

# Document Understanding Evaluation Template (Ver 0.2.3)



**Masaki Kumamoto**  
Sales Engineer  
UiPath Canada



# 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

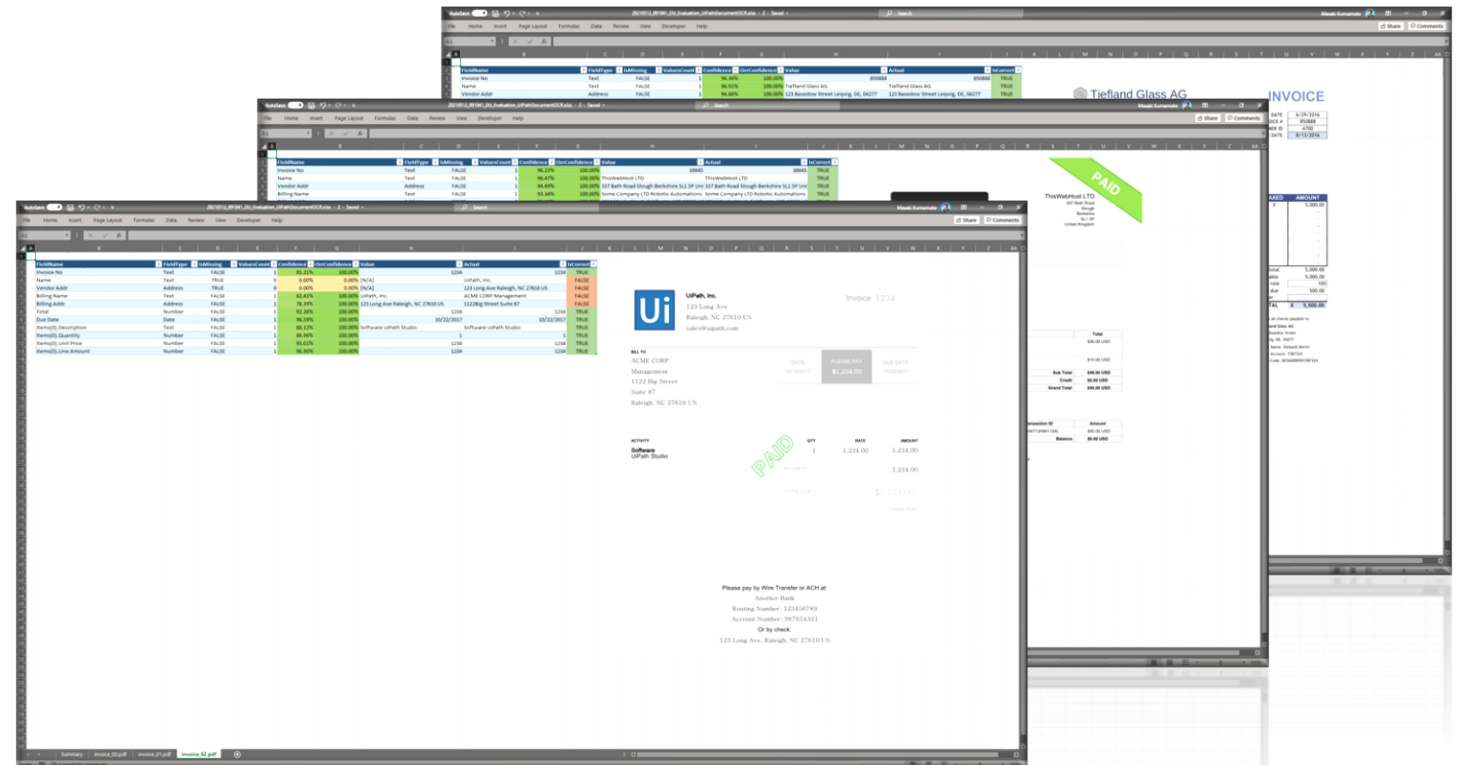
1

## 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%



# Features

## Minimum Effort Development

You only have to edit 3 files!



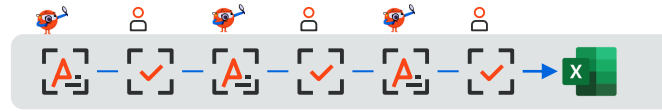
You can just focus on...

