PROJECT

PUBLIC SERVICE ASSISTANT USING RAG

Presented by iSheero

TABLE OF **CONTENTS**

01
CONTEXT

02
PROBLEM

03
SOLUTION

CONTEXT

- Digitalization of administrative services with over 150 services accessible online.

- Adoption in january 2023 of the the National Strategy for Artificial Intelligence and Megadata (SNIAM) making benin a major ai player in west africa by 2027.

PROBLEM

No help tool to make it easier to find the links to these services and get the needed very quickly.
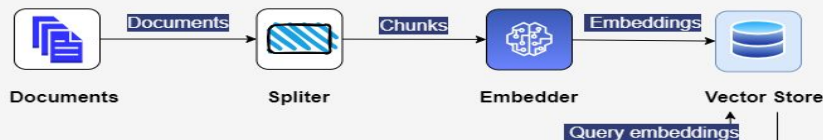
SOLUTION

Develop a conversational agent (chatbot) based on RAG system to answer all questions related to public services in Benin.
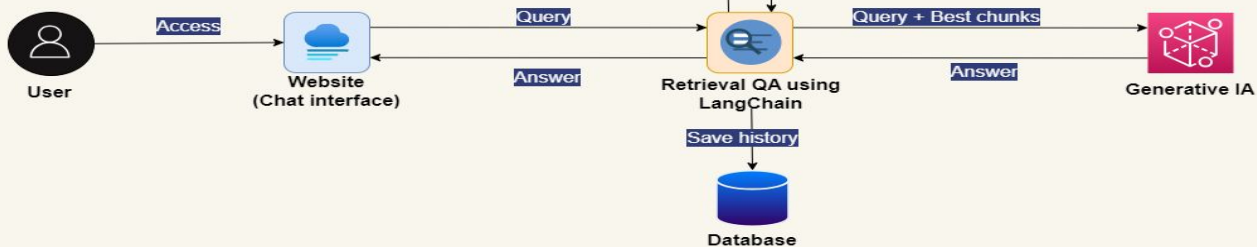
SOLUTION (CONTINUED)

1. Load documents
Phase 1 : Read documents, split in chunks, extract embeddings and store in vector store

Documents → Documents → Spliter → Chunks → Embedder → Embeddings → Vector Store

2. Query documents
Phase 2 : Get user query, extract embeddigs, retrieve chunks by similarity, send question + chunks to Generative AI for summary

Query embeddings
Best chunks

User → Access → Website (Chat interface) → Query → Retrieval QA using LangChain → Query + Best chunks → Generative IA

Answer ← Website (Chat interface)
Answer ← Retrieval QA using LangChain

Save history → Database

SOLUTION (CONTINUED)

01

MODELS SELECTION

02

DATA PROCESSING

03

CHOICE OF VECTOR STORE

04

GENERATOR FINETUNING

05

MODELS EVALUATION

06

SET UP A LANGCHAIN AGENT

07

DEPLOYMENT

RESOURCES

- Public service website https://service-public.bj/

- Build a RAG system using Langchain:

https://learn.deeplearning.ai/courses/langchain-chat-with-your-data

- Retrieval with Langchain

https://python.langchain.com/v0.1/docs/modules/data_connection/

- Retrievers evaluation metrics https://hal.science/hal-03467001/document

- Embeddings models leaderboard https://huggingface.co/spaces/mteb/leaderboard

- LLM Chatbot leaderboard

https://huggingface.co/spaces/lmsys/chatbot-arena-leaderboard

- LLM finetuning https://www.leewayhertz.com/parameter-efficient-fine-tuning/

- LLM evaluation metrics

https://fabianofalcao.medium.com/metrics-for-evaluating-summarization-of-texts-performed-by-transformers-how-to-evaluate-the-b3ce68a309c3

GOOD LUCK