Document Understanding Evaluation Template (Ver 0.2.3)

Masaki Kumamoto
Sales Engineer
UiPath Canada

UiPath Reboot™ Work.

Slide update

Date	Document Version	Description
Jun 3, 2021	0.0.1_doc0.1	Initial document
Jun 8, 2021	0.0.3_doc0.1	Update for v0.0.3
Jun 8, 2021	0.0.4_doc0.1	Update for v0.0.4
Jun 8, 2021	0.0.5_doc0.1	Update for v0.0.5
Jun 9, 2021	0.0.7_doc0.1	Update for v0.0.7
Jul 16, 2021	0.2.1_doc0.1	Update for v0.2.1
Mar 24, 2022	0.2.3_doc0.1	Update for v0.2.3

Agenda

1. **Overview**
   - What is Document Understanding Evaluation Template?
   - Use cases
   - Features

2. **Quick start guide**
   - Preparation Steps
   - Development Steps
   - Execution Steps
   - Improvement Steps

3. **Details of the reporting files**
   - DU_Evaluation.xlsx
   - ActionList.xlsx

4. **Other useful features**
   - Multi OCR execution
   - Auto verification confidence threshold
   - Show OcrConfidence in validation action
   - Use existing Action.xlsx as an actual value reference

5. **Specifications**

6. **Release note**

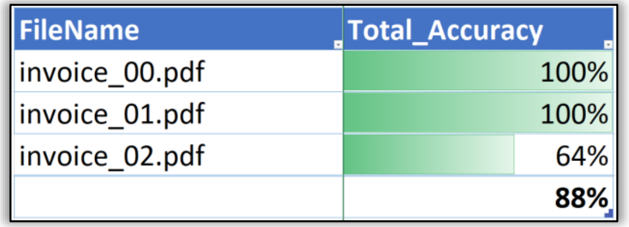
Document Understanding Evaluation Template

1

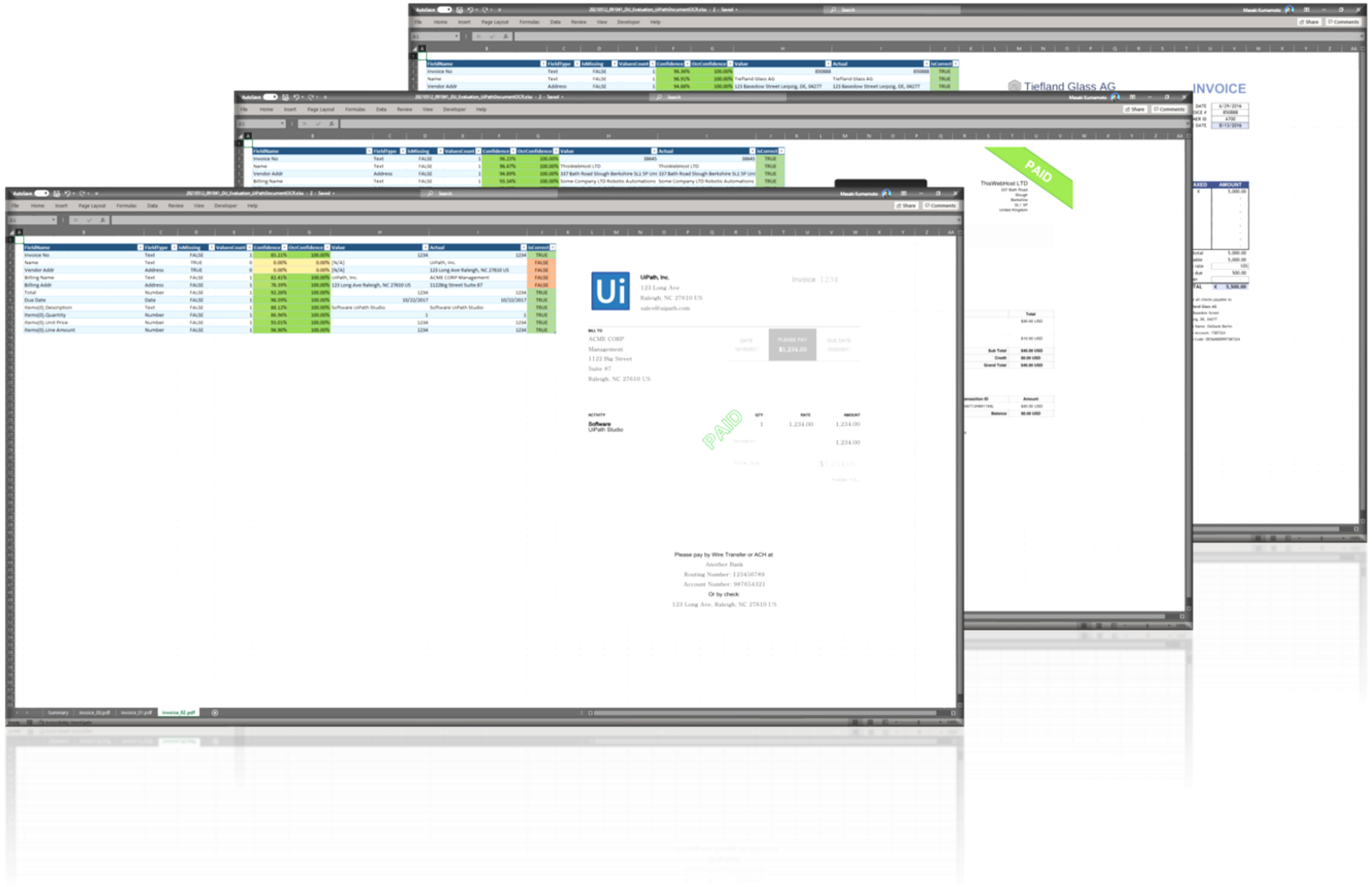
Overview

What is Document Understanding Evaluation Template?

This template project facilitates the efficient development of workflows that output information about the extraction accuracy of Document Understanding in a beautiful Excel format automatically.



FileName	Total_Accuracy
invoice_00.pdf	100%
invoice_01.pdf	100%
invoice_02.pdf	64%
	88%



FieldName	FieldType	IsMissing	ValuesCount	Confidence	OcrConfidence	Value	Actual	IsCorrect
Invoice No	Text	FALSE	1	85.21%	100.00%	1234	1234	TRUE
Name	Text	TRUE	0	0.00%	0.00%	[N/A]	UiPath, Inc.	FALSE
Vendor Addr	Address	TRUE	0	0.00%	0.00%	[N/A]	123 Long Ave Raleigh, NC 27610 US	FALSE
Billing Name	Text	FALSE	1	82.41%	100.00%	UiPath, Inc.	ACME CORP Management	FALSE
Billing Addr	Address	FALSE	1	78.39%	100.00%	123 Long Ave Raleigh, NC 27610 US	1122Big Street Suite 87	FALSE
Total	Number	FALSE	1	92.26%	100.00%	1234	1234	TRUE
Due Date	Date	FALSE	1	96.59%	100.00%	10/22/2017	10/22/2017	TRUE
Items[0].Description	Text	FALSE	1	88.12%	100.00%	Software UiPath Studio	Software UiPath Studio	TRUE
Items[0].Quantity	Number	FALSE	1	86.96%	100.00%	1	1	TRUE
Items[0].Unit Price	Number	FALSE	1	93.01%	100.00%	1234	1234	TRUE
Items[0].Line Amount	Number	FALSE	1	96.90%	100.00%	1234	1234	TRUE

Features

Minimum Effort Development

You only have to edit 3 files!