- Taxonomy definition
- Classification/Extraction logic

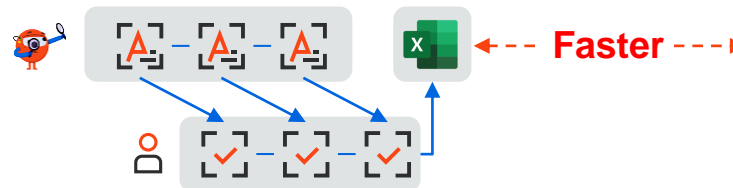
You do not even need to add any additional activities, valuables and arguments!

## Faster Validation Flow

Without the template

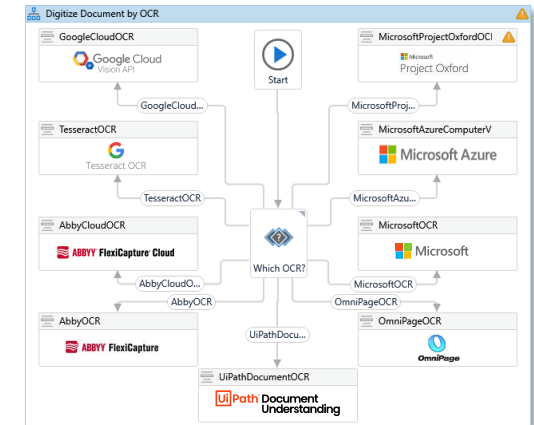


With the template



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

**Ui Action Center**

Demonstration for building Document Validation Actions in the Action Center

# Document Understanding Evaluation Template

2

## Details of the reporting files



# 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%

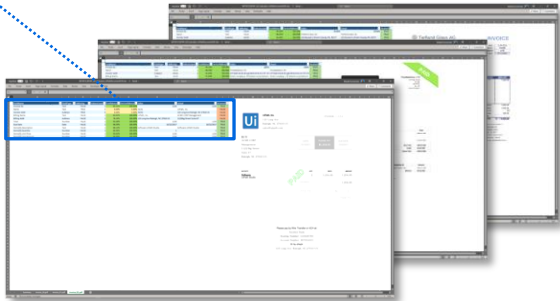
This sheet renders the percentage of correct extractions for all target documents extracted by DU.

## 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
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

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

# 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



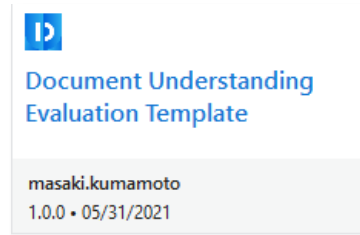
## Development



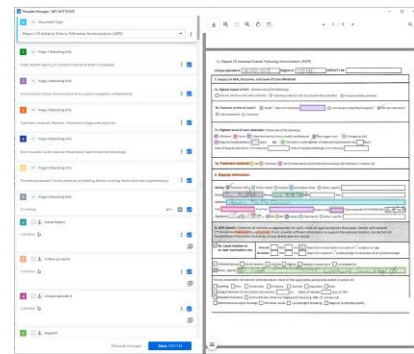
## Execution



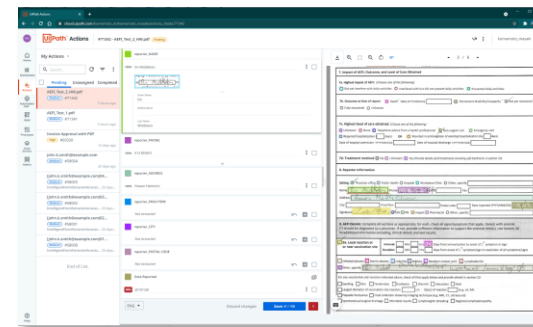
## Improvement



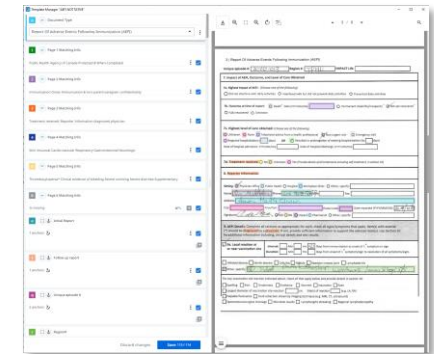
1. Place template project nuget package in the template folder
2. Create a new project from the template



