NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Deep Learning based Question Answering System SCSE20156

Presented by: Tan Ching Fhen

Supervised by: Assoc Prof Yeo Chai Kiat

Co-supervisor: Dr Yan Shi Xing

Mentor: Sunil Sivadas, Josey Mathew

Transformers answers your Covid-19 Questions!

Motivation

The Covid-19 pandemic has brought about many rules and regulations in Singapore to mitigate viral spread. Sometimes, we may have questions about these rules. For instance, a tour organizer may ask 'What is the maximum number of people that can be in a tour group?' Though some questions are easily googled, other questions require analyzing through long documents which can be a hassle!

What if we could get the exact answers to our Covid-19 questions by asking a Question Answering System?

Methodology

A Question Answering (QA) System has two portions:

1. Retriever – finds documents that may contain answer

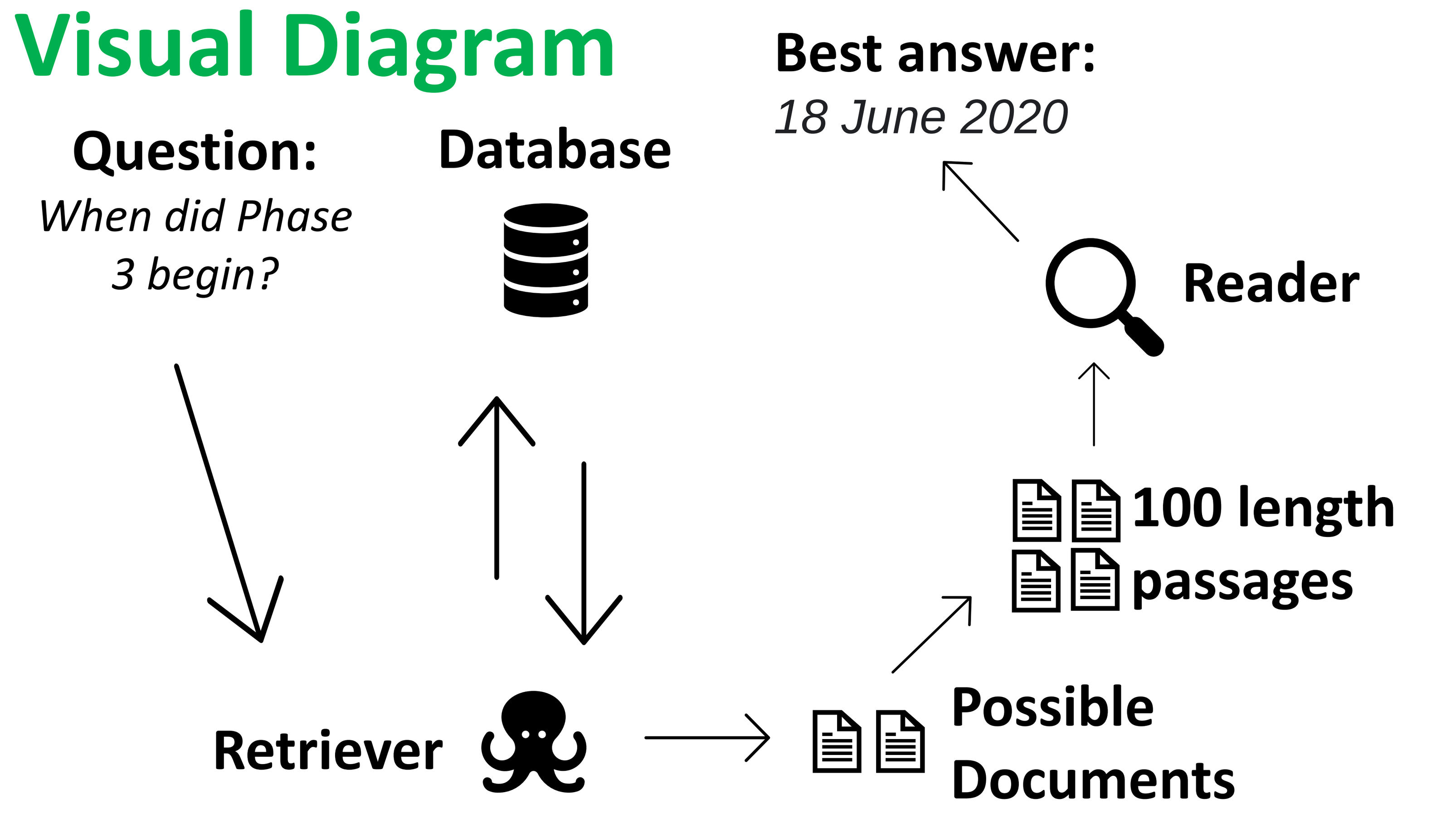
BM25	Dense Passage Retriever
Non-neural network model that uses word matching	Neural network model that finds documents using vectors
Pros: Does not need pre-training	Pros: Better performance
Cons: Poorer performance	Cons: Requires pre-training

