

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 Retrieval:

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:

Paragraph answer

✓ [46] cnt

'Novel coronavirus (2019-nCoV) outbreak. The sequence data confirmed that SARS-CoV is a previously unrecognized coronavirus. The Centers for Disease Control and Prevention (CDC) is closely monitoring an outbreak of respiratory illness caused by a novel (new) coronavirus (named coronavirus disease 2019 or COVID-19) that was first detected in Wuhan, Hubei Province, China and which continues to expand. Reported illnesses have ranged from mild symptoms to severe illness and death for confirmed coronavirus disease 2019 (COVID-19) cases. Crowded travel settings, like airports, may increase chances of getting COVID-19, if there are other travelers with coronavirus infection. However, cases of coronavirus disease (COVID-19) have been reported in all states, and some areas are experiencing community spread of the disease. On April 14, 2020, the Centers for Disease Control and Prevention (CDC) announced the completion of the full-length genetic sequencing of the genome of the SARS-associated coronavirus.'

Answer key Generation

✓ [48] question = 'Where is coronavirus first detected'

```
contexts = cnt
answer = extract_answer(contexts, question)
print(answer)
```

wuhan, hubei province, china

✓ [50] question = 'how has COVID-19 affected Canada'

```
contexts = cnt
answer = extract_answer(contexts, question)
print(answer)
```

community spread of the disease

✓ [51] cnt

'Novel coronavirus (2019-nCoV) outbreak. The sequence data confirmed that SARS-CoV is a previously unrecognized coronavirus. The Centers for Disease Control and Prevention (CDC) is closely monitoring an outbreak of respiratory illness caused by a novel (new) coronavirus (named coronavirus disease 2019 or COVID-19) that was first detected in Wuhan, Hubei Province, China and which continues to expand. However, cases of coronavirus disease (COVID-19) have been reported in all states, and some areas are experiencing community spread of the disease. The sequence data confirmed that SARS-CoV is a previously unrecognized coronavirus. Results of laboratory testing at the Centers for Disease Control and Prevention (CDC) have provided additional evidence that a 27-year-old patient in Singapore was infected with SARS-associated coronavirus (SARS-CoV). Results of laboratory testing at the Centers for Disease Control and Prevention (CDC) have provided additional evidence that a 27-year-old patient in Singapore was infected with SARS-associated coronavirus (SARS-CoV). Reported illness...'

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