

BERT/RAG-Based Question-Answer Generator

Akhil Karra, Charis Graham, Frank Sacco,
Ram Potham



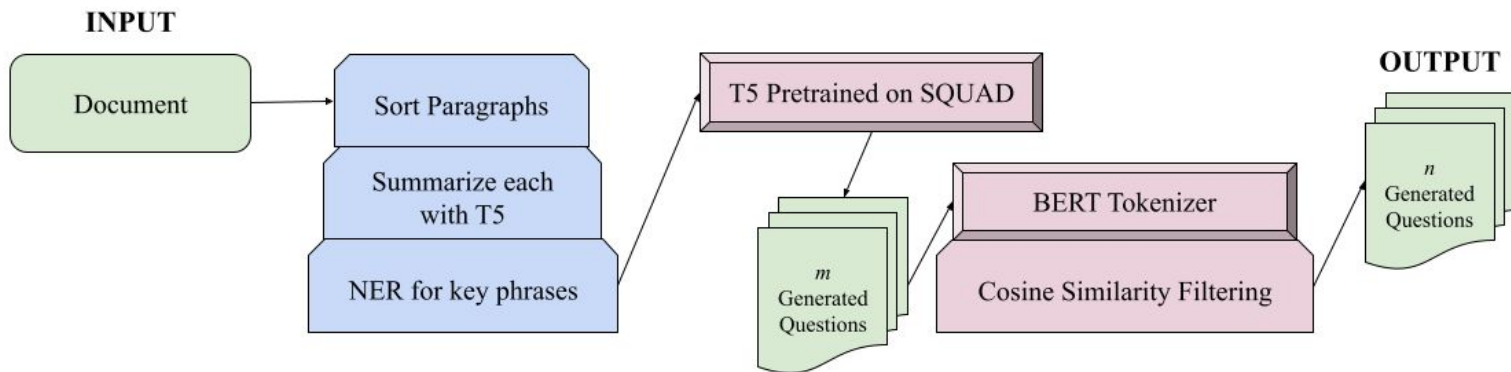


Naive Question Generation Approach

- Methodology: Part-of-speech tagging piping into templates
- Limitations: Lack of contextual understanding and question complexity
- Results: Not very good, mostly because there are not very many templates

Question Generation

- Use T5 base model to summarize each paragraph
- Named Entity Recognition to find key phrases shared between summary and paragraph
- T5 pre-trained on SQUAD to generate question using key phrase and summary
- Filter questions using cosine similarity on BERT embeddings





Limitations / Potential Improvements

The Naive Approach is good as is, it was never intended to be good, but we could potentially make it better by just making more templates. (But thats boring!)

The Advanced Approach had 2 critical error types. Sometimes, there would be issues where questions are too similar to each other or questions about tangential / unrelated subjects. Both of these are hard to remove or even track without running another NLP level on the stack, which would strain our resources budget.



Example Outputs (cherry-picked)

Naive Answers

What do kangaroos have?

What do kangaroos fare?

What do kangaroos release?

What do kangaroos adapt?

What do kangaroo use?

Advanced Answers

What is the far-northern equivalent of the eastern and western grey kangaroos?

The antilopine is far-northern equivalent of what other kangaroo?

What type of kangaroo can be 2 meters tall?

What does Adam McDonald Jr think the ability to travel long distances at moderate speed is crucial to survival?

What is the collective noun for a mob, troop or court?

