REPORT Group 21 (T14)

TASKS

The 2020 EPIC-QA track involves two tasks:

Task A: Expert QA There is a separate dataset for Experts which contain Research articles so that answers can provide information that is useful to researchers, scientists, or clinicians

Task B: Consumer QA There is other dataset for consumers which has general articles so that answers can be understandable by the general public.

We have taken data from preliminary round to perform the tasks.

Data Preprocessing:

The Dataset is cleaned and removed the unnecessary Symbols and we made a csv file having Column names: doc_id , Title , context_text

The csv file is "expert_context.csv" for Expert data and "consumer_context.csv" for Consumer data

TOP 10 DOCs Retreival:

The Query and Document Titles similarity is taken into consideration and returning the top 10 document ids using BM25 Model. We get the sentences from the top 10 documents.

Sentences ReRanking:

The sentences retrieved are reranked according to the question context using BERT Reranking and top ranked sentences is the passage annotated answer for the query.

BERT Question Answering Model for Answer Key Generation:

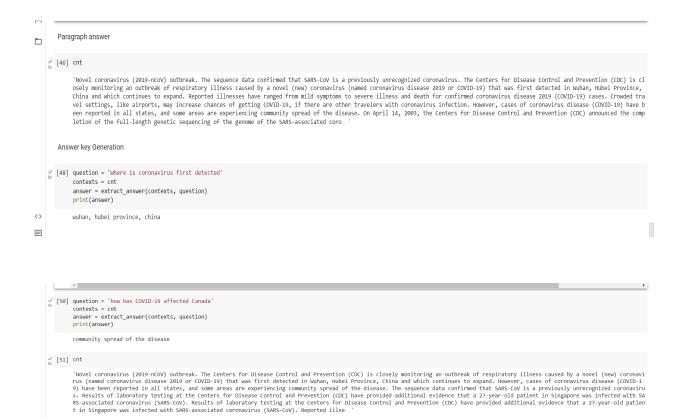
We take the top ranked sentences as context and feed the query and context into a pretrained BERT question answering model. We get the required answer.

Results:	,
----------	---

Query:

PASSAGE ANNOTATION:

ANSWER-KEY GENERATION:



Novelty:

Next Possible Question Generation:

Based on the Extracted Context and the Passage Annotated answer We have generated next possible questions the user may ask for his understanding.

You Can see the implementation in Expert QA.ipynb