

## Weekly Work Report

### 1. URL Testing - PaddleOCR Table Structure Recognition:

This week, I tested the table structure recognition feature of PaddleOCR using the URL provided

("https://www.paddleocr.ai/main/en/version3.x/module\_usage/table\_structure\_recognition.html"). The tool returns coordinates and an HTML representation of the detected table structure.

The screenshot shows the JSON Path Finder web application. The main area displays a JSON array of HTML table elements, each with its bounding box coordinates. The array is indexed from 4246 to 4274. The elements are HTML table rows and cells. The right sidebar shows the path `x.bbox[375]` and a list of coordinates for the selected path. Below the coordinates, the structure of the JSON is shown, indicating the path is an array of objects with a `structure` property.

Index	Value
0	14
1	1413
2	220
3	1417
4	213
5	1459
6	13
7	1456

structure:

Index	Value
0	<html>
1	<body>
2	<table>
3	<tbody>
4	<tr>

I attempted to map the HTML output, and here's a preview of the result:

[illegible]

Upon comparing the results with the data from page 28 of the document, I found discrepancies in the cell count. The number of cells detected by PaddleOCR did not match the expected count, highlighting a potential issue with its table structure recognition. This misalignment suggests that the tool is either incorrectly identifying or misrepresenting the table structure in certain cases.

## 2. Layout Detection - PaddleOCR:

I also explored the layout detection feature of PaddleOCR, which helps identify the general structure of a document. However, due to the varying table sizes in the images, we found that dynamic allocation based on layout detection was not feasible. The tool could not reliably handle the varying dimensions of the tables, which made it unsuitable for our current use case

Figure title 0.54  
Group 00—ANIMALS AND ANIMAL PRODUCTS, EDIBLE

Commodity & Quantity	Commodity Description and Economic Class	Unit or Quantity	Rate or Duty	Tariff
			100 Tons Ad. (except as noted)	
			Trade Agreement	
<b>ANIMALS, EDIBLE (EXCEPT FOR BREEDING)</b>				
0010 000 •	Cattle.....	No *		
	Weighting less than 200 pounds each (calves).....	Lb. 25¢ lb.	1½ lb. Can., Mex., bound GATT.	
0010 700 •	Weighting 200 pounds and less than 700 pounds each.....	Lb. 25¢ lb.	1½ lb. Mex.	701
0010 800 •	Cows for dairy purposes.....	No *	1½ lb. Can., bound GATT.	
0010 900 •	Cattle, n. s. p. f.....	No *	1½ lb. Can., Mex., bound GATT.	
0012 000 •	Sheep and lambs.....	No 35 each	1½ lb. Can., bound GATT.	
0012 300 •	Goats.....	No 35 each	1½ lb. Can., bound GATT.	702
0013 000 •	Poultry, live.....	Lb. 2¢ lb.	4¢ lb. Can.	
0014 000 •	Turkeys.....	Lb. 4¢ lb.	2¢ lb. GATT.	711
0015 000 •	Chickens, ducks, geese, and guinea.....	Lb. 4¢ lb.	2¢ lb. GATT.	
0015 500 •	Baby chicks of poultry.....	No. 4¢ each	2¢ lb. GATT.	
<b>MEAT PRODUCTS</b>				
0018 000 •	Fresh, chilled, or frozen.....	Lb. 6¢ lb.	3¢ lb. Can.	
0019 000 •	Veal.....	Lb. 6¢ lb.	3¢ lb. GATT.	701
0020 100 •	Pork.....	Lb. 25¢ lb.	1½ lb. Can., bound GATT.	703
0020 200 •	Ham.....	Lb. 25¢ lb.	1½ lb. GATT.	
0021 000 •	Mutton.....	Lb. 25¢ lb.	1½ lb. GATT.	702
0022 000 •	Lamb.....	Lb. 25¢ lb.	1½ lb. GATT.	
0023 100 •	Beef meat.....	Lb. 25¢ lb.	1½ lb. GATT.	704
0023 200 •	Veal meat.....	Lb. 25¢ lb.	1½ lb. GATT.	
0023 300 •	Game (except birds), n. s. p. f. (specify by name).....	Lb. 25¢ lb.	1½ lb. GATT.	
0023 400 •	Edible offal: Livers, kidneys, tongues, hearts, sweetbreads, tripe, and brains.....	Lb. 25¢ lb.	1½ lb. Can.	706
0023 500 •	Prep. legs (C. D. 610-4/6/42).....	Lb. 10¢	75¢ min. GATT.	1558
0023 900 •	Meats, n. s. p. f.....	Lb. 6¢ lb.	4¢ lb. bound Cuba GATT.	706
0024 000 •	Birds, including poultry.....	Lb. 10¢ lb.	5¢ lb. Can.	
0024 100 •	Ducks and geese.....	Lb. 10¢ lb.	5¢ lb. Can.	
0024 200 •	Birds, n. s. p. f. (including game birds).....	Lb. 10¢ lb.	5¢ lb. GATT.	712
0025 100 •	Prepared or preserved.....	Lb. 10¢ lb.	5¢ lb. GATT.	
0025 200 •	Prep. parts of 0025 000.....	Lb. 10¢ lb.	5¢ lb. GATT.	
0025 300 •	Beef, including poultry, prepared or preserved, n. s. p. f. (formerly part of 0025 000).....	Lb. 10¢ lb.	5¢ lb. GATT.	
0025 400 •	Canned beef, including corned beef.....	Lb. 10¢ lb.	5¢ lb. GATT.	
0025 500 •	Beef and veal, pickled or cured.....	Lb. 10¢ lb.	5¢ lb. GATT.	