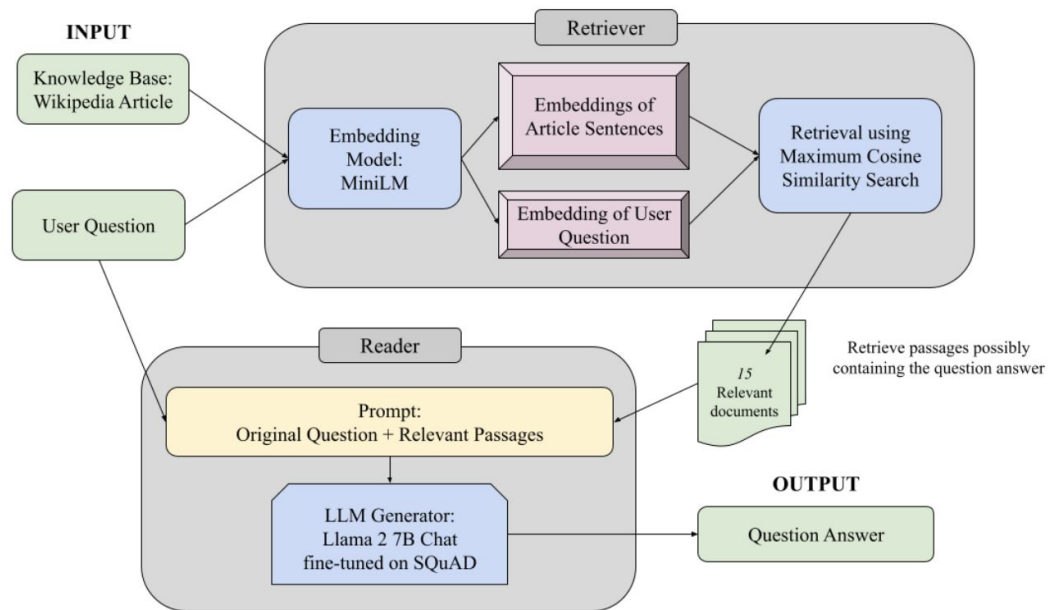


Naive Answer Generation Approach

- Methodology: Cosine similarity between sentence vectors
- Limitations: Uses Bag-of-Words implementation so loses syntactic relationships as well as bases off of true similarity rather than semantic similarity
- Results: Not very good, mostly due to loss of syntactic information and lack of true semantic relationship representation

Answer Generation

- Use MiniLM to embed both Wikipedia article and user question
- Maximum cosine similarity search to find top 15 most similar passages to query
- LLaMa 2 Chat 7B fine-tuned on SQuAD generates answer based on question and relevant documents





Limitations / Potential Improvements

The Naive Approach is good as is, it was never intended to be good. If we were to improve it, we would take syntactic relationships into consideration and abandon the Bag-of-Words model.

The Advanced Approach had the major issue of generating answers that were too long. The system produced answers that were semantically similar to the true answers, but more wordy. This could be improved via prompt engineering, post-processing on the answer, and further fine-tuning on SQuAD.



Example Outputs for Naive Version (cherry-picked)

Question	Generated Answer
Are turtles related to snakes?	In some species, shells may have red, orange, yellow, or grey markings and these markings are often spots, lines, or irregular blotches.
What is a turtle?	* Pritchard, Pether C H (1979).
Define "turtle."	Turtles are reptiles of the order Testudines (the crown group of the superorder Chelonia), most of whose body is shielded by a special bony or cartilaginous shell developed from their ribs.
What two major groups are turtles divided into?	In some species, shells may have red, orange, yellow, or grey markings and these markings are often spots, lines, or irregular blotches.



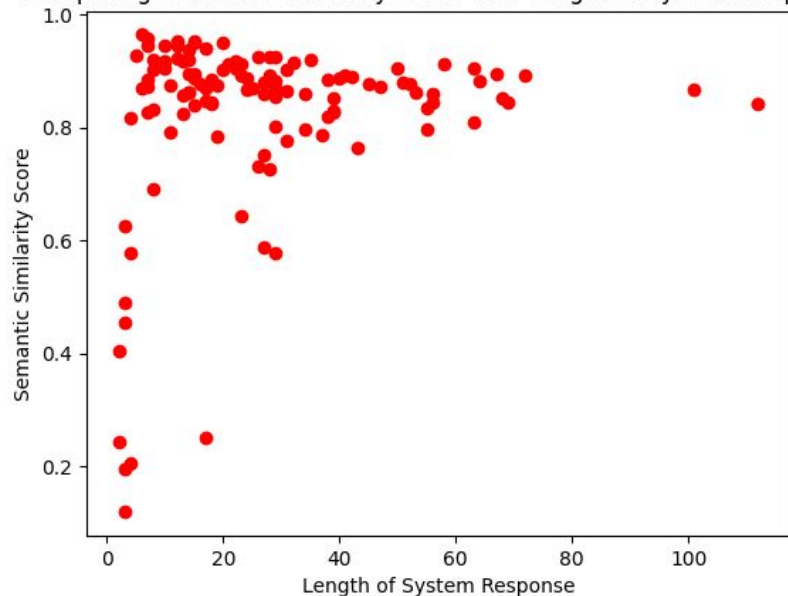
Example Outputs for Advanced Version Without Fine-Tuning (cherry-picked)

Question	Reference Answer	Generated Answer
Do ants belong to the Hymenoptera order?	yes	Yes, ants do belong to the Hymenoptera order. According to the context, ants, along with wasps and bees, belong to the order Hymenoptera.
Are ants used in cuisine?	yes	Yes, ants are used in cuisine in many human cultures, as mentioned in the context.
Does an ant's head contain sensory organs?	yes	Yes, an ant's head contains many sensory organs. According to the text, an ant's head contains sensory organs such as compound eyes, antennae, and metapleural glands.

