



TARPON
HEALTH

a healthcare automation community

OCR and UiPath Document Understanding



What letter is this and how do you know?

R

R

R

R

R

R

R

R

R

R

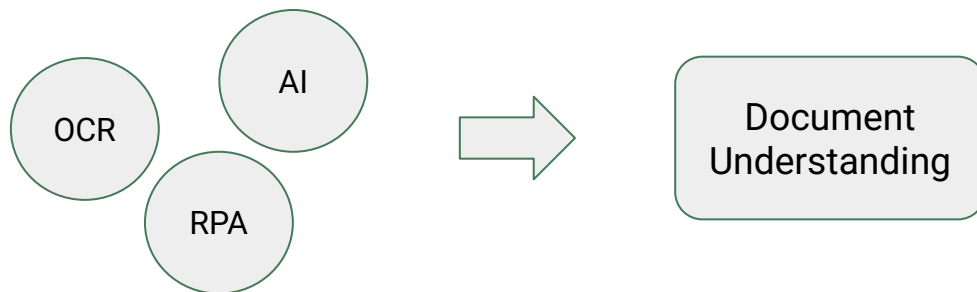
R

R

R

Definitions

1. OCR - Optical Character Recognition
 - a. Printed characters -> ASCII characters (machine-readable)
2. Pattern Recognition (*what letter is this?*)
 - a. Think of one font only - this type of pattern is this letter
3. Feature Detection (*how do you know?*)
 - a. / \ -> is this an A or a W?
4. UiPath Document Understanding
 - a. Intelligent document processing within automations
 - b. This framework takes RPA and AI to interpret all kinds of documents (handwriting, images, etc.)

