Mortgage Document Extraction – UiPath

-By Yashwant Sai Kishore Kandregula

About Automation:

This automation is used to extract the required fields from Closing Disclosure document.

User Guide:

1. **INPUT:** Place the pdfs in the input folder of the automation.
2. **OUTPUT:** Trigger the job, output csv contains the extracted fields of all the input files.
3. Logs and errors can be observed form the orchestrator logs.

Version Control:

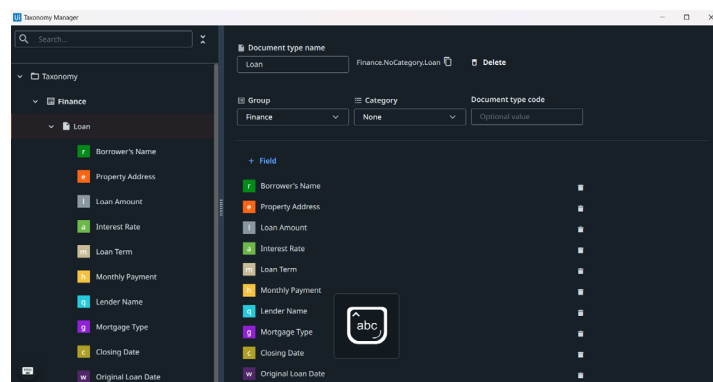Github: https://github.com/yashwant43/MortgageDocumentExtractRPA

Architecture:

1. This automation uses UiPath Document Understanding.
2. Automation is a linear process where each input file is read, and fields are extracted to csv file.

Dependencies:

1. UiPath.DocumentUnderstanding.ML.Activities=1.17.0
2. UiPath.IntelligentOCR.Activities=6.5.0
3. UiPath.OmniPage.Activities=1.17.0

Automation Steps:

1. For each file in the input folder, bot iterates and stores the file path into str_InputFolder.
2. Taxonomy is created, fields and their respective data types are added.

3. Load Taxonomy activity is used to load the created taxonomy and the output is stored in Taxonomy variable.
4. Str_DocumentPath is assigned to the file path.
5. Digitize Document activity is used to digitize the pdf to text. The output is saved to str_DocumentText and meta data to DOM variable. OmniPage OCR is used.
6. Classify Document Scope is used to classify the type of document which is loan.

Keyword based classifier is used and necessary keywords are added. Results are stored to classificationResults variable.

7. Data Extraction Scope activity is used to extract the fields. Extracted results are stored in extractionResults variable. Form Extractor is used extract the fields.

   API value is retrieved from UiPath Admin->Licences.

   Template Manager is used to annotate the fields. Each filed location is marked and saved.

8. Present Validation Station is used to validate the results.
9. Export Extraction Results activity is used to extract the results to dataset variable.
10. Output datable of extracted fields is present in dt_Output variable.
11. File Path and Timestamp columns are added and data is stored in output.csv.

Screenshots:

Input:

Output:

Borrower's Name	Property Address	Loan Amount	Interest Rate	Loan Term	Monthly Payment	Lender Name	Mortgage Type	Closing Date	Original Loan Date	File Path
Michael Jones and Mary Stone	456 Somewhere Ave Anytown, ST 1234	$162,000	3.88%	30 years	$761.78	Ficus Bank	Fixed Rate	4/15/2013	4/15/2013	F:\OneDrive\Documents\UiPath\MortgageDocument_Extraction
Tim David and Alex Brown	317 DaVinci Ave Anytown, ST 54321	$178,000	4.19%	40 years	$542.78	LIB Bank	Fixed Rate	8/18/2023	6/15/2020	F:\OneDrive\Documents\UiPath\MortgageDocument_Extraction