1. Place target documents in Input folder
2. Configure [Config.xlsx](#)
3. Define Taxonomy
4. Build [DU\\_GetExtractionResult.xaml](#)



1. Run [01\\_ExtractDocumentsData.xaml](#)
2. Complete the Document Validation Action task in Action Center
3. Run [02\\_CopyActualValuesToReport.xaml](#)

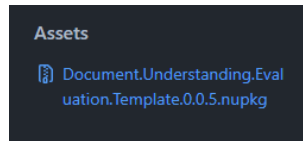


1. Use existing [ActionList.xlsx](#) for further development to improve the DU logic

# Preparation Steps

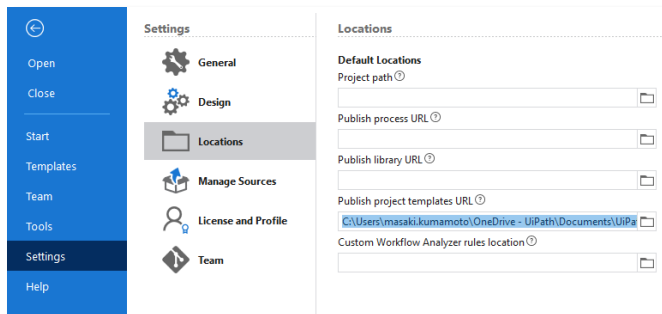
## 1. Place template project nuget package in the template folder

- Download [Document.Understanding.Evaluation.Template.X.X.X.nupkg](#) (Download from Assets)



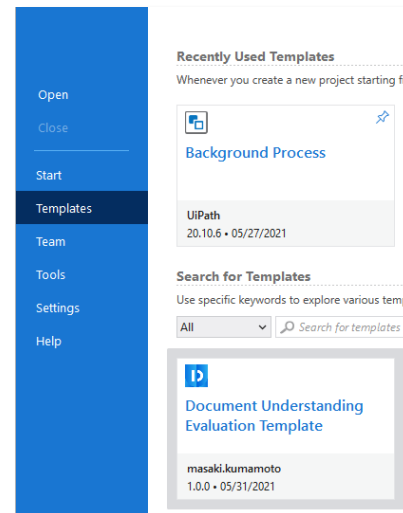
- You can identify the template folder below.  
UiPath Studio > Settings > Locations > Publish project templates URL

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



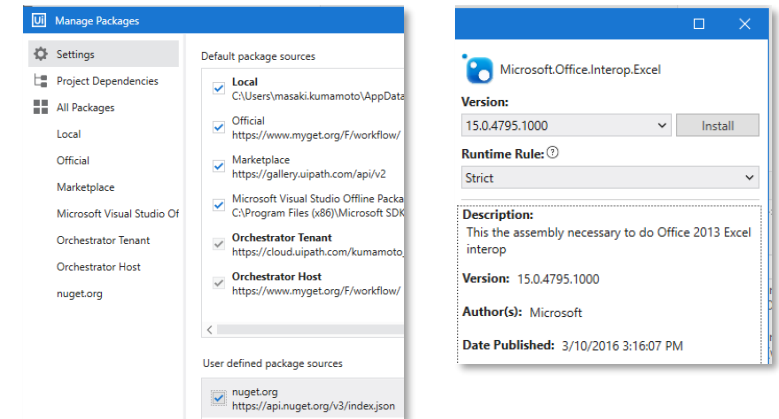
## 2. Create a new project from the template

- UiPath Studio > Templates > Document Understanding Evaluation Template



## 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.



# Development Steps

## 1. Place target documents in **Input** folder

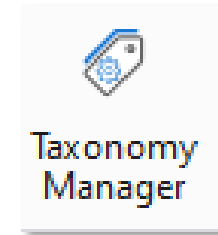
- If you use Form Extractor/Intelligent Form Extractor, you can also place template documents in "TemplateDocument" Folder).
- The file name placed in the Input Folder should not exceed 31 characters including the file extension. If the number of characters exceeds 31, an error will occur when creating 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 taxonomy in step 3.)
- **"ActionSettings"** Sheet
  - AC\_AssignUserEmail
  - OC\_FolderPath (Orchestrator Folder name that your Studio/Robot is deployed in)
  - SB\_BucketName (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.