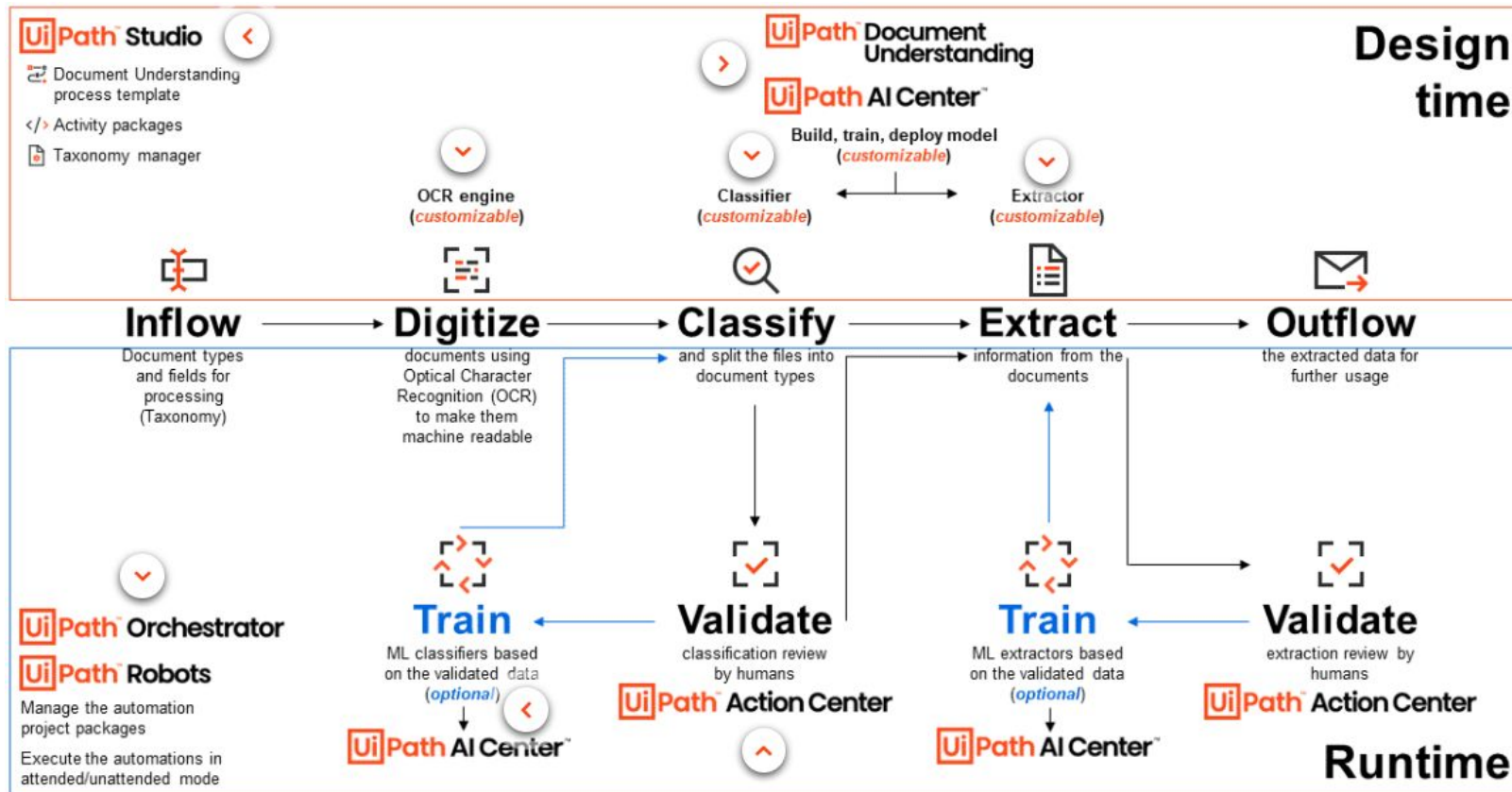


Potential Use Cases

1. Medical Forms
2. Medical Bills
3. Health Records
4. Claim Processing
5. **Remittance Advice**
6. Provider Bill Processing
7. **Order Intake**
8. Medical Records

NOTE: The conceivability of each use case is dependent on the document types, layouts, etc. If the process needs to read a wide range of documents from various facilities, it may not be the best candidate for automation. No fixed document format means more training, which means more development, which means more money.

Document Understanding Standard Workflow



Digitization (Document Understanding Step 1)

Digitization is the process of obtaining machine-readable text from a given file, allowing a robot to understand its contents and act accordingly.

Digitization is done with an OCR engine.

UiPath OCR Usage

Document Understanding	RPA
• Importing documents for classification/extraction • Engines available for this step are UiPath Document OCR, Google Cloud Vision OCR, and Microsoft Read OCR.	• At run time when calling models from RPA workflows • All engines listed on left plus Abbyy Finereader, Microsoft OCR (legacy), Microsoft Project Oxford OCR, and Tesseract. • Activities like Click OCR Text, Get OCR Text, and Find OCR Text Position

NOTE: If you're getting poor results from a single model, it's a good idea to try other models to see if they yield better results

Classification (Document Understanding Step 2)

What type of document is this?

Structured

1. Focuses on collecting information in a precise format
2. Can contain handwritten text, signatures, and checkboxes
3. Examples: forms, passports, licenses

Semi-structured

1. Repetitive information each time
2. Examples: invoices, receipts, purchase orders, medical bills

Unstructured

1. No fixed format
2. Examples: contracts, agreements, leases, emails

Form	1040-X	Department of the Treasury—Internal Revenue Service Amended U.S. Individual Income Tax Return ▶ Go to www.irs.gov/Form1040X for instructions and the latest information.	OMB No. 1545-0074
(Rev. January 2020)			
This return is for calendar year ☐ 2019 ☐ 2018 ☐ 2017 ☐ 2016 Other year. Enter one: calendar year or fiscal year (month and year ended):			
Your first name and middle initial		Last name	Your social security number
Dean E		Vargas	5 7 4 4 2 1 1 7 9
If joint return, spouse's first name and middle initial		Last name	Spouse's social security number
Mary		Vargas	6 7 4 1 6 9 3 4 5
Current home address (number and street). If you have a P.O. box, see instructions.			Apt. no.
5016 Musgrave Street			Your phone number
City, town or post office, state, and ZIP code. If you have a foreign address, also complete spaces below. See instructions.			404-226-7298
Foreign country name			Foreign postal code
Foreign province/state/county			
Amended return filing status. You must check one box even if you are not changing your filing status. Caution: In general, you can't change your filing status from a joint return to separate returns after the due date.			
☐ Full-year health care coverage (or, for amended 2018 returns only, exempt). If amending a 2019 return, leave blank. See instructions.			
☐ Single ☒ Married filing jointly ☐ Married filing separately (MFS) ☐ Qualifying widow(er) (QW) ☐ Head of household (HOH)			
If you checked the MFS box, enter the name of spouse. If you checked the HOH or QW box, enter the child's name if the qualifying person is a child but not your dependent. ▶			
Use Part III on the back to explain any changes			
Income and Deductions		A. Original amount reported or as previously adjusted (see instructions)	B. Net change—amount of increase or (decrease)—explain in Part III
C. Correct amount			
1 Adjusted gross income. If a net operating loss (NOL) carryback is			

Sample Merchandise License Agreement – This sample agreement should not be used without prior consultation with an experienced intellectual property attorney

THIS AGREEMENT is made this _____, by and between _____ of _____, (the "Licensor"), and _____ of _____, (the "Licensee").