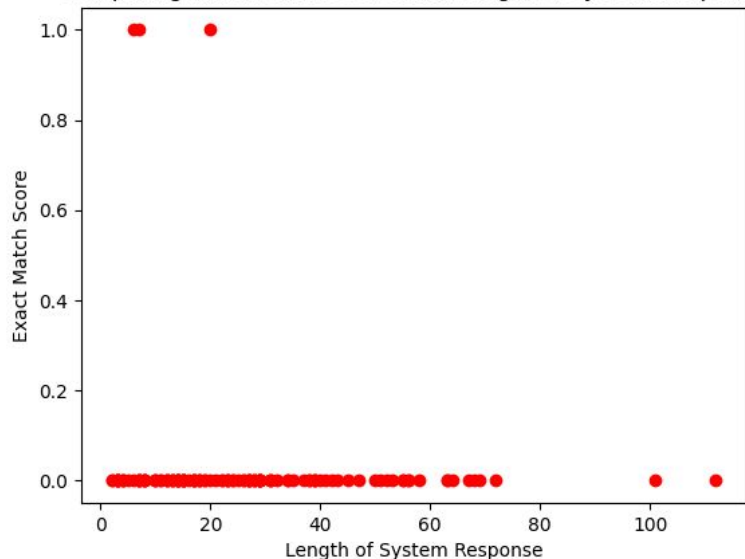


RAG-Based Answer Generation Approach Without Fine-tuned LLaMa 2

Comparing Semantic Similarity Score with Length of System Response



Comparing Exact Match Score with Length of System Response





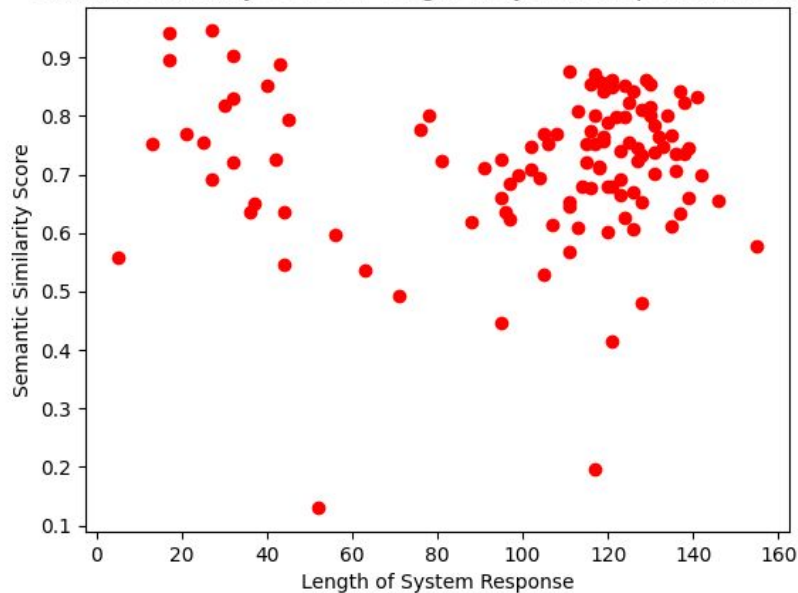
Example Outputs for Advanced Version With Fine-Tuning (cherry-picked)

Question	Reference Answer	Generated Answer
Do ants belong to the Hymenoptera order?	yes	Yes, ants are social insects of the family Formicidae (), and along with the related wasps and bees, they belong to the order Hymenoptera. These chemical signals are more developed in ants than in other hymenopteran groups. Ants fossilised in Baltic amber

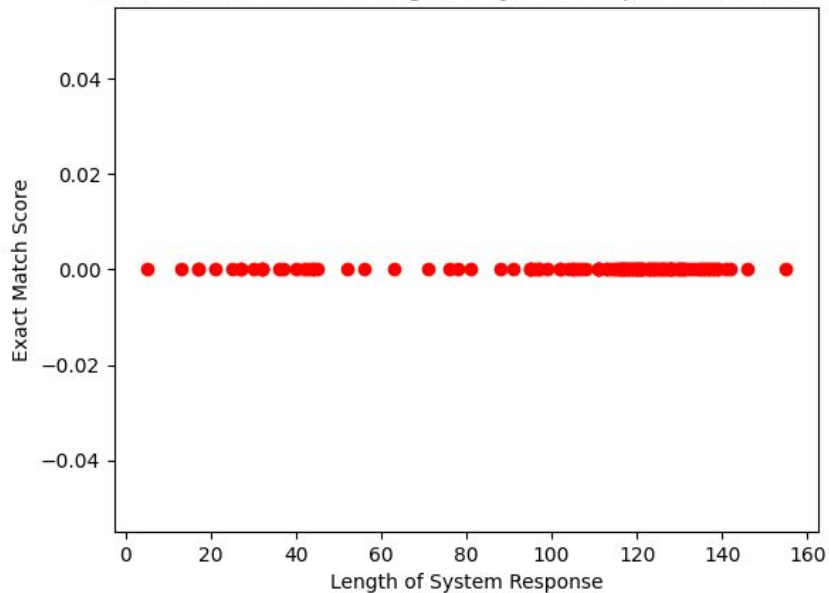


RAG-Based Answer Generation Approach With Fine-tuned LLaMa 2

Semantic Similarity Score vs. Length of System Response, Fine Tuned



Exact Match Score vs. Length of System Response, Fine Tuned





Thank you!