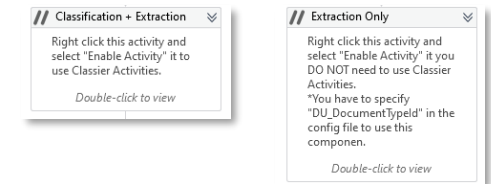
## 3. Define Taxonomy

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

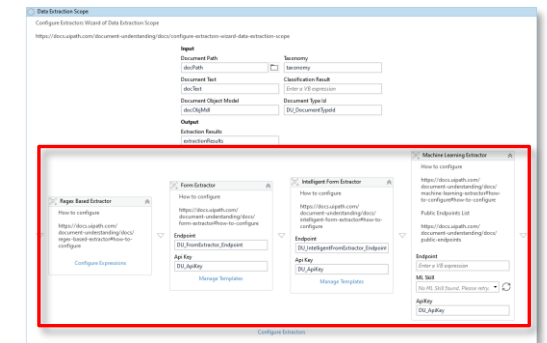


## 4. Build **DU\_GetExtractionResult.xaml**

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



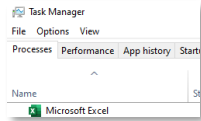
- Delete Classification/Extraction Activities which you do not use



# 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.

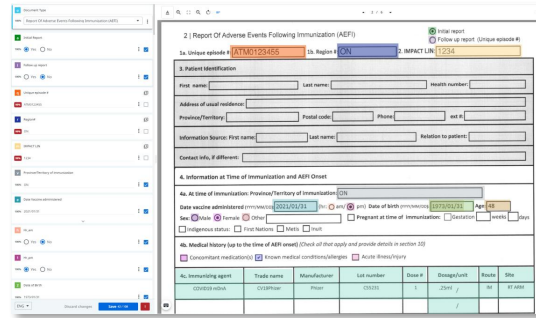


- It takes about 1-2 minutes to process each document. (It would take more with “Debug” so “Run” is recommended)
- After the execution is complete, Excel reports for each OCR set in Config.xlsx and the Document Validation Action in Action Center will be generated.



[01\\_ExtractDocumentsData.xaml](#)

## 2. Complete the Document Validation Action task in Action Center



The screenshot shows the UiPath Action Center interface. On the left is a sidebar with various action categories. The main area displays a form for '2) Report Of Adverse Events Following Immunization (AEFI)'. The form includes sections for '3. Patient identification' (with fields for first name, last name, health number, address, and phone), '4. Information at Time of Immunization and AEFI Onset' (with fields for date, time, location, and immunization details), and '5. Medical history up to the time of AEFI event'. At the bottom, there is a table for '5. Immunization history' with columns for trade name, manufacturer, lot number, dose, and date.

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.

