Brance <AI Applied Researcher - Intern Hiring RAG>Task

Name: Swetanshu Pandey

Linkedin Profile:
https://www.linkedin.com/in/swetanshu-pandey-54b08a1b8/

Date Challenge Received: 06-July-2023
Date Solution Delivered: 09-July-2023

---

1. Problem Statement

What was the task and how, you understood it ?
Ans:-
The task was to build a chatbot using a Retrieval Augmented Generation,which is a language generation model.It consists of two phases, one is the Retrieval Phase and the other is the Generation Phase.
So, I have to build the chatbot using both the phases.

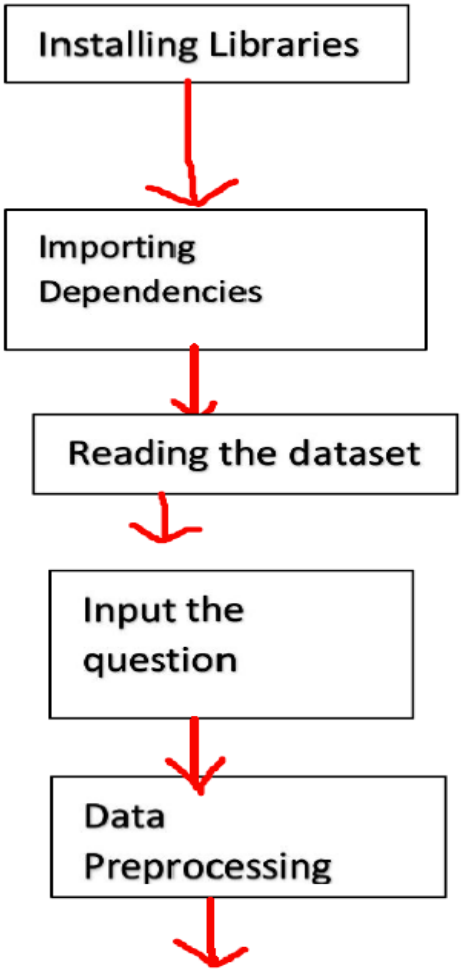
2. Approach

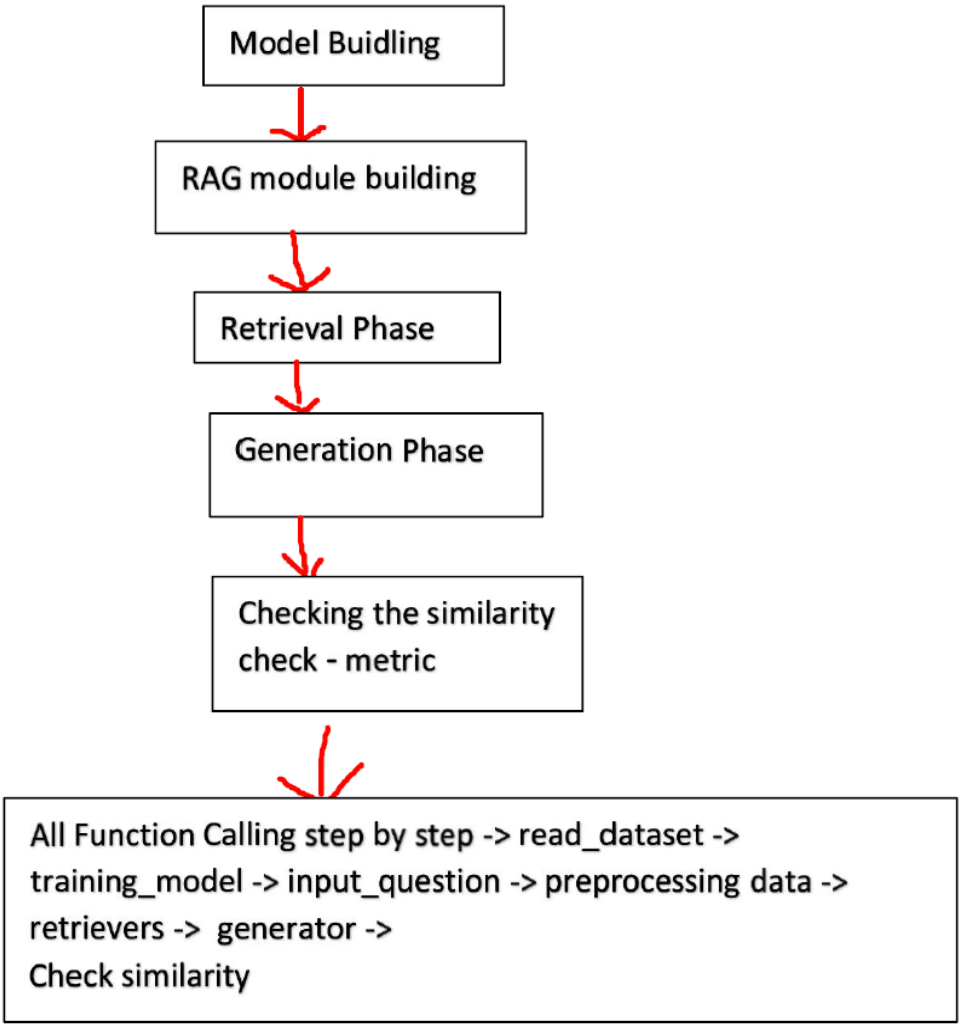
Your approach to the problem. Mention any assumptions made.
Ans.:-
I firstly divided the task into two phases, Retrieval and Generation.Did some preprocessing of the provided data and finally builded the RAG module and passed that preprocessed data as an input to the RAG module.
Prepared a function so that a user can type its own question and based on that ,the answer to the user will be provided.

3. Solution

Details about your solution. Illustrate performance and design with diagrams.
Ans:-
I performed the entire coding and building in the Google Colab. Started up with installing some necessary libraries which aren't pre-present in google colab. Created functions to read the dataset.After that created a function to take input of a user.Then created function for data pre-processing.Now I came to the RAG model building part.Created a function “training_model” that when called, could initialize the model.After that created a function “creating_input_dict_and_outputs” ,this function will perform seq2_seq tokenizing on the input dataset.After that created Retrieval Phase, which will retrieve the relevant answers from the data

that best aligns to the user question.Then the generator question, that will provide the generated string for the asked question from user side.
Then created a check_similarity function that will keep in check of how much the generated answer aligns the factual answers.
Finally, one by one ,I called all the functions.
The performance was not so good, it was about 35% in similarity.





4. Future Scope

Thoughts on how you could have improved the solution.

Ans:-
There are more fine tuning methods that I haven't used in building the RAG chatbot.
More than that, I have trained this RAG model on only custom data of 50 rows only (provided data), so I can use a large amount of dataset for training the RAG model.
This would have definitely improved the solution.