

Arquitectura del Proyecto

1. Introducción

El proyecto busca implementar un AI Agent que interactúe con los usuarios en tiempo real, manteniendo el contexto de la conversación y optimizando el almacenamiento de memoria a corto y largo plazo. Para lograr esto, se utiliza una combinación de **LangGraph**, **FastAPI**, **Redis**, **MongoDB**, **FAISS** y **The Good Trip**.

FAISS proporciona capacidades de **búsqueda semántica** y **RAG (Retrieval-Augmented Generation)**, permitiendo al AI Agent acceder a documentos internos de la compañía para mejorar la precisión de sus respuestas. The Good Trip, una API desarrollada internamente, permite al agente realizar acciones sobre el **PSS (Passenger Service System)** de la aerolínea, como consulta de vuelos, emisión de tickets y cambios en reservas.

Objetivo del Documento

Este documento describe la arquitectura del sistema, con un enfoque en cómo se gestiona el estado de las conversaciones utilizando estos componentes clave.

2. Componentes Tecnológicos

LangGraph

- Se utiliza para estructurar el flujo de conversación del AI Agent.
- Permite manejar múltiples caminos conversacionales y estados dentro del diálogo.
- Optimiza la toma de decisiones dentro de la interacción del usuario.

FastAPI

- Framework de backend basado en Python para construir APIs rápidas y escalables.
- Gestiona las solicitudes del usuario y la interacción con Redis y MongoDB.
- Facilita la integración con LangGraph para manejar las respuestas del AI Agent.

Redis (Memoria de Corto Plazo)

- Se usa como **almacén temporal** para guardar el estado de las conversaciones activas.
- Almacena el contexto del usuario en **memoria RAM** con expiración automática.
- Evita consultas innecesarias a la base de datos y mejora el rendimiento.

MongoDB (Memoria de Largo Plazo)

- Se usa para almacenar conversaciones a largo plazo.
- Permite consultar interacciones pasadas cuando el usuario regresa tras un largo periodo.
- Soporta datos semiestructurados y es ideal para almacenar diálogos en JSON.

FAISS (Búsqueda Semántica y RAG)

- Proporciona la capacidad de realizar **RAG (Retrieval-Augmented Generation)** sobre documentos internos de la compañía.
- Permite al AI Agent recuperar información relevante de grandes volúmenes de datos no estructurados.
- Mejora la precisión de las respuestas al acceder a conocimiento especializado en tiempo real.

The Good Trip (API de Integración con PSS de la Aerolínea)

- API desarrollada internamente que permite la interacción del AI Agent con el **PSS (Passenger Service System)** de la aerolínea.
- Facilita acciones como **consulta de vuelos, emisión de tickets, cambios en reservas, y gestión de pasajeros**.
- Actúa como una capa de integración segura entre el agente y los sistemas internos de la aerolínea, mejorando la automatización de procesos operativos.

3. Arquitectura del Sistema

Flujo de Datos

1. **El usuario inicia una conversación** → FastAPI recibe la solicitud.
2. **FastAPI consulta Redis** para recuperar el estado de la conversación.
3. **Si existe un contexto en Redis**, se carga y se envía al AI Agent con LangGraph.
4. **El AI Agent responde** y el estado actualizado se almacena nuevamente en Redis.

5. **Si la conversación termina**, se almacena en MongoDB para consultas futuras.

Diagrama de Flujo

Usuario → FastAPI → Redis (si existe contexto) → LangGraph → Respuesta