

Tales Araujo Leonidas

ITAI 2372 - Artificial Intelligence Applications

Prof. Patricia McManus

October 31st, 2024

Extracting Form Data in Document Intelligence Studio: Lab Report

Introduction

Microsoft Azure Lab 4 of the AI 900 certification, "Extract form data in Document Intelligence Studio", focused on using Azure's Document Intelligence, formerly Azure Form Recognizer, to analyze and extract structured data from a receipt. Through this lab, I gained hands-on experience with a prebuilt model in Document Intelligence, which can capture key data points from unstructured documents and store them in an organized format. This report summarizes my experience and reflections on the lab.

Description of Experience

The lab started with setting up the Document Intelligence resource in Azure. After creating this resource, I downloaded a sample receipt for the fictitious Northwind Traders and uploaded it to Document Intelligence Studio. Once the image was uploaded, I clicked "Run analysis" after checking that the "Receipt" option was selected, which used the prebuilt model specifically trained to recognize receipt data.

The analysis returned specific fields such as the merchant's name, address, phone number, transaction date, and line items. A confidence score was provided in each field to indicate the likelihood of accurate recognition. The model identified the purchased items, subtotal, tax, and total amount. This lab illustrated how Document Intelligence builds upon Optical Character Recognition (OCR) by extracting raw text, interpreting it, and structuring it into meaningful data fields.

Personal Reflection

This lab provided a valuable introduction to how Azure Document Intelligence can automate document processing. Unlike traditional OCR, which extracts unstructured text, Document Intelligence structures the data, making it suitable for storage and further analysis. I was impressed by the efficiency of the tool, especially in recognizing various data fields on a receipt, which would otherwise require manual entry.



4

**Extract form data in Document Intelligence Studio** (Expected Duration 1 hours) [Details](#)

AI-900T00-A Microsoft Azure AI Fundamentals [Cloud Slice Provided], Learning Path 04 (CSS)

Required: Yes

Status: Complete

Started: Wednesday, October 30, 2024 7:24 PM (Central Standard Time)

Ended: Wednesday, October 30, 2024 7:48 PM (Central Standard Time)

[Launch](#)

9 of 10 launch attempts remaining

Conclusion

Lab 4 was a practical exercise in exploring Azure's Document Intelligence, demonstrating how AI can streamline data extraction from documents like receipts. The lab underscored the difference between OCR and Document Intelligence by showing how AI can read and organize text, significantly enhancing data usability, and highlighting the potential of AI in automating document-based tasks.

References

Microsoft. "AI-900T00-A Microsoft Azure AI Fundamentals: Extract form data in Document Intelligence Studio.", 2024, <https://github.com/MicrosoftLearning/mslearn-ai-services>.