OCR/MTO Desktop Flutter Application

Purpose:

The purpose of this application is to use OCR to automatically extract important information from a PDF that contains a diagram and text.

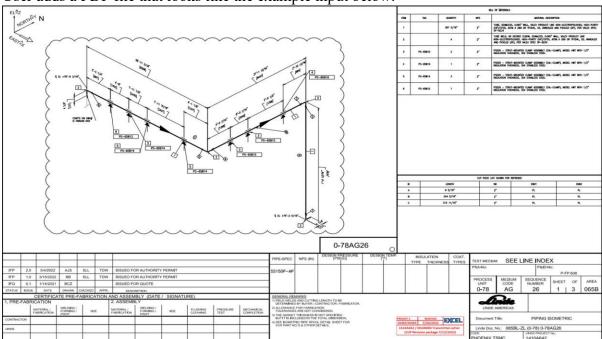
Input/Output

The input is a PDF image that contains a drawing, and different text fields in boxes.

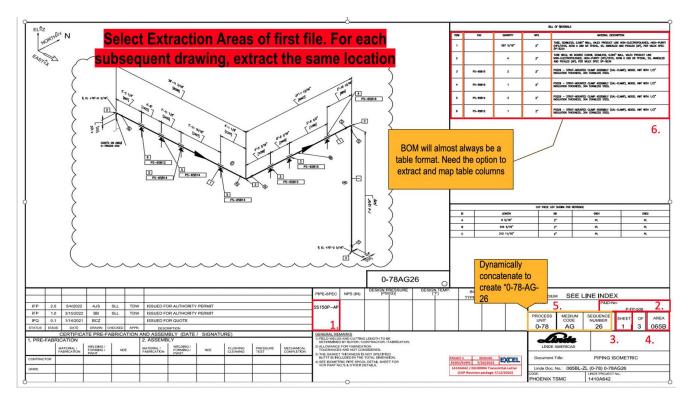
The output of this is a CSV file that contains rows and columns with the correct information, that is extracted from the PDF input.

Application Flow

1. User adds a PDF file that looks like the example input below.



2. A user can create box shapes on top of the input image to perform OCR to extract data. Extraction areas can change depending on the input PDF, which is why the app supports box creation in any place and any size. In the lower right corner with the orange boxes, OCR supports concatenation of the fields to generate an ID. In the example below, there are 6 boxes created in order to extract information.



3. There exists a basic Material Take-Off (MTO) Basic Table that contains the schema for all the column names possible. This schema will be provided in the format of a CSV or a JSON file. It contains column names similar to those shown below.

Material Take-Off Basic Table

This will be our basic table template

	Drav	ving Inform	ation		BOM (Bill of Materials)					
Drawing #	Sheet	Sheet Pipe P&ID Spec		Area	Item	Tag	Quantity NPS		Material Description	
							1			

4. When each box is created to extract data, a popup will appear to ask the user to select the column name from one that exists in the MTO Table. All the possible names are listed in the drop-down menu.

Markup #	Map To Column					
1	Pipe Spec					
2	Drawing # Sheet					
3	Pipe Spec P&ID					
4	Area Item Tag Quantity NPS Material Description					
5						
6						
6	Tag					
6	Quantity					
6	NPS					
6	Material Description					

5. After all the 6 boxes are done performing OCR, the MTO table will be populated with data. Currently it prints out this table. We would like to store it as a CSV file.

Material Take-Off Expected Output

	Dı	awing Informat	tion		BOM (Bill of Materials)					
Drawing #	Sheet	Pipe Spec	P&ID	Area	Item	Tag	Quantity	NPS	Material Description	
0-78-AG-26	1	SS150P-AP	P-FP-508	065B	1		567 5/16"	2"	TUBE, SEAMLESS, 0.065" WALL, VALEX PRODUCT LINE NON-ELECTROPOLISHED, HIGH-PURITY (HP)/CFOS, ASTM A 269 GR TP304L, SS, ANNEALED AND PICKLED (AP), PER VALEX SPEC SP-9234	
0-78-AG-26	1	SS150P-AP	P-FP-508	065B	2		4	2"	TUBE WELD, 90 DEGREE ELBOW, SEAMLESS, 0.065" WALL, VALEX PRODUCT LINE NON-ELECTROPOLISHED, HIGH-PURITY (HP)/CFOS, ASTM A 269 GR TP304L, SS, ANNEALED AND PICKLED (AP), PER VALEX SPEC SP-9234	
0-78-AG-26	1	SS150P-AP	P-FP-508	065B	3	PS-65B15	2	2"	PSD29 - STRUT-MOUNTED CLAMP ASSEMBLY (CAL-CLAMP), MODEL HWT WITH 1/2" INSULATION THICKNESS, 304 STAINLESS STEEL	
0-78-AG-26	1	SS150P-AP	P-FP-508	065B	4	PS-65B16	1	2"	PSD29 - STRUT-MOUNTED CLAMP ASSEMBLY (CAL-CLAMP), MODEL HWT WITH 1/2" INSULATION THICKNESS, 304 STAINLESS STEEL	
0-78-AG-26	1	SS150P-AP	P-FP-508	065B	5	PS-65B14	3	2"	PSD29 - STRUT-MOUNTED CLAMP ASSEMBLY (CAL-CLAMP), MODEL HWT WITH 1/2" INSULATION THICKNESS, 304 STAINLESS STEEL	
0-78-AG-26	1	SS150P-AP	P-FP-508	065B	6	PS-65B12	1	2"	PSD29 - STRUT-MOUNTED CLAMP ASSEMBLY (CAL-CLAMP), MODEL HWT WITH 1/2" INSULATION THICKNESS, 304 STAINLESS STEEL	

6. Group the data based on these columns, and save as a new CSV.

Material Take-Off Expected Output

- Now group the data and combine quantities (Only for BOM Data)
 - Group by 'Material Description' and Size(NPS)
 - Sum quantities

	Drawing Information					BOM (Bill of Materials)					
Drawing #	Sheet	Pipe Spec	P&ID	Area	Item	Tag	Quantity	NPS	Material Description		
0-78-AG-26	1	SS150P-AP	P-FP-508	065B	[Array]		567 5/16"	2"	TUBE, SEAMLESS, 0.065" WALL, VALEX PRODUCT LINE NON- ELECTROPOLISHED, HIGH-PURITY (HP)/CFOS, ASTM A 269 GR TP304L, SS, ANNEALED AND PICKLED (AP), PER VALEX SPEC SP-9234		
0-78-AG-26	1	SS150P-AP	P-FP-508	065B	[Array]		4	2"	TUBE WELD, 90 DEGREE ELBOW, SEAMLESS, 0.065" WALL, VALEX PRODUCT LINE NON-ELECTROPOLISHED, HIGH-PURITY (HP)/CFOS, ASTM A 269 GR TP304L, SS, ANNEALED AND PICKLED (AP), PER VALEX SPEC SP-9234		
8-AG-26	1	SS150P-AP	P-FP-508	065B	[Array]	[Array]	7		PSD29 - STRUT-MOUNTED CLAMP ASSEMBLY (CAL-CLAMP), MODEL HWT WITH 1/2" INSULATION THICKNESS, 304 STAINLESS STEEL		