Enabling Multilingual RAG Tasks Using OpenAI Whisper

Project Objective:

To develop a multilingual speech recognition model without training, utilizing a pre-trained multilingual speech recognition model like OpenAI Whisper, to empower RAG to execute tasks in multiple languages.

Background:

RAG (Retrieval-Augmented Generation) is a generative model capable of various tasks, including speech recognition, translation, and summarization. However, RAG's current capabilities are limited to a single language. By creating a multilingual speech recognition model without training, we can expand RAG's task repertoire to multiple languages without requiring additional training.

Methodology:

1. Leveraging OpenAI Whisper for Multilingual Speech Recognition:
   OpenAI Whisper is a state-of-the-art multilingual speech recognition model capable of transcribing and translating speech into over 98 languages. By integrating OpenAI Whisper into our approach, we can bypass the need for training a separate speech recognition model for each language.

2. Adapting RAG for Multilingual Tasks:
   To enable RAG to perform tasks in multiple languages, we will adapt its input processing pipeline to handle multilingual input. This involves utilizing OpenAI Whisper's transcription capabilities to convert multilingual speech input into a unified representation that RAG can process effectively.

3. Evaluation and Refinement:
   The proposed approach will be evaluated using benchmark datasets and real-world tasks to assess its performance and identify areas for improvement. Based on the evaluation results, we will refine the approach to further enhance its accuracy and robustness.

Expected Outcomes:

- Development of a multilingual speech recognition model without training.
- Integration of OpenAI Whisper to enable RAG's multilingual capabilities.
- Evaluation of the proposed approach using benchmark datasets and real-world tasks.
- Refinement of the approach based on evaluation results.

Project Deliverables:

1. Implementation of a multilingual speech recognition model using OpenAI Whisper.

2. Integration of the multilingual speech recognition model with RAG.
3. Evaluation report detailing the performance of the proposed approach.
4. Refined approach incorporating insights from the evaluation.

Resources:

- OpenAI Whisper pre-trained multilingual speech recognition model
- RAG generative model

Conclusion:

By leveraging OpenAI Whisper's multilingual speech recognition capabilities, we can expand RAG's capabilities to multiple languages without the need for extensive training. This approach has the potential to significantly enhance RAG's versatility and applicability in a wide range of real-world scenarios.