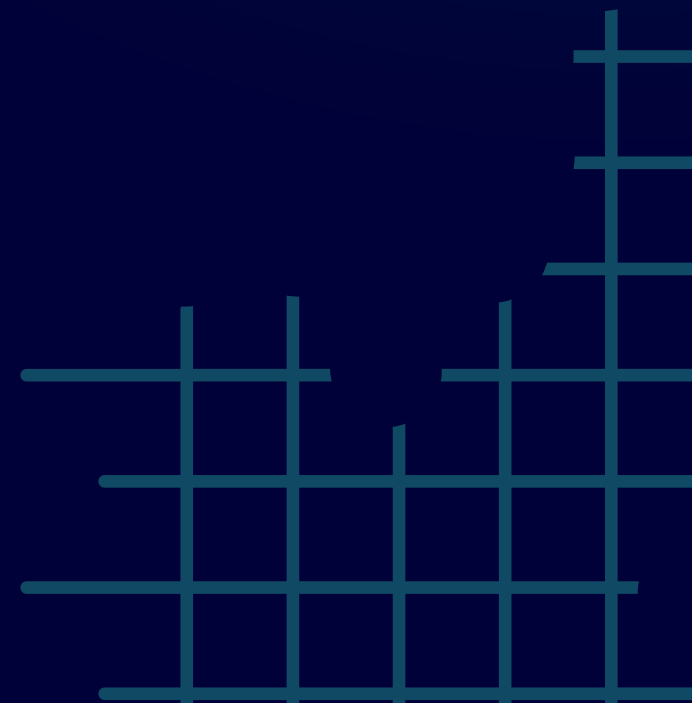


Assistant IA des services public - utilisation du RAG



Sommaire

1

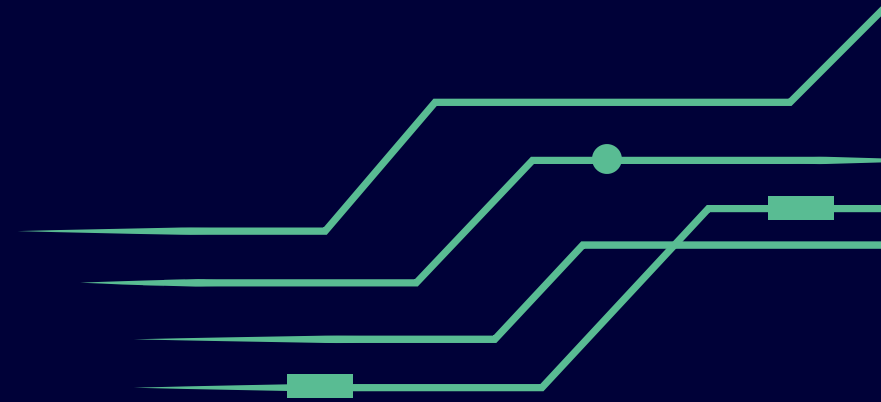
Contexte

2

Problème

3

Solution



Contexte



- Numérisation des services administratifs à travers le continent, permettant à des centaines de services d'être accessibles en ligne, améliorant ainsi la qualité des services publics.
- Adoption de Stratégies Nationales pour l'Intelligence Artificielle et les Big Data dans plusieurs pays africains, visant à positionner l'Afrique comme un acteur majeur dans le paysage mondial de l'IA en exploitant les technologies innovantes et en renforçant l'expertise régionale

Problème



Aucun outil d'aide ne permet de trouver plus facilement les liens vers ces services et d'obtenir très rapidement les informations nécessaires.

Solution

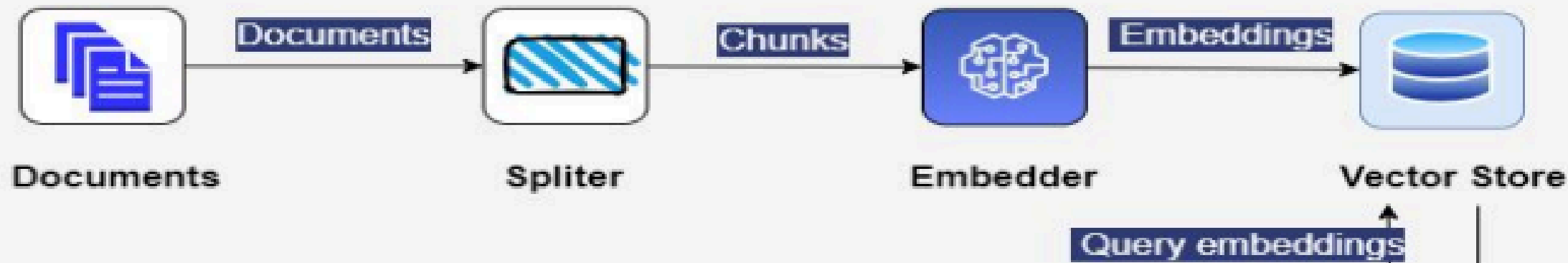


Développer un agent conversationnel (chatbot) basé sur la méthode RAG pour répondre à toutes les questions liées aux services publics.

Solution

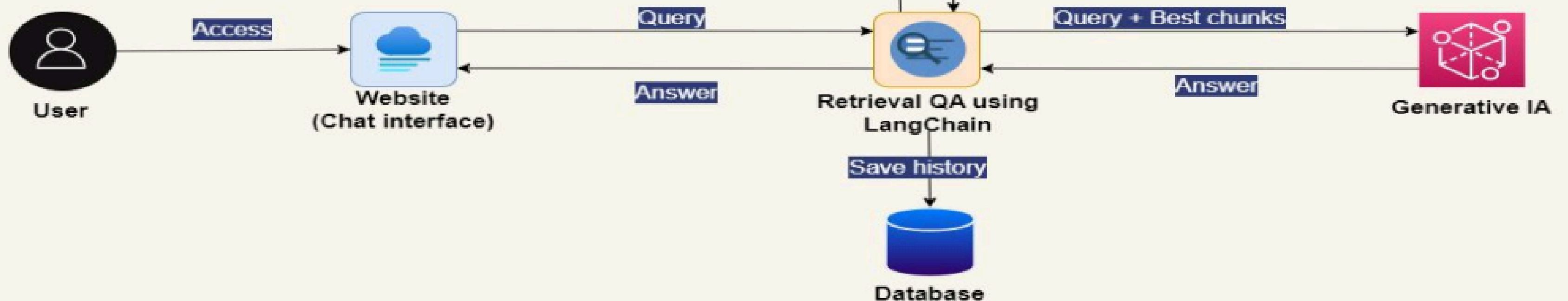
1. Load documents

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



2. Query documents

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



Solution



Résumé des flux principaux

- Préparation des données : Lecture → Division → Embedding → Stockage vectoriel.
- Interrogation utilisateur : Embedding de la requête → Recherche vectorielle → Enrichissement du contexte → Réponse générée.

Solution

1

choix du modèle

2

Préparation des
données

3

Choix du Vector
Store

4

Fituning du modele
de génération de
texte

5

Evaluation du
modèle

6

Configurer un
agent Langchain

7

Déploiement



Ressources



- 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>



Bonne Chance