

# Real Estate Document Processing System

## Overview

This application processes real estate documents using OCR and NLP, extracts relevant data, and stores it in a MySQL database. It uses Azure Blob Storage for file handling, Azure OpenAI for NLP, and LangChain with Azure AI Search for semantic indexing.

## Architecture

- User uploads real estate documents (PDF/Images).
- Azure Blob Storage stores raw files.
- Azure Function App downloads files, applies OCR (PyMuPDF), extracts text.
- Extracted text is parsed via LangChain & OpenAI for entity recognition.
- Processed data stored in MySQL.
- Azure AI Search indexes the documents.
- Results saved back to Blob Storage for downstream systems.

## Tech Stack

- Python (FastAPI, PyMuPDF, LangChain)
- Azure (Blob Storage, OpenAI, Cognitive Search)
- MySQL (Relational Database)
- Docker (Containerization)