, 2009, pp.1-10.
6. Priscila Cedillo, Andrés García, Juan Diego Cárdenas, Alexandra Bermeoz, "A systematic literature review of Electronic invoicing, platforms and notification systems," in 5th ICEDEG/IEEE, Ambato,Ecuador, 2018, pp. 150-157.
7. Benjamin Pessi, "The impact of implementation of the electronic purchase invoice system on a company example of Hahle group," 2017.
8. J. Luo and J. Cross, "Advanced Image Processing Techniques for Maximum Information Recovery," 2007 Thirty-Ninth Southeastern Symposium on System Theory, Macon, GA, 2007, pp. 58-62, doi: 10.1109/SSST.2007.352317.
9. Russ J.C. (1990) Image Processing. In: Computer-Assisted Microscopy. Springer, Boston, MA. https://doi.org/10.1007/978-1-4613-0563-7_3
10. J. M. White and G. D. Rohrer, "Image Thresholding for Optical Character Recognition and Other Applications Requiring Character Image Extraction," in IBM Journal of Research and Development, vol. 27, no. 4, pp. 400-411, July 1983, doi: 10.1147/rd.274.0400.
11. R. Smith, "An Overview of the Tesseract OCR Engine," Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Parana, 2007, pp. 629-633, doi: 10.1109/ICDAR.2007.4376991.
12. Leptonica library for image support <https://github.com/DanBloomberg/leptonica>