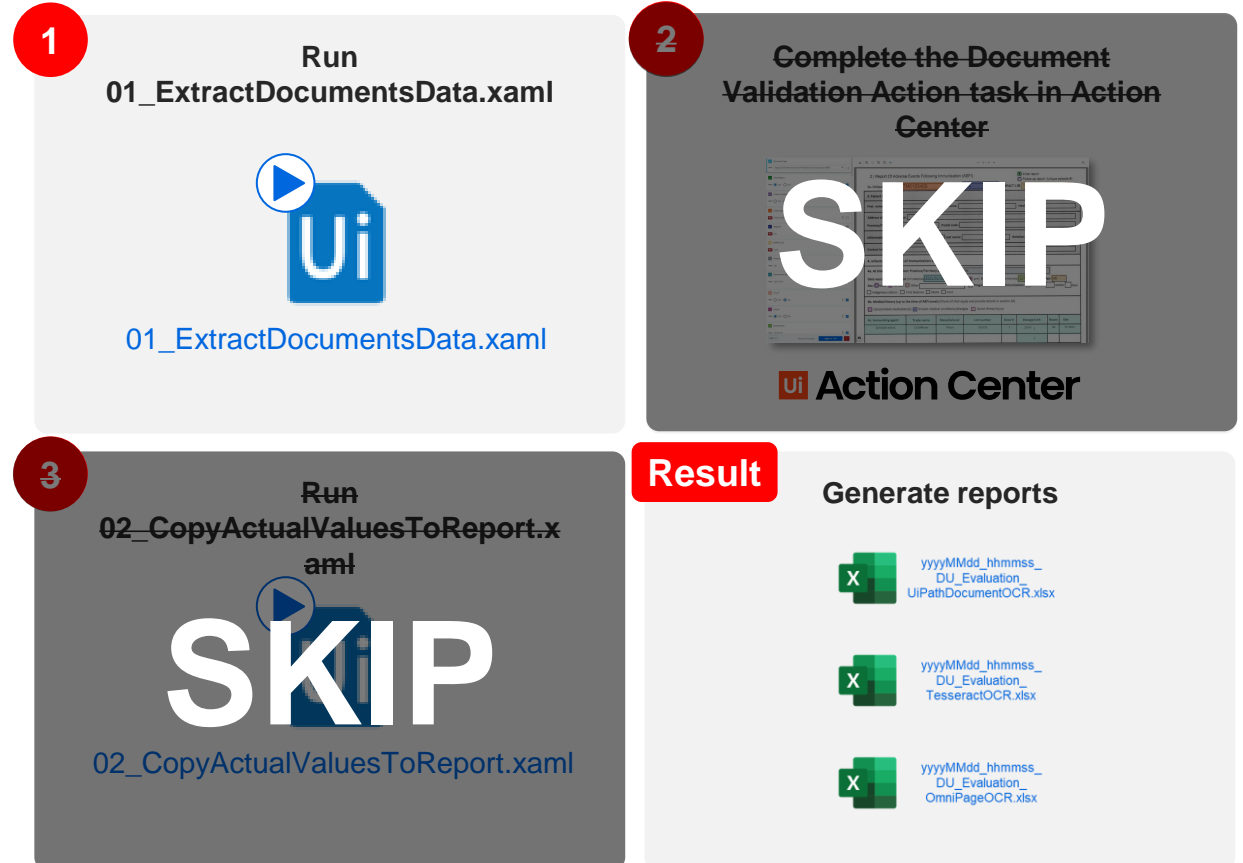
- **Configure [Config.xlsx](#)**

- "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)





# Document Understanding Evaluation Template

4

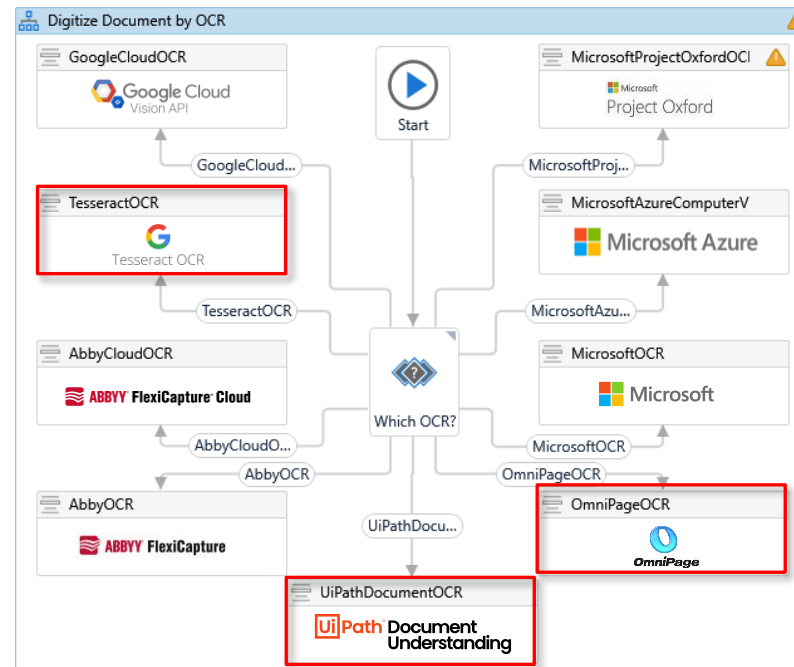
## Other useful features

# 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)



4c. Immunizing agent	Trade name	Manufacturer	Lot number	Dose #	Dosage/unit	Route	Site
COVID19 mRNA	CV19Pfizer	Pfizer	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)