2. Reader – finds the best answer among the documents

BERT	LSTM
Transformers variant model	Recurrent Neural Network model
Pros: Achieves state-of-the-art levels of performance	

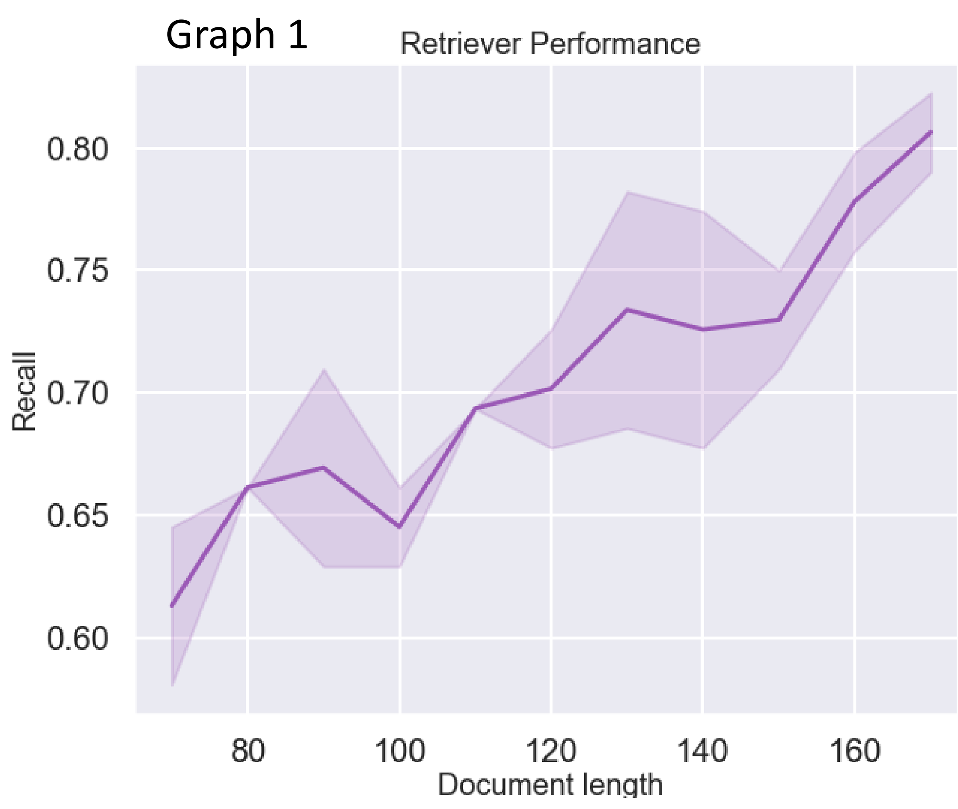
Out of the many options available, I chose BM25 to minimize amount of pre-training and BERT for best performance