Config.xlsx

taxonomy.json
(Taxonomy Manager)

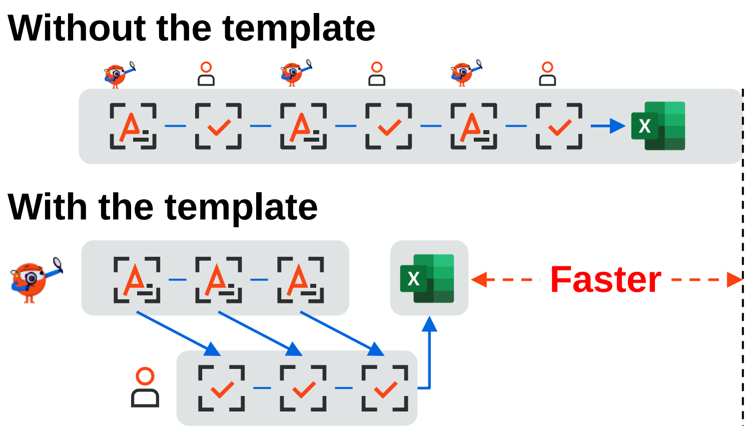
DU

You can just focus on…

- Taxonomy definition
- Classification/Extraction logic

You do not even need to add any additional activities, valuables and arguments!

Faster Validation Flow

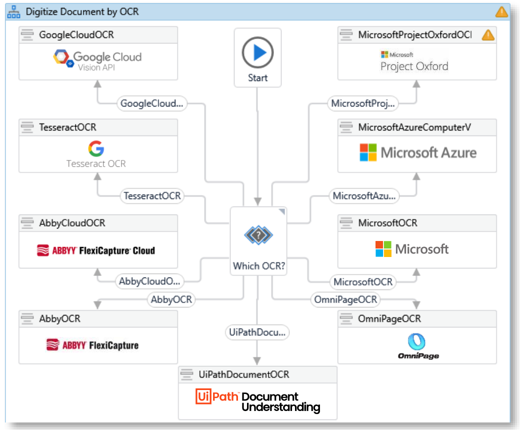


Without the template

With the template

Since Validation Actions will be uploaded in bulk, the lead time can be shortened because the “Extraction by Robots” and “Validation by human" can be processed in parallel for multiple documents.

Multiple OCRs Evaluation



By enabling specific OCRs in Config.xlsx, you can apply multiple OCRs to the DU and create evaluation reports for each OCR simultaneously.

Use cases

DU accuracy reporting

Rendering the extraction accuracy of a DU for given documents in a beautiful Excel format automatically to assist in evaluating the functionality of the DU

OCR Comparison

Comparing the extraction accuracy of each OCR applied to DU

Improve DU accuracy

Optimizing the development process to improve DU's document extraction accuracy based on the benchmarks

Action Center Demo

Demonstration for building Document Validation Actions in the Action Center

Document Understanding Evaluation Template 2

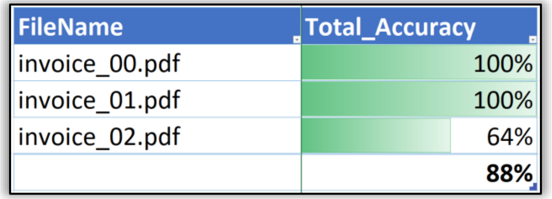
Details of the reporting files

UiPath Reboot™ Work.

DU_Evaluation.xlsx

This file contains the percentage of correct extractions for all target documents and detailed extraction results. Files will be generated for the number of OCRs defined in Config.xlsx.

Summary sheet

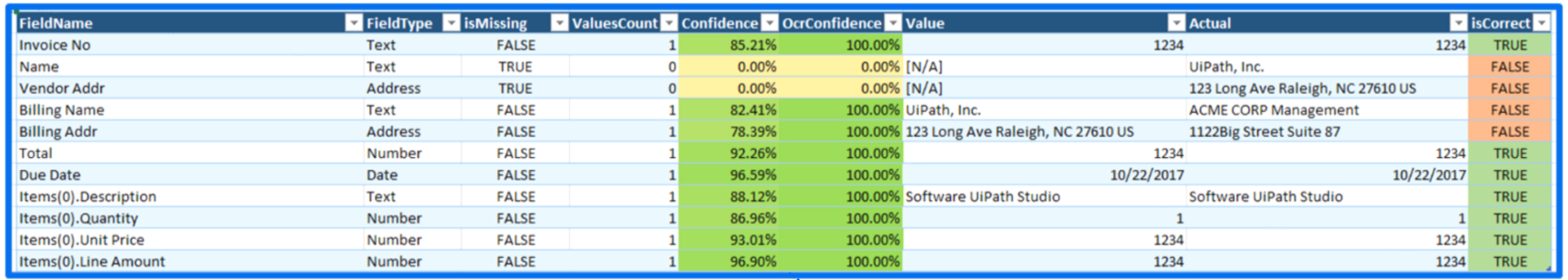


FileName	Total_Accuracy
invoice_00.pdf	100%
invoice_01.pdf	100%
invoice_02.pdf	64%
	88%

This sheet renders the percentage of correct extractions for all target documents extracted by DU.

Header name	Description
Boolean_Accuracy	Correct extraction rate for “Boolean” field
Others_Accuracy	Correct extraction rate for “Text”, “Number”, “Date”, “Name”, “Address” and “Set” fields
Total_Accuracy	Correct extraction rate all the fields

Extraction/Actual value report sheets (per target documents)



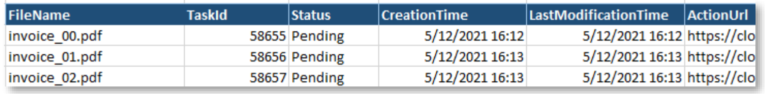
FieldName	FieldType	isMissing	ValuesCount	Confidence	OcrConfidence	Value	Actual	isCorrect
Invoice No	Text	FALSE	1	85.21%	100.00%	1234	1234	TRUE
Name	Text	TRUE	0	0.00%	0.00%	[N/A]	UiPath, Inc.	FALSE
Vendor Addr	Address	TRUE	0	0.00%	0.00%	[N/A]	123 Long Ave Raleigh, NC 27610 US	FALSE
Billing Name	Text	FALSE	1	82.41%	100.00%	UiPath, Inc.	ACME CORP Management	FALSE
Billing Addr	Address	FALSE	1	78.39%	100.00%	123 Long Ave Raleigh, NC 27610 US	1122Big Street Suite 87	FALSE
Total	Number	FALSE	1	92.26%	100.00%	1234	1234	TRUE
Due Date	Date	FALSE	1	96.59%	100.00%	10/22/2017	10/22/2017	TRUE
Items(0).Description	Text	FALSE	1	88.12%	100.00%	Software UiPath Studio	Software UiPath Studio	TRUE
Items(0).Quantity	Number	FALSE	1	86.96%	100.00%	1	1	TRUE
Items(0).Unit Price	Number	FALSE	1	93.01%	100.00%	1234	1234	TRUE
Items(0).Line Amount	Number	FALSE	1	96.90%	100.00%	1234	1234	TRUE

Header name	Description
FieldName	Field name
FieldType	Field type
isMissing	If extractor missed the field or not
ValuesCount	Numbers of values which was extracted
Confidence	Confidence level for location
OcrConfidence	Confidence level for OCR
Value	Extracted value by DU
Actual	Validated value (Actual value)
isCorrect	If the extracted value is correct or not

ActionList.xlsx

This file contains information of the generated Document Validation Actions and the values.
This file will be used by Robots to get the validation results.

