

Homework 1: Receipt Image Recognition

Student name:Chen Junhao

SID:1155251265

Environment Configuration and Dependency Management

Install the Python package “langchain-openai” and use the OpenAI interface to call Google Gemini.

Model Data Acquisition and Preprocessing

Download the receipt image dataset via Google Drive and implement one-click downloading using the gdown library to ensure data accessibility. Unzip the file to obtain the receipt images.

Use the `image_to_base64()` function to convert images into Base64-encoded strings, ensuring secure transmission of image data through APIs.

Use the `get_image_data_url()` function to construct Data URLs in the format required by the API.

Implementation Process

1. Receipt Recognition

Use the ChatOpenAI interface provided by LangChain to initialize a multimodal large language model for receipt content parsing and Q&A.

Define the `llmchat(images, question)` function (where images is a list of receipt image paths, and question is a user's natural language query):

First, construct a SystemMessage to define the model's overall behavioral role.

Then, construct a HumanMessage whose content is a list containing:

A text segment (i.e., the user's question).

Multiple `image_url` type entries, each corresponding to a Data URL of a receipt image.

Pass the message list to `llm.invoke(messages)`, and the model automatically performs multimodal understanding, returning structured or semi-structured text results. The function ultimately returns the model's output content as the basis for subsequent parsing or direct display. This encapsulation decouples the model invocation logic from the business logic, making subsequent calls more concise.

Define the `get_single_receipt_data(image_path)` function:

Construct a system message instructing the model to act as a "receipt parser."

Use system prompts to constrain the model's output format, explicitly requiring it to output a JSON object containing the actual payment amount and the original price (pre-discount).

Automate the processing of multiple receipts by iteratively calling the single receipt parsing function.

2.Intent Recognition

Determine which type of amount field to aggregate by analyzing semantic keywords in the user's question.

The model decides whether to calculate and aggregate the actual payment amount or the original pre-discount price based on the presence of keywords (e.g., discount, original, actual, etc.) in the question.