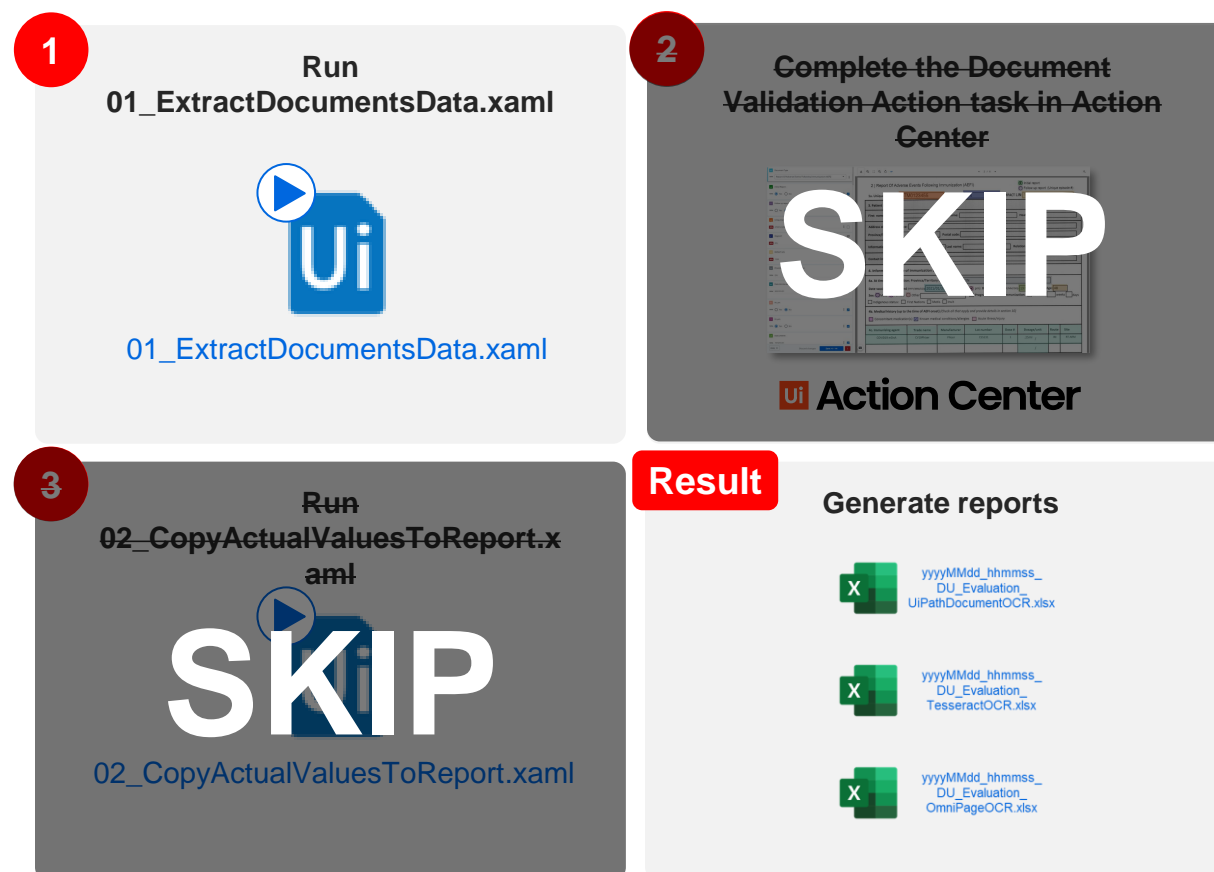
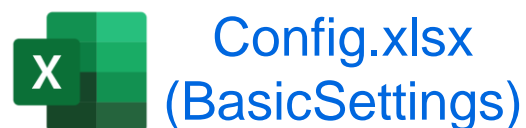
Confidence	OcrConfidence
<div>reporter_PROV/TERR</div> <div>100% ON</div>	<div>reporter_PROV/TERR</div> <div>99% ON</div>
<div>reporter_CITY</div> <div>100% ottawa</div>	<div>reporter_CITY</div> <div>89% ottawa</div>
<div>reporter_POSTAL CODE</div> <div>100% K2H</div>	<div>reporter_POSTAL CODE</div> <div>55% K2H</div>