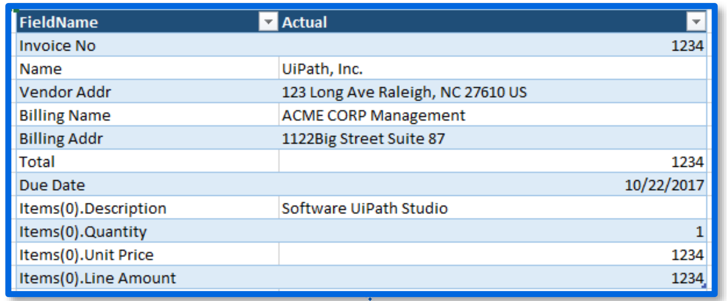
Actions sheet



FileName	TaskId	Status	CreationTime	LastModificationTime	ActionUrl
invoice_00.pdf	58655	Pending	5/12/2021 16:12	5/12/2021 16:12	https://clo
invoice_01.pdf	58656	Pending	5/12/2021 16:13	5/12/2021 16:13	https://clo
invoice_02.pdf	58657	Pending	5/12/2021 16:13	5/12/2021 16:13	https://clo

Header name	Description
FileName	Target document file name
TaskId	Task Id of the Action
Status	Status of the Action
CreationTime	Creation time of the Action
LastModificationTime	Last modification time of the Action
ActionUrl	URL of the Action

Actual value report sheets (per target documents)



FieldName	Actual
Invoice No	1234
Name	UiPath, Inc.
Vendor Addr	123 Long Ave Raleigh, NC 27610 US
Billing Name	ACME CORP Management
Billing Addr	1122Big Street Suite 87
Total	1234
Due Date	10/22/2017
Items(0).Description	Software UiPath Studio
Items(0).Quantity	1
Items(0).Unit Price	1234
Items(0).Line Amount	1234

Header name	Description
FieldName	Field name
Actual	Validated value (Actual value)

Document Understanding Evaluation Template

3

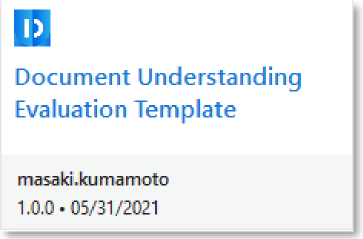
Quick Start Guide

UiPath Reboot™ Work.

Quick Start Guide

After creating a new project from the Document Understanding Evaluation Template, you can accomplish the goal with a minimum of effort by following the steps described below.
(Detailed steps are in the following pages)

Preparation

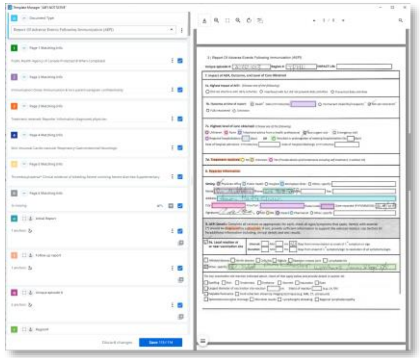


Document Understanding Evaluation Template

masaki.kumamoto
1.0.0 • 05/31/2021

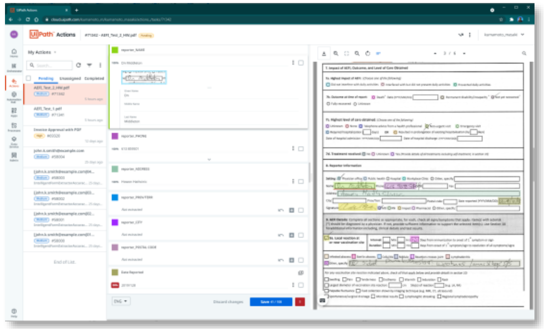
1. Place template project nuget package in the template folder
2. Create a new project from the template

Development



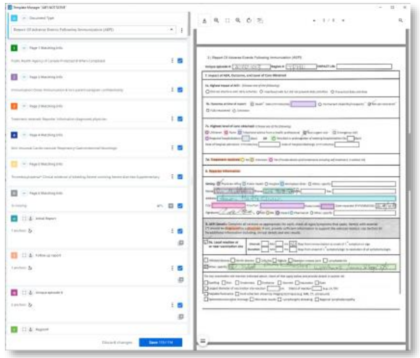
1. Place target documents in Input folder
2. Configure Config.xlsx
3. Define Taxonomy
4. Build DU_GetExtractionResult.xaml

Execution



1. Run 01_ExtractDocumentsData.xaml
2. Complete the Document Validation Action task in Action Center
3. Run 02_CopyActualValuesToReport.xaml

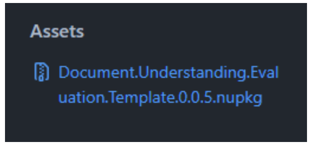
Improvement



1. Use existing ActionList.xlsx for further development to improve the DU logic

Preparation Steps

1. **Place template project nuget package in the template folder**
   - Download Document.Understanding.Evaluation.Template.X.X.X.nupkg (Download from Assets)

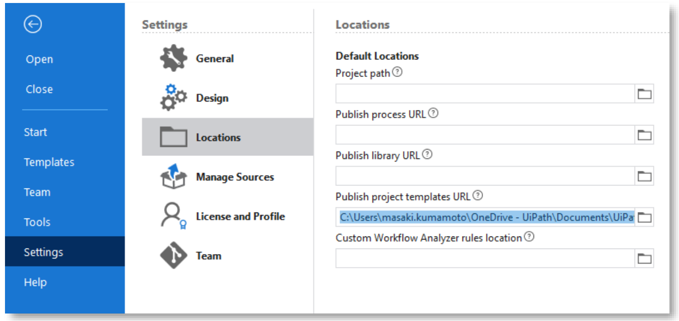


Assets

Document.Understanding.Evaluation.Template.0.0.5.nupkg

   - You can identify the template folder below. UiPath Studio > Settings > Locations > Publish project templates URL

     e.g. “C:\Users\{UserName}\Documents\UiPath\.templates”



Locations

Default Locations

Project path

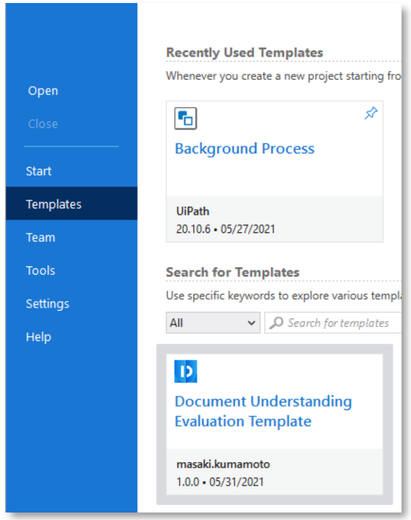
Publish process URL

Publish library URL

Publish project templates URL: C:\Users\masaki.kumamoto\OneDrive - UiPath\Documents\UiPa

Custom Workflow Analyzer rules location

2. **Create a new project from the template**
   - UiPath Studio > Templates > Document Understanding Evaluation Template



Recently Used Templates

Whenever you create a new project starting fro

Background Process

UiPath
20.10.6 • 05/27/2021

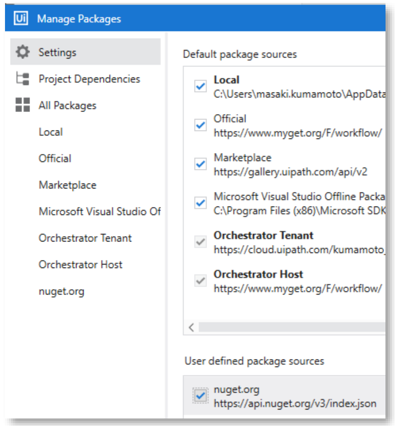
Search for Templates

Use specific keywords to explore various templa

Document Understanding Evaluation Template

masaki.kumamoto
1.0.0 • 05/31/2021

3. **Wait until the project will resolve the dependency**
   - If Studio can’t auto-resolve the dependency for Microsoft.Office.interop.Excel, enable nuget.org (https://api.nuget.org/v3/index.json) as a package source in package manager and install the package.



