**Uipath Demo for Project3 from THETA Team:**

**Objective:**

To automate data extraction from large number of PDF files and consolidate all extracted data into a spreadsheet in a most speedy manner.

The challenge for this project is to accurately extract at least 18 items from 1000 e-Invoices within one minute

**Approaches: 3 programs are developed and evaluated:**

Extract\_PDF\_data\_via\_screen\_scraping.xaml

- Launch each file to scrap data in foreground. Speed is extremely slow.

- 2 Scraping Options are available.

1.Modern Design experience: (set IsClassicMode to False in Main.xaml)

Speed - complete 1.5 files in 45sec. (lowest speed)

For 1st Demo

2.Classic Design experience: (set IsClassicMode to True in Main.xaml)

Speed - complete 3 files in 45sec. (very slow speed)

Below is the video link for the 1st and 2nd Demo:

https://leemunhoi-gmail.tinytake.com/msc/NTc5MzE1OV8xNzgxOTQxMw

Extract\_PDF\_data\_via\_Taxonomy\_Manager.xaml

- Use Taxonomy Manager to build data fields and use Document Understanding tools to digitize (via OmniPage OCR), classify file contents and extract data via various build-in extraction scopes in background. 2 types of Extractor Scopes were tried:

For 2nd Demo

1.Form Extractor: (set IsFormExtractor to True in Manin.xaml)

Note: Need Developer Api Key to access EndPoint site to enable the form template editor which is used to mark the data position via an existing PDF file as template.

Limitation: **Community licensed only allow to run 50 pdf files within an hour.**

Speed – complete 25 files in 40sec. (far from expectation)

For 3rd Demo

2.RegexBasedExtractor: (set IsFormExtractor to False in Main.xaml)

No need Api key but Developer needs to program specific Regex patterns to match each field.

Speed – complete 150 files in 40sec. (fail to beat the challenge)

Extract\_PDF\_data\_via\_Regex\_match.xaml

- Convert all texts of each PDF file in background and use one Regex match pattern to extract all the required data in one go and apply strings manipulation to fine-tuning each extracted data.

- 2 Options are available.

1.Direct Append extracted data to text file in CSV format without data table. (set IsDirecAppendToCSV to True in Main.xaml)

Speed – complete 1000 files in 40sec. (Meet the Challenge’s Goal)

For 4th Demo

2.Export extracted data to excel file via data table. (set IsDirecAppendToCSV to False in Main.xaml)

Speed – complete 1000 files in 40sec. (Meet the Challenge’s Goal)

Below is the video link for the 3nd and 4th Demo:

https://leemunhoi-gmail.tinytake.com/msc/NTc5MzIwMF8xNzgxOTQ4Ng