WHEREAS, the Licensor is the owner of certain rights in the property described and/or illustrated in Schedule A (the "Property") including, but not limited to, the Trademark(s) listed in Schedule B (the "Trademarks");

WHEREAS, the Licensee desires to use the Property and/or the Trademarks on or in connection with the products identified in Schedule C (the "Licensed Products") and in the countries identified in Schedule D (the "Licensed Territory"); and

WHEREAS, the Licensor is willing to grant the Licensee the right to use the Property and/or the Trademarks on such Licensed Products.

Now, THEREFORE, in consideration of the mutual promises, covenants and conditions contained herein, it is hereby agreed as follows.

1. Grant of License

The Licensor is the sole and exclusive owner of all rights in and to the Property and/or the Trademarks and, with the exception of the rights being licensed hereunder, all other rights relating thereto are expressly reserved by the Licensor.

The Licensor hereby grants to the Licensee an exclusive, nontransferable, non-assignable license, without the right to grant sublicenses, to use the Property and/or the Trademarks solely on or in connection with the manufacture, sale, offering for sale, advertising, promotion and distribution of the Licensed Products and solely within the Licensed Territory and, for this purpose only, to affix the Trademarks on or to packaging, advertising and promotional materials sold, used or distributed in connection with the Licensed Products (the "Promotional and Packaging Material").

2. Term and Options

(a) This Agreement shall begin and be effective upon execution of this Agreement by both parties (the "Effective Date") and shall run for a "First Term" of _____ months thereafter, unless terminated sooner pursuant to a provision of this Agreement.

(b) If the Licensee is in full compliance with all terms and conditions hereof during the then in effect Term and the Licensee's sales during the then in-effect Term will have resulted in actual royalty payments to the Licensor of at least One Million United States Dollars (\$1,000,000), the Licensor agrees to grant to the Licensee two separate options (the "Option") to extend the then in

Template License Agreement

[APPROPRIATE RECITALS]

1. Definitions.

1.1 Licensed Patents means University's U.S Patent Application[s], Serial Number _____, filed _____, [ADDITIONAL APPLICATIONS] ("Applications") and (a) all divisionals, continuations and continuation-in-part (excluding new matter and claims containing new matter) applications that claim priority to any of the Applications; (b) any foreign application corresponding to any Application or application identified in (a); and (c) each patent that issues or reissues from any Application or application identified in (a) and (b) including [fill in any issued patents].

1.2 Technology means any information: (a) created prior to the Effective Date; (b) owned by [University]; (c) not infringing any pending or issued claim in the Licensed Patents; and (d) necessary to practice inventions claimed in patents and/or patent applications included in the Licensed Patents.

1.3 Licensed Product means any product that if manufactured, used, sold, offered for sale, or imported would constitute, but for the license granted Licensee under this Agreement, direct infringement, induced infringement and/or contributory infringement of any pending or issued claim of the Licensed Patents.

1.4 Field of Use means _____.

1.5 Territory means [geographic scope of the license].

1.6 Net Sales means the gross amount of all sales or leases of Licensed Products by Licensee, Affiliates, or sublicensees to any customer LESS:

- (a) discounts and rebates actually granted
- (b) excise and sales taxes imposed on the License Products and actually paid;
- (c) import, export and custom duties imposed on the License Products and actually paid; and
- (d) freight, transit and insurance charges actually paid.

1.7 [ADDITIONAL TERMS]

2. License Grant.

2.1 License Grant. [University] hereby grants to Licensee a [exclusive][nonexclusive] license under the Licensed Patents to make, have made for it, use and sell Licensed Products [in

5 Rue Bader
Narbonne, Aude, 11100
Phone: +33 140 260294

TO:
Roger Bigot
bonbono
4 Rue des Cites
Aubervilliers, Seine-Saint-Denis, 93300
Phone: +33 148 340990

COMMENTS OR SPECIAL INSTRUCTIONS:
None

INVOICE

INVOICE # BPXINV-00550
DATE: 23.05.2021

SHIP TO:
Roger Bigot
bonbono
4 Rue des Cites
Aubervilliers, Seine-Saint-Denis, 93300
Phone: +33 148 340990

SALESPERSON	P.O. NUMBER	TERMS
Marianne de la Guillaume	BPXPO-00536	Due after 30 days