Default package sources

☑ Local
C:\Users\masaki.kumamoto\AppData

☑ Official
https://www.myget.org/F/workflow/

☑ Marketplace
https://gallery.uipath.com/api/v2

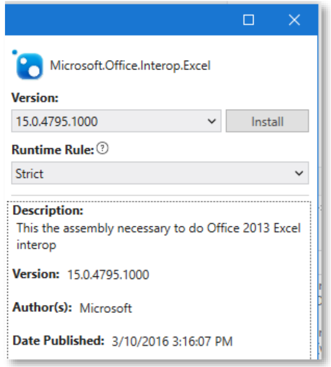
☑ Microsoft Visual Studio Offline Packa
C:\Program Files (x86)\Microsoft SDK

☑ Orchestrator Tenant
https://cloud.uipath.com/kumamoto

☑ Orchestrator Host
https://www.myget.org/F/workflow/

User defined package sources

☑ nuget.org
https://api.nuget.org/v3/index.json



Microsoft.Office.Interop.Excel

Version: 15.0.4795.1000

Runtime Rule: Strict

Description: This the assembly necessary to do Office 2013 Excel interop

Version: 15.0.4795.1000

Author(s): Microsoft

Date Published: 3/10/2016 3:16:07 PM

Development Steps

1. **Place target documents in Input folder**
   - If you use Form Extractor/Intelligent Form Extractor, you can also place template documents in "TemplateDocument" Folder).
   - The file name placed in the Input Folder should not exceed 31 characters including the file extension. If the number of characters exceeds 31, an error will occur when creating the sheet in the Report Excel file.

2. Configure Config.xlsx

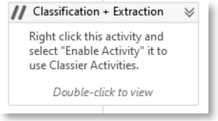
- **"DUSettings"** Sheet
  - DU_ApiKey
  - DU_DocumentTypeId
    (If you use Classification activity, you do NOT need to specify this field. You can find the value in Taxonomy Manager once you create a taxonom in step 3.)
- **"ActionSettings"** Sheet
  - AC_AssignUserEmail
  - OC_FolderPath
    (Orchestrator Folder name that your Studio/Robot is deployed in)
  - SB_BacketName
    (Set the same name Storage Bucket in Orchestrator)
- **"OcrSettings"** Sheet
  - Set TRUE for the OCR to be applied to DU. You can enable multiple OCRs to be applied. UiPath Document Understanding OCR will always be performed.

3. Define Taxonomy

- Set the definition of the field information to be extracted from Ribbon>Design>Taxonomy Manager. (How to use Taxonomy Manager)

4. Build DU_GetExtractionResult.xaml

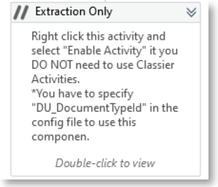
- If you want to use Classification, Enable “Classification + Extraction”
  If not,
  Enable “Extraction Only”



Classification + Extraction

Right click this activity and select "Enable Activity" it to use Classier Activities.

Double-click to view

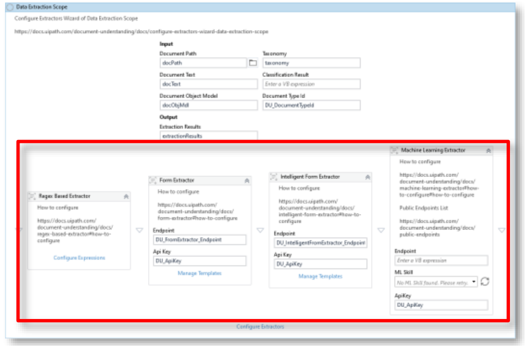


Extraction Only

Right click this activity and select "Enable Activity" it you DO NOT need to use Classier Activities.
*You have to specify "DU_DocumentTypeId" in the config file to use this componen.

Double-click to view

- Delete Classification/Extraction Activities which you do not use



Data Extraction Scope

Configure Extractors Wizard of Data Extraction Scope

https://docs.uipath.com/document-understanding/docs/configure-extractors-wizard-data-extraction-scope

Input

Document Path: docPath
Taxonomy: taxonomy
Document Text: docText
Classification Result: *Enter a VB expression*
Document Object Model: docObjMdl
Document Type Id: DU_DocumentTypeId

Output

Extraction Results: extractionResults

Regex Based Extractor
How to configure
https://docs.uipath.com/document-understanding/docs/regex-based-extractor#how-to-configure
Configure Expressions

Form Extractor
How to configure
https://docs.uipath.com/document-understanding/docs/form-extractor#how-to-configure
Endpoint: DU_FromExtractor_Endpoint
Api Key: DU_ApiKey
Manage Templates

Intelligent Form Extractor
How to configure
https://docs.uipath.com/document-understanding/docs/intelligent-form-extractor#how-to-configure
Endpoint: DU_IntelligentFromExtractor_Endpoint
Api Key: DU_ApiKey
Manage Templates

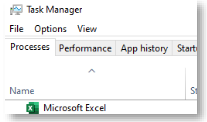
Machine Learning Extractor
How to configure
https://docs.uipath.com/document-understanding/docs/machine-learning-extractor#how-to-configure#how-to-configure
Public Endpoints List
https://docs.uipath.com/document-understanding/docs/public-endpoints
Endpoint: *Enter a VB expression*
ML Skill: *No ML Skill found. Please retry.*
ApiKey: DU_ApiKey

Configure Extractors

Execution Steps

1. Run 01_ExtractDocumentsData.xaml

- You should stop the OneDrive sync function while the process is running otherwise an error may occur.
- Make sure “Microsoft Excel” is not running even in background. This would cause unitability of the workflow somehow.

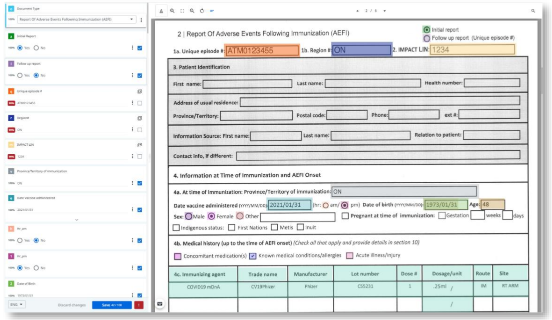


Task Manager
File Options View
Processes Performance App history Start
Name
Microsoft Excel

- It takes about 1-2 minutes to process each document. (It would take more with “Debug” so “Run” is recommended)
- After the execution is complete, Excel reports for each OCR set in Config.xlsx and the Document Validation Action in Action Center will be generated.

01_ExtractDocumentsData.xaml

2. Complete the Document Validation Action task in Action Center



2 | Report Of Adverse Events Following Immunization (AEFI)

● Initial report
○ Follow up report (Unique episode #)

1a. Unique episode #: ATM0123455 1b. Region #: ON 2. IMPACT LIN: 1234

3. Patient Identification

First name: Last name: Health number:

Address of usual residence:

Province/Territory: Postal code: Phone: ext #:

Information Source: First name: Last name: Relation to patient:

Contact info, if different:

4. Information at Time of Immunization and AEFI Onset

4a. At time of immunization: Province/Territory of immunization: ON

Date vaccine administered (YYYY/MM/DD): 2021/01/31 (hr: ○ am/ ● pm) Date of birth (YYYY/MM/DD): 1973/01/31 Age: 48

Sex: ○ Male ● Female ○ Other □ Pregnant at time of immunization: □ Gestation □ weeks □ days

□ Indigenous status: □ First Nations □ Metis □ Inuit

4b. Medical history (up to the time of AEFI onset) *(Check all that apply and provide details in section 10)*

□ Concomitant medication(s) ☑ Known medical conditions/allergies □ Acute illness/injury

4c. Immunizing agent	Trade name	Manufacturer	Lot number	Dose #	Dosage/unit	Route	Site
COVID19 mDnA	CV19Phizer	Phizer	C55231	1	.25ml /	IM	RT ARM
					/		

Ui Action Center

3. Run 02_CopyActualValuesToReport.xaml

- Immediately after the execution, the robot will prompt the user to select a folder where the DU evaluation reports are located.
- After the execution is completed, the results of the Document Validation Actions will be pasted to the ActionList.xlsx and the DU Evaluation Reports for each OCR.
- If there are documents that have not yet been validated by Document Validation Actions when Step 3 is completed, complete the validation in Validation Action and then execute Step 3 again to complete DU evaluation reports.