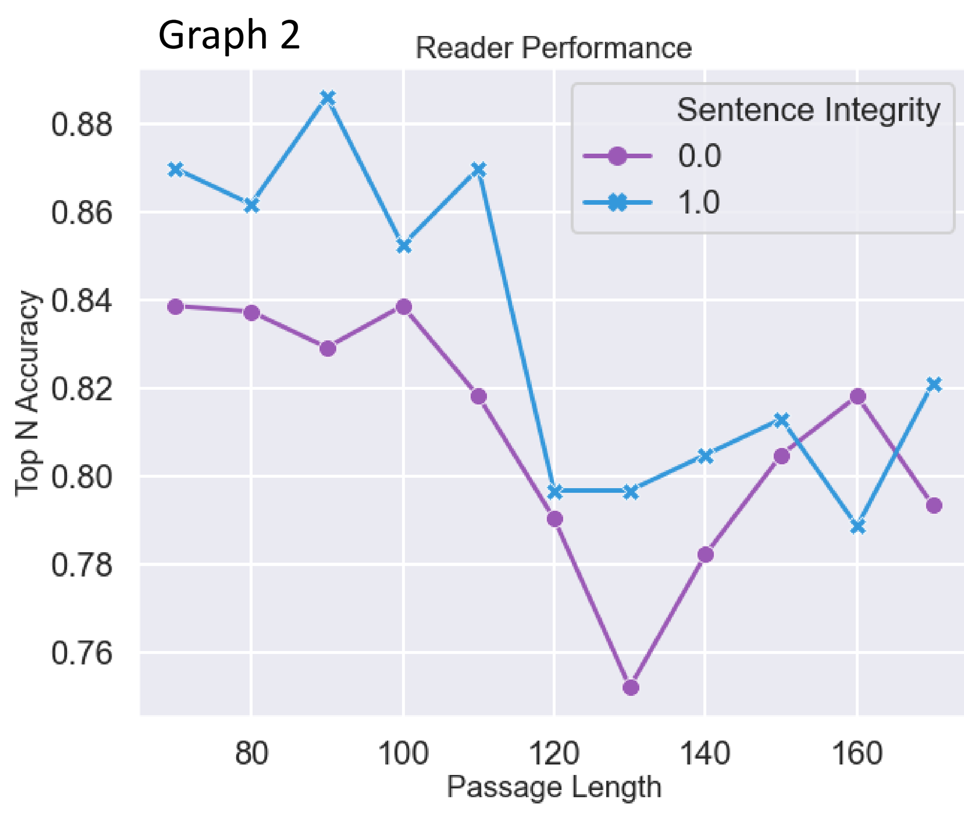
Visual Diagram



Findings

Many parameters need to be tweaked to optimize a QA system. One of which is the length of documents retrieved by the retriever. Based on my findings, **retriever performance increases linearly with document length** (Graph 1). This makes sense because BM25 uses word matching, so when more words are available, the retriever is more likely to retrieve the correct documents.
On the other hand, reader performance tapers off after length 100 (Graph 2). This is consistent with the finding that '**passages with 100 words works the best**' by Wang et. al. (2019). Therefore, an additional step I used was to split the documents retrieved by the retriever into 100 length passages, before feeding into the reader.





Lastly, maintaining sentence integrity is generally better (Graph 2) because the reader can better 'understand' sentences better if they are not split midway.

Conclusion

Through the utilization of BM25 (retriever) and BERT (reader), together with implementation of experiment findings, I was able to build a Covid-19 Question Answering System. This prototype can answer Covid-19 related questions.

QA System Prototype:

Question 1: How many people can attend a singing class?

Answer: One participant

Question 2: How many people are allowed at marriage solemnizations?

Answer: Up to 100

Future Work

In recent research, it has been found that Dense passage retriever (DPR) 'outperforms Lucene-BM25 system by 9-19%' (Karpukhin et. al, 2020). Via the use of vector similarity, it can find correct documents even if there are no matching words with the question. Thus, the next step would be to utilize DPR as the retriever in order to achieve even greater performance!

References:
Wang Z, Ng P, Ma X, Nallapati R, Xiang B, 2019. Multi-passage BERT: A Globally Normalized BERT Model for Open-domain Question Answering. arXiv:1908.08167v2
Karpukhin V, Oguz B, Min S, Lewis P, Wu L, Edunov S, Chen D , Yih WT, 2020. Dense Passage Retrieval for Open-domain Question Answering. arXiv:2004.04906v3