QUANTITY	DESCRIPTION	UNIT PRICE	TOTAL
10	Dextromethorphan polistirex BPXPN-00057	12.45	124.50
25	Venlafaxine Hydrochloride BPXPN-00012	16.00	400.00
25	Metoclopramide Hydrochloride (BPXPO-00537) BPXPN-00002	9.99	249.75

Invoice 4235

Neid

Gillitzerstrasse 10, Rosenheim, BY, 83022



Date	To	Ship To
October 12, 2021	Obacht AG Heidelberger Strasse 86 Bensheim, HE, 64625	Same as recipient

Instructions
None

Make/Model	Product Code	Unit Price
Chrysler/300M Country of origin: France Color: Blue	762-88-0358	45400
Mercedes-Benz/SL-Class Country of origin: Italy Color: Red	630-35-4784	12100
Audi/90 Country of origin: Spain Color: Black	582-33-2563	18100
Ford/Bronco Country of origin: Spain Color: Blue	877-84-8130	30900
Audi/A6	638-13-7553	30800

INVOICE 0002478

Date: 19.03.2020

ConIncorporated

305 Fleet Rd
Fleet, Hampshire County,
GU51 3BU
012 5261 2116
admin@conincorp.co.uk

SHIP TO:

Brad McDonald
Asiatic Solutions
67 Manor Close
DOE LEA, S44 2YY
078 6814 5322
Customer ID: -

BILL TO:

Brad McDonald
Asiatic Solutions
67 Manor Close
DOE LEA, S44 2YY
078 6814 5322
Customer ID: -

ORDER DATE	ORDER NUMBER	DUE DATE
19.03.2020	PO-002444	19.04.2020

Item #	Ordered Service	Item Price	Total
1	8-200 - Wood and Plastic Doors (9)	19.45	175.05
2	3-230 - Anchor Bolts (30)	82.22	2466.6
3	1-000 - Purpose (13)	12.61	163.93
4	13-010 - Air-Supported Structures (13)	7.42	96.46
5	2-782 - Brick Pavers (7)	41.69	291.83

Extraction (Document Understanding Step 3)

Get all the requested fields from the document

1. Multiple types to choose from
 - a. Regex Based Extractor
 - b. Form Extractor
 - c. Machine Learning Extractor
 - d. Generative Extractor

Document Types and Fields	Regex Based Extractor		Form Extractor	
	Framework Alias	Minimum Confidence %	Framework Alias	Minimum Confidence %
▼ Government	☒	80	☐	50
Title	☒		☐	
Form Type	☒		☐	
OMB	☒		☐	
Part 1	☒		☐	
Part 2	☒		☐	
Part 3	☒		☐	