02_CopyActualValuesToReport.xaml

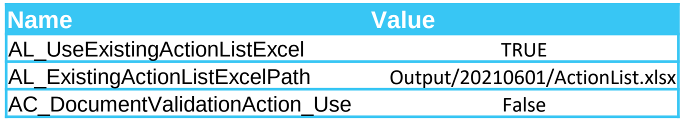
Improvement Steps

Use existing ActionList.xlsx to improve the DU logic

If you have performed the "Execution Steps" and generated ActionList.xlsx for the same list of documents using the same taxonomy in the past, from next time, you can skip step 2 & 3 by following the steps below. You can also disable to create Document Validation Action so the process can run faster

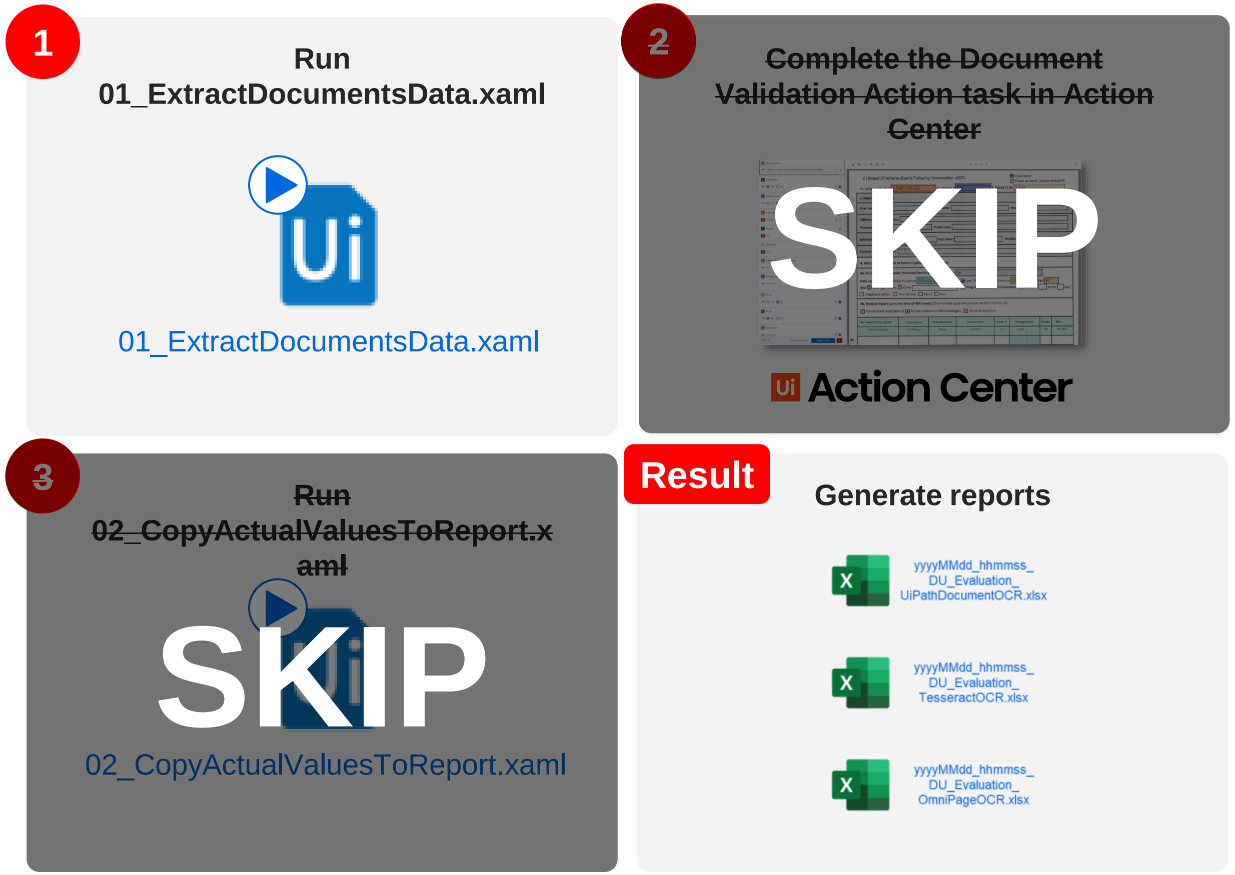
This capability is useful to modify the workflow based on the accuracy rate from previous execution result report, so you can improve the DU's classification/extraction logic.

- **Configure Config.xlsx**
  - "BasicSettings" Sheet
    - AL_UseExistingActionListExcel (= TRUE)
    - AL_ExistingActionListExcelPath
  - "ActionSettings" Sheet
    - AC_DocumentValidationAction_Use (= False)



Name	Value
AL_UseExistingActionListExcel	TRUE
AL_ExistingActionListExcelPath	Output/20210601/ActionList.xlsx
AC_DocumentValidationAction_Use	False

Config.xlsx
(BasicSettings, ActionSettings)



1 Run 01_ExtractDocumentsData.xaml

01_ExtractDocumentsData.xaml

2 ~~Complete the Document Validation Action task in Action Center~~

SKIP

Action Center

3 ~~Run 02_CopyActualValuesToReport.xaml~~

SKIP

02_CopyActualValuesToReport.xaml

Result Generate reports

yyyyMMdd_hhmmss_DU_Evaluation_UiPathDocumentOCR.xlsx

yyyyMMdd_hhmmss_DU_Evaluation_TesseractOCR.xlsx

yyyyMMdd_hhmmss_DU_Evaluation_OmniPageOCR.xlsx

Document Understanding Evaluation Template

4

Other useful features

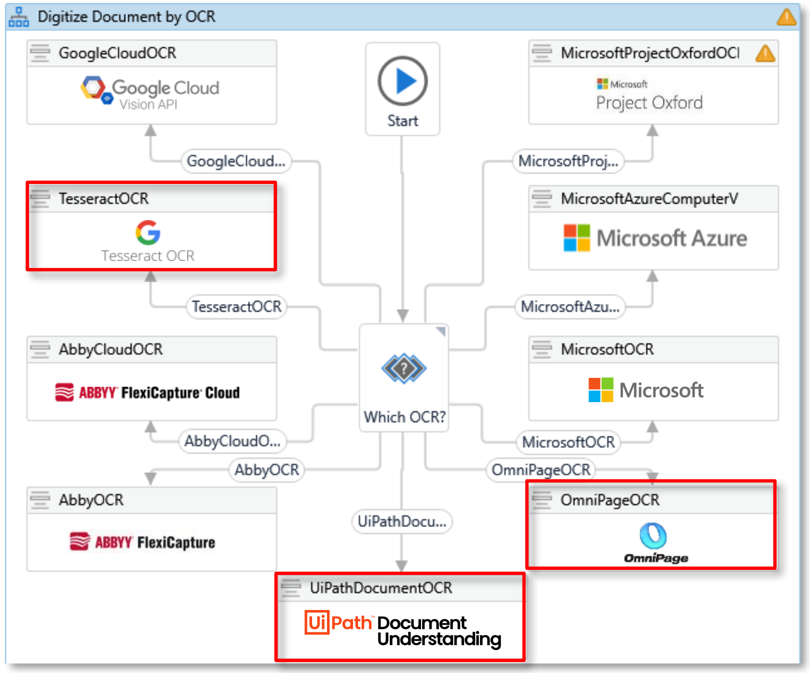
UiPath Reboot™ Work.

