SIVAKUMAR M (20MIA1002)

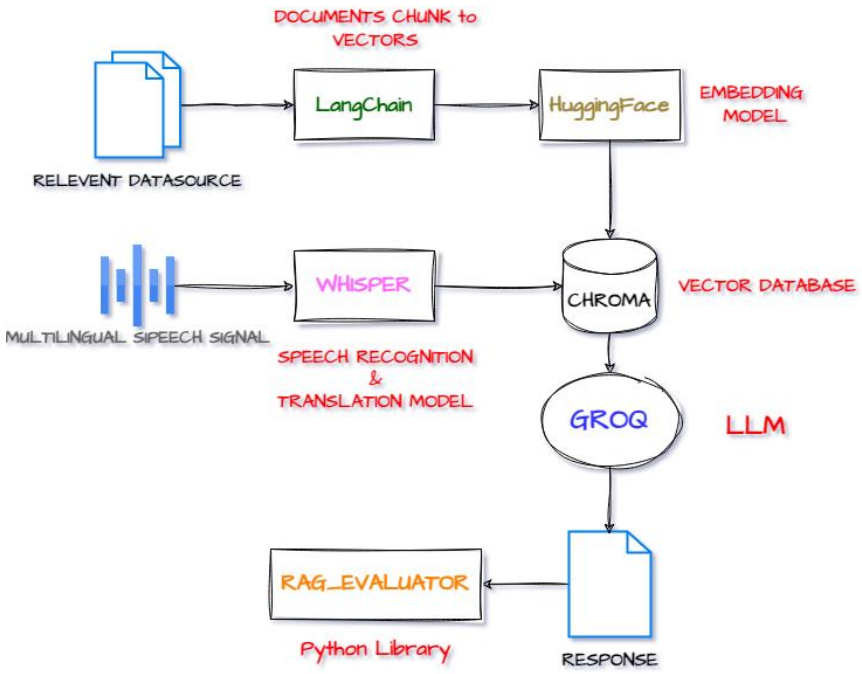
VELLORE INSTITUTE OF TECHNOLOGY – CHENNAI

Building a Multilingual Speech Recognition Model for RAG without Training

Objective:

To build a multilingual speech recognition model without training, using a pre-trained multilingual speech recognition model, to enable RAG to perform task in multiple language.

Architecture:



Procedures:

- Generate a necessary API keys (LangChain & Groq)
- Create a dummy RAG source file on any specific topic
- Use LangChain TextLoader to load the data. Then the contents and metadata of the file is stored
- The document chunks are converted into numerical representation using HuggingFaceEmbedding Model.
- The embeddings are stored in Chroma vector database for retrieval based on cosine similarity.

- Multilingual Speech Query is parsed to Whisper for Speech recognition and Translation process.
- Based on the Translation input, Groq LLM model will retrieval relevant data from Chroma.
- The retrieved text is then compared with the manual Ground Truth using Rag_evaluator.

Results:

- BLEU: Measures the overlap between the generated output and reference text based on n-grams.
- ROUGE-1: Measures the overlap of unigrams between the generated output and reference text.
- BERT Score: Evaluates the semantic similarity between the generated output and reference text using BERT embeddings.
- Perplexity: Measures how well a language model predicts the text.
- Diversity: Measures the uniqueness of bigrams in the generated output.
- Racial Bias: Detects the presence of biased language in the generated output.

Metrics	**French**	**German**	**Hindi**
BLEU	14.144	28.533	**35.006***
ROUGE – 1	0.461	0.435	**0.619***
BERT Precision	0.732	0.723	**0.777***
BERT Recall	0.791	0.860	**0.871***
BERT F1	0.760	0.786	**0.821***
Perplexity	32.929	**13.619***	23.759
Diversity	**0.965***	0.947	**0.965***
Racial Bias	0.491	**0.474***	0.482

Interpretation:

- Hindi Speech Sample has the Highest Model Accuracy compared with French and German on parameters BLEU, ROUGE-1 and BERT Score.
- German Speech Sample has the minimum perplexity which means it has more coherent text to the reference.
- Both the French and Hindi Speech has high Diversity value which infer it has large vocabulary and structure
- German has the minimum Racial Bias

GOOGLE COLAB LINK:

https://colab.research.google.com/drive/1J2pfi6QiRkyjyRaG1TSG5PsVgaS9ADES?usp=sharing