For explanation of superior notations against commodity numbers and GATT rates, see page 1000.

\* Includes beef imported into Puerto Rico (not for consumption in U. S. except by special permission).

Next page 0.55

Figure title 0.76  
Group 00—ANIMALS AND ANIMAL PRODUCTS, EDIBLE—Continued

Commodity & Quantity	Commodity Description and Economic Class	Unit or Quantity	Rate or Duty	Tariff
			100 Tons Ad. (except as noted)	
			Trade Agreement	
<b>MEAT PRODUCTS—Continued</b>				
<b>Prepared or preserved—Continued</b>				
0030 100 •	Pork.....	Lb. 25¢ lb.	1½ lb. Can., bound GATT.	
0030 200 •	Ham.....	Lb. 25¢ lb.	1½ lb. GATT.	
0030 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	703
0031 100 •	Other.....	Lb. 25¢ lb.	1½ lb. Can., bound GATT.	
0031 200 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0031 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0032 100 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0032 200 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0032 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0033 100 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0033 200 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0033 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0034 100 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0034 200 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0034 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0035 100 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0035 200 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0035 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0036 100 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0036 200 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0036 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0037 100 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0037 200 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0037 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0038 100 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0038 200 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0038 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0039 100 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0039 200 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	
0039 300 •	Other.....	Lb. 25¢ lb.	1½ lb. GATT.	

For explanation of superior notations against commodity numbers and GATT rates, see page 1000.

\* Includes beef imported into Puerto Rico (not for consumption in U. S. except by special permission).

Next page 0.77

### 3. Table Cell Detection - PaddleOCR:

I tested the Table Cell Detection feature in PaddleOCR, and while it did provide the cell count, the results were not usable for our purposes. The detection did not align with the expectations we had for accurately identifying individual cells in a table, making it difficult to rely on for further processing.



convert this information into a CSV file for further processing. This is something I plan to explore further.

5. I reached out to Suzie, a student of Kristy, to discuss possible improvements and strategies. Unfortunately, I have not received a response yet, but I will continue to follow up on this for further collaboration.

### **Next Steps:**

- As of now, we are still using the QWEN LLM for OCR tasks. However, Sparrow OCR has recently released a new feature that utilizes local Llama models. I plan to test this new feature in the coming week and provide updates based on its performance. This could potentially offer improvements in accuracy and processing speed.