Multi OCR execution

By enabling specific OCRs in Config.xlsx, you can apply multiple OCRs to the DU and create evaluation reports for each OCR simultaneously. (UiPath Document OCR will be always used)

Name	Value
OCR_TesseractOCR_Use	TRUE
OCR_TesseractOCR_Language	eng
OCR_OmniPageOCR_Use	TRUE
OCR_OmniPageOCR_Language	eng

Config.xlsx (OcrSettings)



yyyyMMdd_hhmmss_ DU_Evaluation_ UiPathDocumentOCR.xlsx

yyyyMMdd_hhmmss_ DU_Evaluation_ TesseractOCR.xlsx

yyyyMMdd_hhmmss_ DU_Evaluation_ OmniPageOCR.xlsx

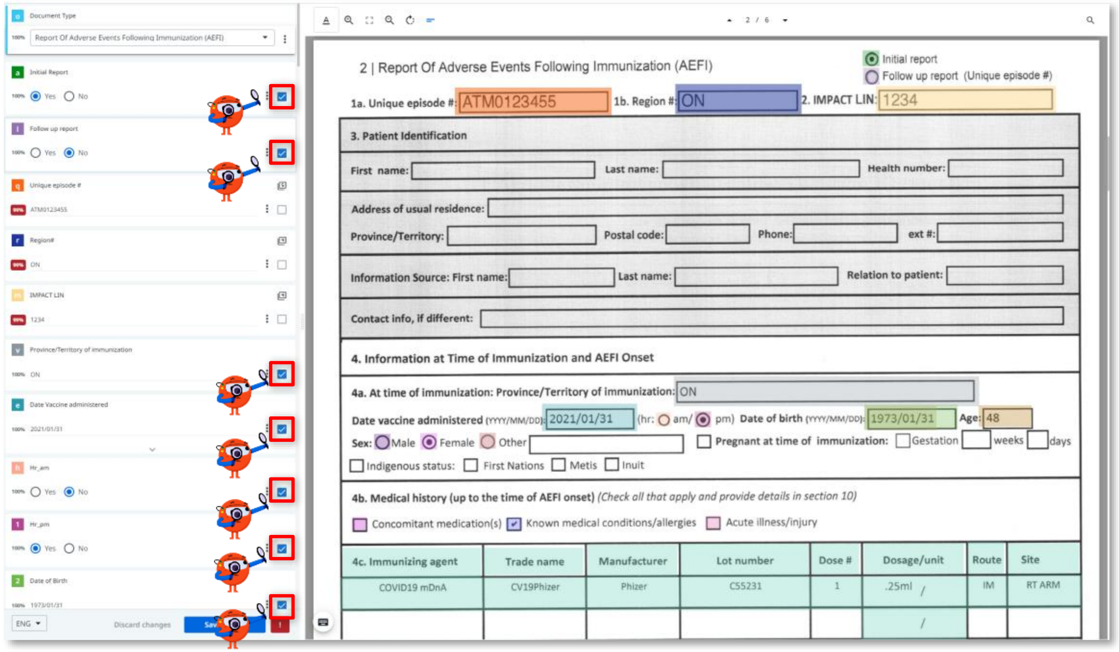
Auto verification confidence levels

You can use extraction Confidence and OcrConfidence as a threshold for auto verification.
If both of them are above or equal to thresholds, the fields will be automatically verified by Robots.

Confidence = Confidence level for location
OcrConfidence = Confidence level for OCR

Name	Value
DU_AutoVerifyMinimumThreshold_Confidence	99.98%
DU_AutoVerifyMinimumThreshold_OcrConfidence	95.99%

Config.xlsx
(DuSettings)



Document Type
100% Report Of Adverse Events Following Immunization (AEFI)

Initial Report
100% ● Yes ○ No ☑

Follow up report
100% ○ Yes ● No ☑

Unique episode #
99% ATM0123455 □

Region#
99% ON □

IMPACT LIN
99% 1234 □

Province/Territory of immunization
100% ON ☑

Date Vaccine administered
100% 2021/01/31 ☑

Hr_am
100% ○ Yes ● No ☑

Hr_pm
100% ● Yes ○ No ☑

Date of Birth
100% 1973/01/31 ☑

2 | Report Of Adverse Events Following Immunization (AEFI)

● Initial report
○ Follow up report (Unique episode #)

1a. Unique episode #: ATM0123455 1b. Region #: ON 2. IMPACT LIN: 1234

3. Patient Identification

First name: Last name: Health number:

Address of usual residence:

Province/Territory: Postal code: Phone: ext #:

Information Source: First name: Last name: Relation to patient:

Contact info, if different:

4. Information at Time of Immunization and AEFI Onset

4a. At time of immunization: Province/Territory of immunization: ON

Date vaccine administered (YYYY/MM/DD): 2021/01/31 (hr: ○ am/ ● pm) Date of birth (YYYY/MM/DD): 1973/01/31 Age: 48

Sex: ○ Male ● Female ○ Other □ Pregnant at time of immunization: □ Gestation □ weeks □ days

□ Indigenous status: □ First Nations □ Metis □ Inuit

4b. Medical history (up to the time of AEFI onset) *(Check all that apply and provide details in section 10)*

□ Concomitant medication(s) ☑ Known medical conditions/allergies □ Acute illness/injury

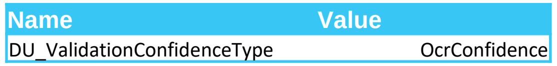
4c. Immunizing agent	Trade name	Manufacturer	Lot number	Dose #	Dosage/unit	Route	Site
COVID19 mDnA	CV19Phizer	Phizer	C55231	1	.25ml /	IM	RT ARM
					/		

Ui Action Center

Show OcrConfidence in validation action

You can select Confidence or OcrConfidence as the value to be displayed in the Action Center.
Depends on the documents set you deal with, chose the proper one.

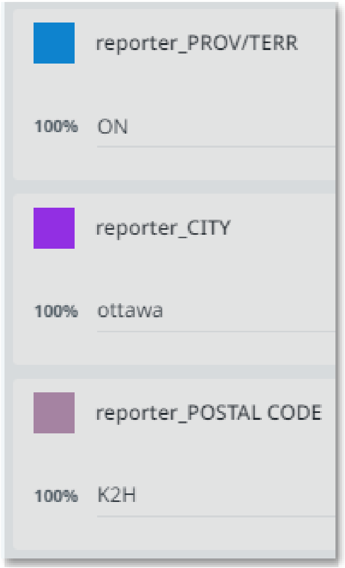
Confidence = Confidence level for location
OcrConfidence = Confidence level for OCR



Name	Value
DU_ValidationConfidenceType	OcrConfidence

Config.xlsx
(DuSettings)

