

Coding Assignment - RAG Mini Pipeline

Objective

You are tasked to build a simple Retrieval-Augmented Generation (RAG) pipeline.

You have been provided a ZIP file containing 30 small knowledge base documents.

Task Description

1. Embed all documents using any embedding model (e.g., Sentence Transformers).
2. Accept a user query as input.
3. Retrieve the top-3 most relevant documents based on cosine similarity.
4. Generate a synthetic answer by summarizing the retrieved content.
5. (Optional Bonus) Visualize the embeddings in 2D using t-SNE or UMAP.

Dataset

The dataset consists of 30 small text files related to FutureTech Innovations' services, case studies, leadership, governance, and initiatives.

Dataset ZIP File: knowledge_base.zip

Sample Queries

- What industries has FutureTech helped with AI solutions?
- What are FutureTech's sustainability practices?
- What is the GenAI initiative focused on?
- How can customers get technical support from FutureTech?
- Describe a case where FutureTech improved retail customer experience.
- Who are FutureTech's strategic cloud partners?
- What measures does FutureTech take for Responsible AI?

Submission Requirements

- Python scripts or Jupyter Notebook (.py or .ipynb).

Coding Assignment - RAG Mini Pipeline

- README file briefly explaining your approach.
- (Optional) Screenshots of visualizations if bonus attempted.

Evaluation Criteria

- Correctness and completeness of the RAG pipeline.
- Code readability and modularity.
- Proper retrieval and accurate answer generation.
- (Bonus) Embedding visualization.
- (Bonus) Performance optimizations or creative improvements.