13520062 - Rifqi Naufal Abdjul
13520085 - Ubaidillah Ariq Pratama

Chatbot for Indonesia High School Curriculum

Problem Statement

In Indonesia, the challenge of providing equal access to quality education across its vast archipelago is pressing. Many remote and underserved areas lack sufficient qualified teachers and educational resources. To address this issue, a chatbot tailored to the Indonesia high school curriculum can be developed, leveraging advanced technology to bridge the educational gap.

The chatbot, based on a Retrieval-Augmented Generation (RAG) system, integrates large language models (LLMs) with a retrieval component that taps into Indonesian educational books and resources. This integration ensures that the chatbot's responses are not only accurate but also relevant to the Indonesian educational landscape. It can provide region-specific examples, cultural references, and pedagogical approaches that resonate with Indonesian high school students.

By using a chatbot with a RAG system, Indonesia can significantly enhance educational outcomes. The chatbot can offer personalized learning pathways, immediate feedback, and up-to-date educational content. This solution has the potential to improve learning experiences, bridge educational gaps, and ultimately contribute to a more equitable education system nationwide.

Related Work

Re2G: Retrieve, Rerank, Generate (Glass et al., 2022)

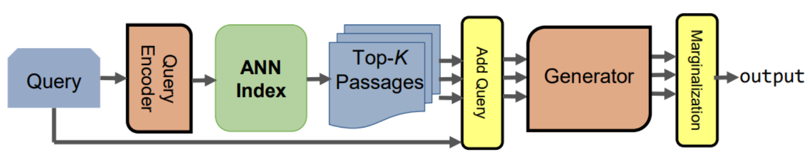


Figure 2: RAG Architecture

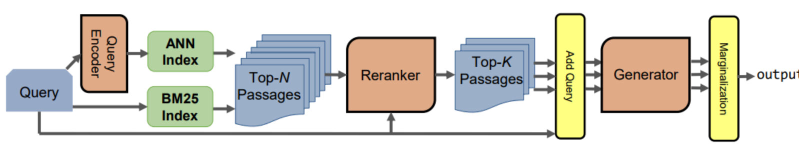


Figure 3: Re^{2}G Architecture

In order to improve the performance of retrieval augmented generation, a reranking process is added to the pipeline. A BM25 index and a reranking model (text encoder model) is used to rearrange the retrieved documents and sort it based on the relevancy.

Chatbot application in a 5th grade science course

This study has shown that the chatbot application designed for science courses positively affected students' online education learning process; students found the chatbot useful and entertaining and wanted to use it in other lessons. It was a good assistant in learning outside the classroom. It allowed them to repeat the lesson and facilitated learning by being interactive. Since the application enabled students to see each other's questions and answers, it also encouraged them to complete the missing information they had on the topic. Also, while the teacher interacts with the students at a certain time and in a certain environment, the chatbot continuously provides information to students on the topic thanks to the flexibility of place and time.

Input/Output Behavior

Input: Natural language query

Output: Generated answer based on the relevant documents

Example:

- Input (Query): “Who is Indonesia’s first president?”
- Relevant Document: PPKN Bab 3, Hal 54
- Output (Answer): “Indonesia’s first president is Ir. Soekarno with his Moh. Hatta as his Vice President”

Evaluation Metric

Retrieval & Reranking Evaluation

We will do accuracy measurement on the most relevant document. We will choose an evaluation query that is related only to a single document. Based on the retrieval results, we will calculate the percentage of successful retrieval that have the target documents being the most relevant document.

Content Generation Evaluation

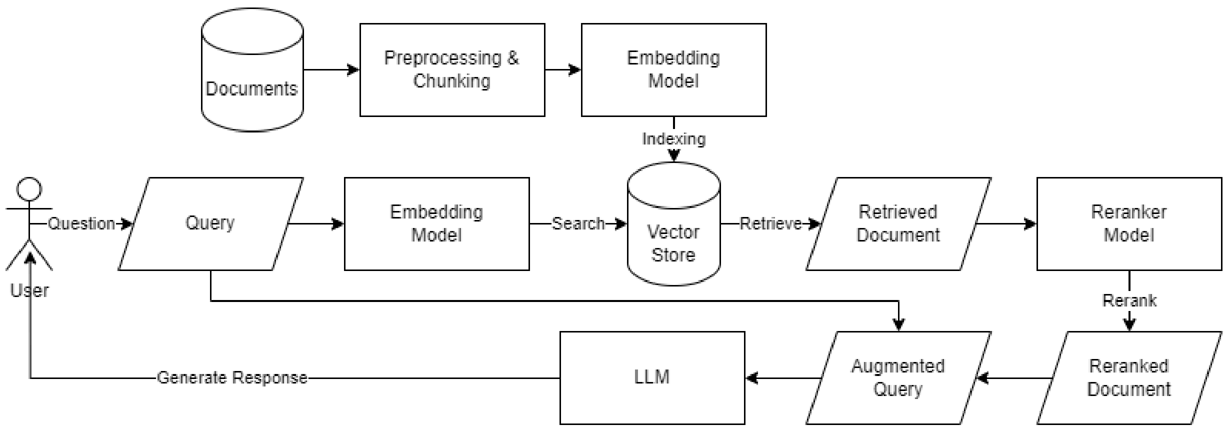
Because of the complexity of the evaluation process. We will simplify the evaluation to use manual human evaluation based on 4 aspects of the generated content.

1. Answer Relevancy: How relevant is the answer to the question at hand
2. Faithfulness: How factually accurate is the answer given the context
3. Correctness: How accurate is the answer against the ground truth data
4. Semantic similarity: How closely does the answer match the context in terms of meaning (semantics)

We will score them 0-4, where 0 is the lowest score and 4 is the highest score. To make it fair, we will have multiple people (2 people) scoring every result.

Methodology

Here is the overview flow of the system, adapted from the ReG architecture on doing retrieval augmented generation.



To simplify things, we will implement the system using third-party solutions like Cohere, OpenAI, and Claude since they perform very well as SOTA solutions and are cheap. Here are the overview of the resource that will be used on this system,

1. Document source. Sourced from official kemdikbud's website.
2. Embedding model. OpenAI text-embedding-3-small model.
3. Vector store. SQLite, indexed using SQLite-VSS extension
4. Reranking model. Cohere Rerank-2 model fine-tuned.
5. Generative LLM. Anthropic Claude 3 Haiku model.

The system consists of 3 main processes: vector store creation, content retrieval, and content generation.

1. Vector store creation
   The document will be preprocessed and chunked per chapter per chapter. The chunked document will be transformed into embedding and then stored on vector store
2. Content retrieval
   User's query will be transformed into embedding using the same embedding model and used to do similarity search on the vector store. Search results then will be fed into the reranking model to output the most relevant document.
3. Content generation
   Putting the relevant documents into context. Generative LLM will answer the user's query using the relevant documents.

References

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... Kiela, D. (2021). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv preprint arXiv:2005.11401.

Glass, M., Rossiello, G., Chowdhury, M. F. M., Naik, A. R., Cai, P., & Gliozzo, A. (2022). Re2G: Retrieve, Rerank, Generate. arXiv preprint arXiv:2207.06300.

Deveci Topal, A., Dilek Eren, C. & Kolburan Geçer, A. Chatbot application in a 5th grade science course. Educ Inf Technol 26, 6241–6265 (2021). https://doi.org/10.1007/s10639-021-10627-8