Confidence

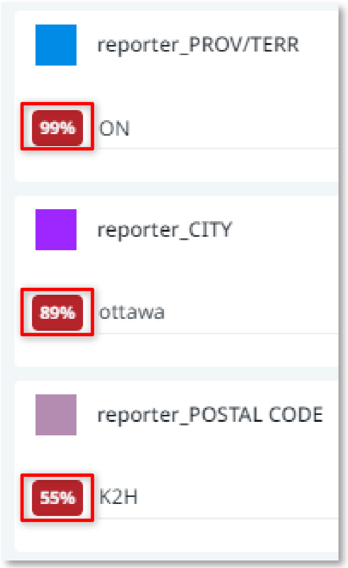


reporter_PROV/TERR
100% ON

reporter_CITY
100% ottawa

reporter_POSTAL CODE
100% K2H

OcrConfidence



reporter_PROV/TERR
99% ON

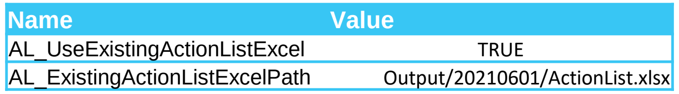
reporter_CITY
89% ottawa

reporter_POSTAL CODE
55% K2H

Ui Action Center

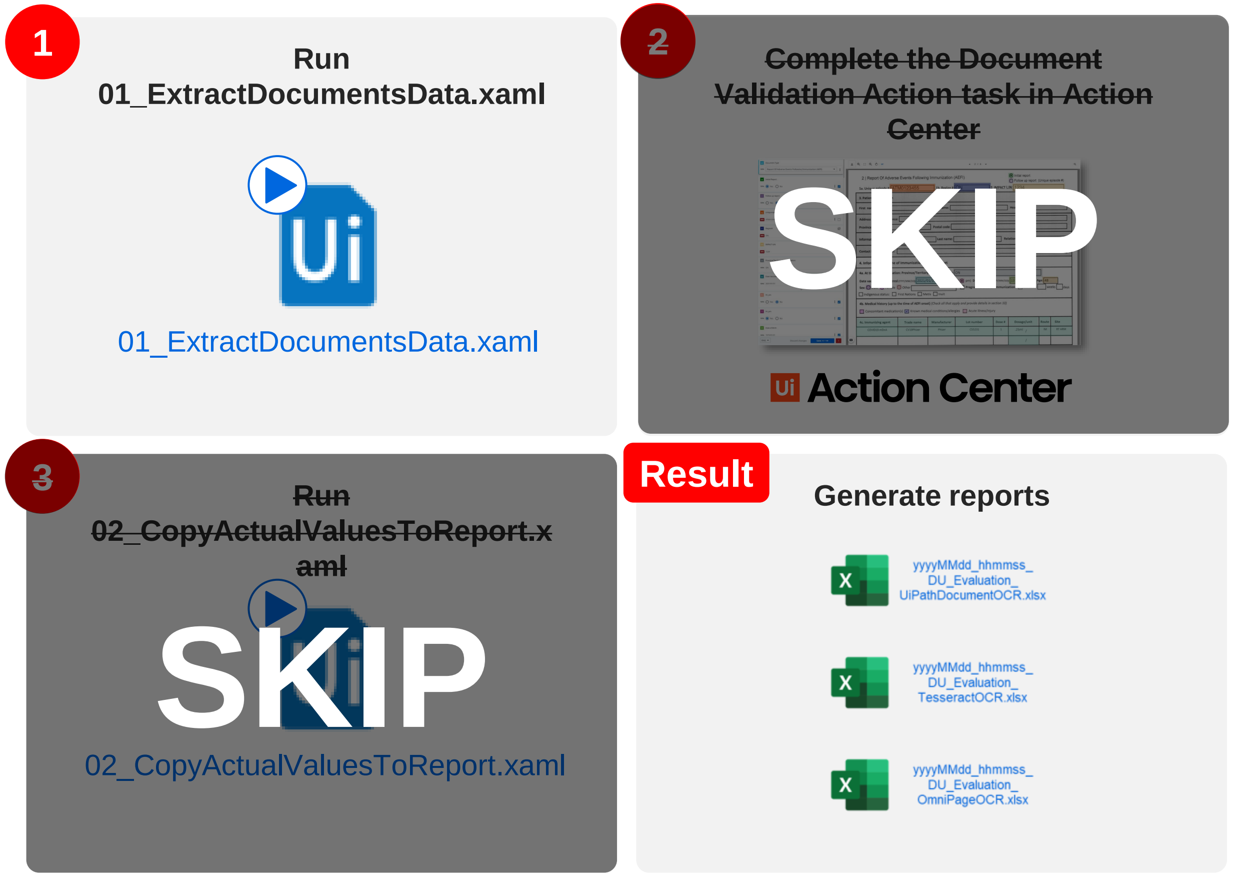
Use existing Action.xlsx as an actual value reference

If you have performed the "Execution Steps" and generated ActionList.xlsx for the same list of documents using the same taxonomy in the past, from next time, you can skip step validation in Action Center and execution of 02_CopyActualValuesToReport.xaml.



Name	Value
AL_UseExistingActionListExcel	TRUE
AL_ExistingActionListExcelPath	Output/20210601/ActionList.xlsx

Config.xlsx (BasicSettings)



1 Run 01_ExtractDocumentsData.xaml

01_ExtractDocumentsData.xaml

~~2~~ ~~Complete the Document Validation Action task in Action Center~~

SKIP

Action Center

~~3~~ ~~Run 02_CopyActualValuesToReport.xaml~~

SKIP

02_CopyActualValuesToReport.xaml

Result Generate reports

yyyyMMdd_hhmmss_DU_Evaluation_UiPathDocumentOCR.xlsx

yyyyMMdd_hhmmss_DU_Evaluation_TesseractOCR.xlsx

yyyyMMdd_hhmmss_DU_Evaluation_OmniPageOCR.xlsx

Document Understanding Evaluation Template

5

Specification & Known issues

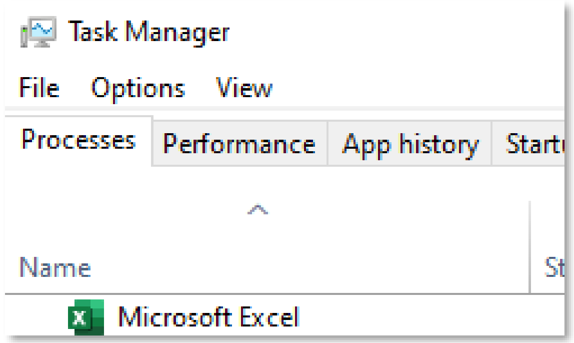
Specification& Known issues

Specification

- If there are multiple candidate extraction results in one field, the result with the highest Confidence and OcrConfidence will be listed in the reports.

Known issue

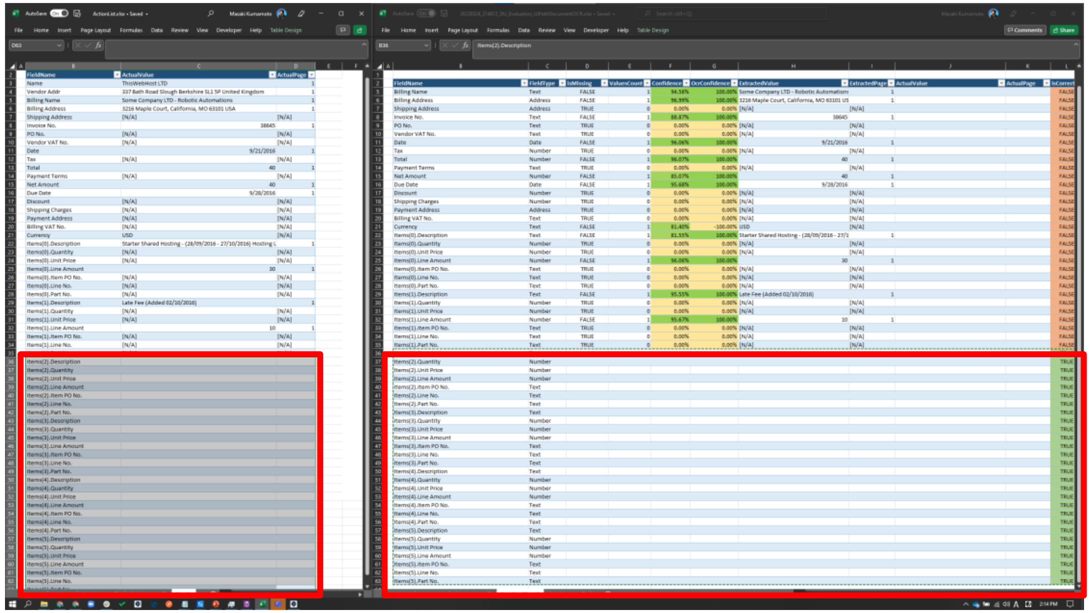
- If “Microsoft Excel” is running even in background when executing the workflow, it would cause unitability of the workflow somehow.



Specification& Known issues

Known issue

- If you use ML Extractor and there is Table fields in the taxonomy, it occurs error when creating reports because the workflow currently can't deal with dynamic number of field in case the numbers of rows got extracted is smaller than the actual numbers of rows in the document.
  You can work around the issue by adding rows with proper field name and empty values manually before running the workflow. Contact Masaki.Kumamoto@uipath.com if help is needed.