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



# Document Understanding Evaluation Template

5

## Specification & Known issues

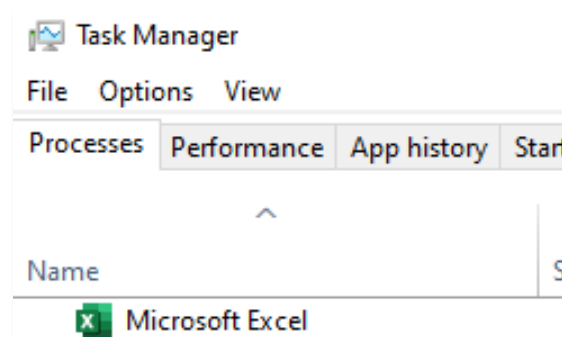
# Specification & Known issues

## 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 unavailability 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](mailto:Masaki.Kumamoto@uipath.com) if help is needed.

The image displays two side-by-side screenshots of the UiPath ML Extractor interface. Both screenshots show a table of extracted data with columns for field name, data type, and extracted value. The right screenshot includes additional columns for confidence scores. A red box in both screenshots highlights the bottom section of the table, which contains multiple rows of data, illustrating the issue where the number of rows extracted is smaller than the actual numbers of rows in the document.



# Document Understanding Evaluation Template

6

Release note

# Release note (1/2)

You can find the packages => [Packages](#)

You can find the release => [Releases](#)

Date	Version	Description	Repository Link
Jun 3, 2021	0.0.1	<ul style="list-style-type: none"><li>Initial project</li></ul>	<a href="#">Link</a>
Jun 8, 2021	0.0.3	<ul style="list-style-type: none"><li>Added default taxonomy for UiPath public endpoint's ML models (Invoices, Purchase Orders, Receipts, Utility Bills)</li></ul>	<a href="#">Link</a>
Jun 8, 2021	0.0.4	<ul style="list-style-type: none"><li>Bug fix. When an extracted value was 0, the "isCorrect" column in the Excel report would display Correct with empty value in "Actual" column.</li></ul>	<a href="#">Link</a>
Jun 9, 2021	0.0.7	<ul style="list-style-type: none"><li>Close disabled activity</li><li>Remove "Table" and "TableColumn" from Summary sheet in DU_Evaluation.xlsx since they are not used</li><li>Freeze 2nd row of the excel sheets so headers will be always on top</li><li>Added ID Cards and Passports taxonomy</li></ul>	<a href="#">Link</a>
Jul 16, 2021	0.2.1	<ul style="list-style-type: none"><li>Auto fill bug fix(from 0.0.7)</li><li>Fix bug: When there is "Pending Action", 02_Process occurs an error.</li><li>Fixed an issue where the overall result of the summary sheet would shift. (Issue from 0.0.7)</li><li>Close disabled activity</li><li>Remove "Table" and "TableColumn" from Summary sheet in DU_Evaluation.xlsx since they are not used</li><li>Freeze 2nd row of the excel sheets so headers will be always on top</li><li>Added ID Cards and Passports taxonomy</li><li>Bug Fixed: behavior when AL_UseExistingActionListExcel is True.</li><li>Include ExtractedPage and ActualPage in the reports</li><li>Include timestamp as a storage bucket folder name</li><li>Not to create a new Excel file when AL_UseExistingActionListExcel is True</li><li>Bugfix: The order of documents images in the excel report is not correct when it's equal or more than 10 pages</li></ul>	<a href="#">Link</a>

# Release note (2/2)

You can find the packages => [Packages](#)

You can find the release => [Releases](#)

Date	Version	Description	Repository Link
Mar 24, 2022	0.2.3	<ul style="list-style-type: none"><li>Added "Delay" activity for stability</li><li>Updated packages</li><li>Added "OC_FolderPath" valueable in "Orchestrator Folder" property of Get Task Data.</li></ul>	<a href="#">Link</a>

# Document Understanding Evaluation Template

## Appendix

### Road map

# 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.