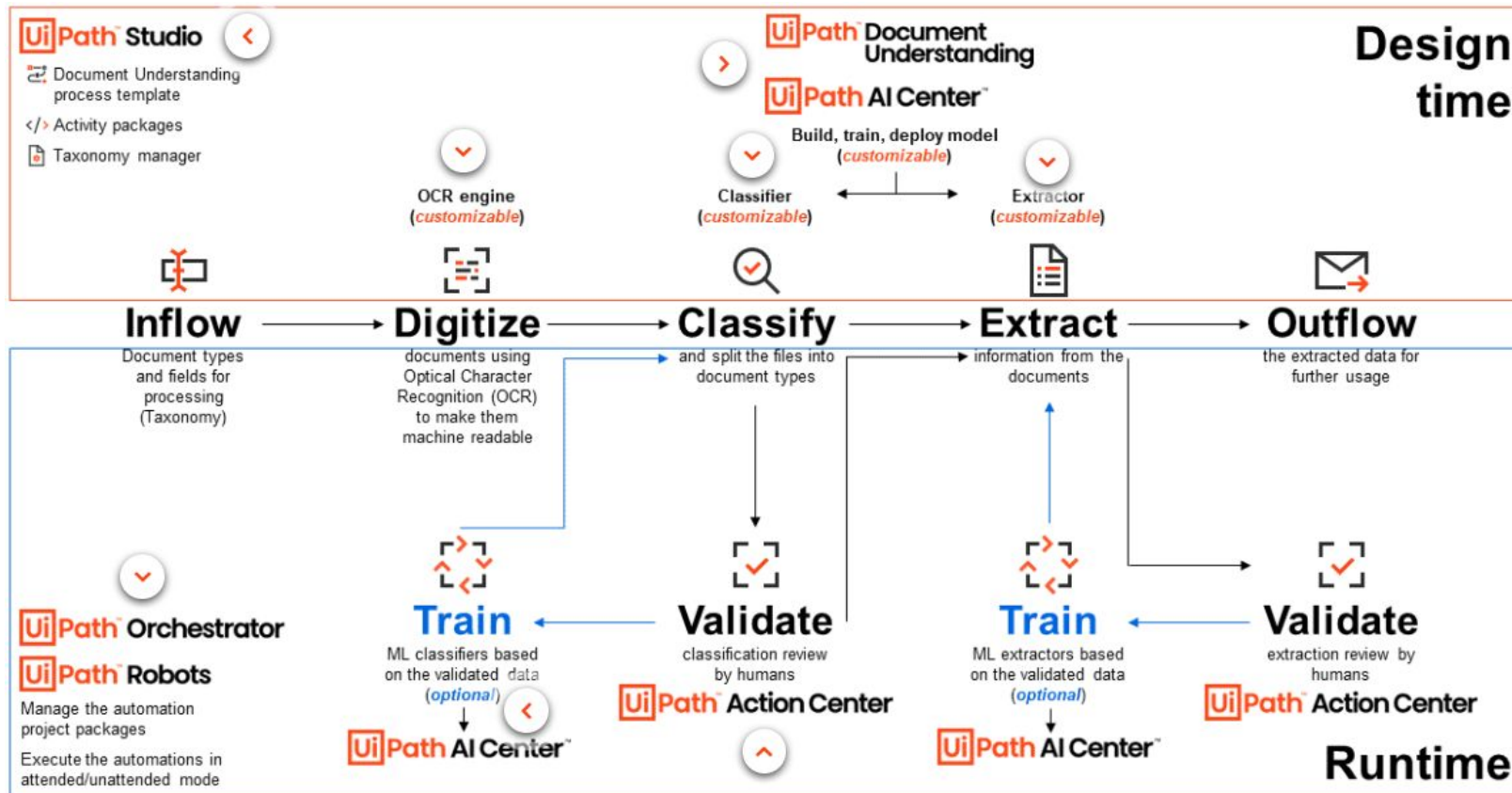
2. Configurable - select the most suitable extractor for each field
3. Use a combination of approaches to ensure smooth and accurate processing of different documents
4. You can even implement fall-back rules for data extraction. If a certain extractor does not report an acceptable value for a given field, then a backup extractor can be called.
 - a. Confidence threshold

Validation (after Classification and Extraction steps)

Human review

- Process of verifying and ensuring the accuracy of the extracted data from documents. Important because it allows users to confirm that the extracted information is correct before proceeding to the next steps in the automation process.
- Done by the end user/SME (Human in the loop) to help the automation get better at reading documents over time
- Essential part of any Document Understanding project
- Classification - did the automation classify the document as the correct type?
- Extraction - did the automation extract the correct data for each field?

Document Understanding Standard Workflow



UiPath Automation Cloud Demo



1. Create a new project
2. Choose Modern
3. Create a classification for your specific document types
4. Upload documents
5. Click Measure and begin annotating your documents
6. Once enough documents have been annotated, you will receive a project score and recommendations on how to improve

NOTE: In order to use this newly trained model in an API or RPA workflow, you will need to click Publish and create a new project version

Examples of how the DU Interface helps you train a model:



Upload more documents

Vehicle Damage Report has too few samples. Add at least 150 documents for an optimal data set size.

Vehicle Damage Report



Annotate more documents

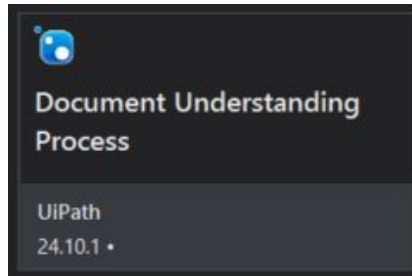
Annotate a minimum of 5 pages. Currently, there are not enough samples to begin model training.

UiPath Studio Demo

*DO NOT USE THIS TYPE OF WORKFLOW IN A PRODUCTION SETTING

1. Create a new process
2. Manage packages and add **UiPath.DocumentUnderstanding.ML.Activities** and **UiPath.IntelligentOCR.Activities**
3. Build Document Type in Taxonomy Manager
4. Digitize Document
5. Classify Document
6. Extract fields
7. Return Excel file

NOTE: If you decide to use DU within UiPath Studio for a production automation, I would highly recommend using the Document Understanding Process Template within UiPath Studio.



Questions?

Document Understanding in Automation Cloud:

[Active Learning in Modern Projects of UiPath Document Understanding Course | UiPath Academy](#)

Document Understanding in Studio:

[Build a Document Understanding Automation in Studio Course | UiPath Academy](#)

Publishing a DU project from Automation Cloud:

[Document Understanding - Publish](#)