AutoSave On
ActionList.xlsx • Saved
Masaki Kumamoto
File Home Insert Page Layout Formulas Data Review View Developer Help Table Design
D63
20220324_214812_DU_Evaluation_UiPathDocumentOCR.xlsx • Saved
Search (Alt+Q)
Comments
Share
B36
Items(2).Description

FieldName	ActualValue	ActualPage
Name	ThisWebHost LTD	1
Vendor Addr	337 Bath Road Slough Berkshire SL1 5P United Kingdom	1
Billing Name	Some Company LTD - Robotic Automations	1
Billing Address	3216 Maple Court, California, MO 63101 USA	1
Shipping Address	[N/A]	[N/A]
Invoice No.	38645	1
PO No.	[N/A]	[N/A]
Vendor VAT No.	[N/A]	[N/A]
Date	9/21/2016	1
Tax	[N/A]	[N/A]
Total	40	1
Payment Terms	[N/A]	[N/A]
Net Amount	40	1
Due Date	9/28/2016	1
Discount	[N/A]	[N/A]
Shipping Charges	[N/A]	[N/A]
Payment Address	[N/A]	[N/A]
Billing VAT No.	[N/A]	[N/A]
Currency	USD	[N/A]
Items[0].Description	Starter Shared Hosting - (28/09/2016 - 27/10/2016) Hosting L	1
Items[0].Quantity	[N/A]	[N/A]
Items[0].Unit Price	[N/A]	[N/A]
Items[0].Line Amount	30	1
Items[0].Item PO No.	[N/A]	[N/A]
Items[0].Line No.	[N/A]	[N/A]
Items[0].Part No.	[N/A]	[N/A]
Items[1].Description	Late Fee (Added 02/10/2016)	1
Items[1].Quantity	[N/A]	[N/A]
Items[1].Unit Price	[N/A]	[N/A]
Items[1].Line Amount	10	1
Items[1].Item PO No.	[N/A]	[N/A]
Items[1].Line No.	[N/A]	[N/A]
Items[2].Description		
Items[2].Quantity		
Items[2].Unit Price		
Items[2].Line Amount		
Items[2].Item PO No.		
Items[2].Line No.		
Items[2].Part No.		
Items[3].Description		
Items[3].Quantity		
Items[3].Unit Price		
Items[3].Line Amount		
Items[3].Item PO No.		
Items[3].Line No.		
Items[3].Part No.		
Items[4].Description		
Items[4].Quantity		
Items[4].Unit Price		
Items[4].Line Amount		
Items[4].Item PO No.		
Items[4].Line No.		
Items[4].Part No.		
Items[5].Description		
Items[5].Quantity		
Items[5].Unit Price		
Items[5].Line Amount		
Items[5].Item PO No.		
Items[5].Line No.		

FieldName	FieldType	IsMissing	ValuesCount	Confidence	OcrConfidence	ExtractedValue	ExtractedPage	ActualValue	ActualPage	IsCorrect
Billing Name	Text	FALSE	1	94.58%	100.00%	Some Company LTD - Robotic Automation	1			FALSE
Billing Address	Address	FALSE	1	96.99%	100.00%	3216 Maple Court, California, MO 63101 US	1			FALSE
Shipping Address	Address	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Invoice No.	Text	FALSE	1	88.87%	100.00%	38645	1			FALSE
PO No.	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Vendor VAT No.	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Date	Date	FALSE	1	96.06%	100.00%	9/21/2016	1			FALSE
Tax	Number	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Total	Number	FALSE	1	96.07%	100.00%	40	1			FALSE
Payment Terms	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Net Amount	Number	FALSE	1	85.07%	100.00%	40	1			FALSE
Due Date	Date	FALSE	1	95.68%	100.00%	9/28/2016	1			FALSE
Discount	Number	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Shipping Charges	Number	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Payment Address	Address	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Billing VAT No.	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Currency	Text	FALSE	1	81.40%	-100.00%	USD	[N/A]			FALSE
Items(0).Description	Text	FALSE	1	81.55%	100.00%	Starter Shared Hosting - (28/09/2016 - 27/1	1			FALSE
Items(0).Quantity	Number	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(0).Unit Price	Number	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(0).Line Amount	Number	FALSE	1	96.06%	100.00%	30	1			FALSE
Items(0).Item PO No.	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(0).Line No.	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(0).Part No.	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(1).Description	Text	FALSE	1	95.55%	100.00%	Late Fee (Added 02/10/2016)	1			FALSE
Items(1).Quantity	Number	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(1).Unit Price	Number	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(1).Line Amount	Number	FALSE	1	95.67%	100.00%	10	1			FALSE
Items(1).Item PO No.	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(1).Line No.	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(1).Part No.	Text	TRUE	0	0.00%	0.00%	[N/A]	[N/A]			FALSE
Items(2).Quantity	Number									TRUE
Items(2).Unit Price	Number									TRUE
Items(2).Line Amount	Number									TRUE
Items(2).Item PO No.	Text									TRUE
Items(2).Line No.	Text									TRUE
Items(2).Part No.	Text									TRUE
Items(3).Description	Text									TRUE
Items(3).Quantity	Number									TRUE
Items(3).Unit Price	Number									TRUE
Items(3).Line Amount	Number									TRUE
Items(3).Item PO No.	Text									TRUE
Items(3).Line No.	Text									TRUE
Items(3).Part No.	Text									TRUE
Items(4).Description	Text									TRUE
Items(4).Quantity	Number									TRUE
Items(4).Unit Price	Number									TRUE
Items(4).Line Amount	Number									TRUE
Items(4).Item PO No.	Text									TRUE
Items(4).Line No.	Text									TRUE
Items(4).Part No.	Text									TRUE
Items(5).Description	Text									TRUE
Items(5).Quantity	Number									TRUE
Items(5).Unit Price	Number									TRUE
Items(5).Line Amount	Number									TRUE
Items(5).Item PO No.	Text									TRUE
Items(5).Line No.	Text									TRUE
Items(5).Part No.	Text									TRUE

Document Understanding Evaluation Template

6

Release note

UiPath Reboot™ Work.

Release note

You can find the packages => Packages, Releases => Releases

Date	Version	Description	Repository Link
Jun 3, 2021	0.0.1	• Initial project	Link
Jun 8, 2021	0.0.3	• Added default taxonomy for UiPath public endpoint's ML models (Invoices, Purchase Orders, Receipts, Utility Bills)	Link
Jun 8, 2021	0.0.4	• **Bug fix:** When an extracted value was 0, the "isCorrect" column in the Excel report would display Correct with empty value in "Actual" column.	Link
Jun 9, 2021	0.0.7	• Close disabled activity • Remove “Table” and “TableColumn” from Summary sheet in DU_Evaluation.xlsx since they are not used • Freeze 2nd row of the excel sheets so headers will be always on top • Added ID Cards and Passports taxonomy	Link
Jul 16, 2021	0.2.1	• Auto fill bug fix(from 0.0.7) • Fix bug: When there is "Pending Action", 02_Process occurs an error. • Fixed an issue where the overall result of the summary sheet would shift. (Issue from 0.0.7) • Close disabled activity • Remove “Table” and “TableColumn” from Summary sheet in DU_Evaluation.xlsx since they are not used • Freeze 2nd row of the excel sheets so headers will be always on top • Added ID Cards and Passports taxonomy • **Bug Fix:** behavior when AL_UseExistingActionListExcel is True. • Include ExtractedPage and ActualPage in the reports • Include timestamp as a strorage bucket folder name • Not to create a new Excel file when AL_UseExistingActionListExcel is True • **Bugfix:** The order of documents images in the excel report is not correct when it's equal or more than 10 pages	Link
Mar 24, 2022	0.2.3	• Added "Delay" activity for stability • Updated packages • **Bugfix:** Added "OC_FolderPath" valueable in "Orchestrator Folder" property of Get Task Data.	Link

Document Understanding Evaluation Template

Apendix

Road map

UiPath Reboot™ Work.

Roadmap

- “Process Template” to “Test Automation” to enable UiPath’s testing capability
- Straight Through Processing (STP) rate in the report
  - Ability to add business logic verification in the template
  - Business logic verification result for each field level
- Add overview reasons why it failed to extract the right text from documents
  - Not being able to Identify the right one
  - Identified as minor options
  - Partially wrong matching
  - Missing some
- Auto-rename file instead of throw exception on file name over 31 chars (excel sheet name limit)
- Add Classification & Extraction processing time for in the report
- Training function for ML Classifier / Extractor
- Progress dialog (Status bar)
- More visualized report (HTML) <= This would make the workflow more stable.