Text Summarization and Question Answering System for COVID-19 Data

- **102017070**

PART I

An unsupervised extractive text summarization system is developed that generates a summary for a given document.

First of all, the document is pre-processed by removing the stop-words. Summary is generated based on weights of sentences using different features.

The following features have been added to the system:

(a) cue phrases for COVID data

(b) numerical data

(c) Heading match (Number of words matching the heading)

(d) Sentence Length

(e) Sentence Position

(f) Word Frequency

(g) Upper case

(h) Proper Noun

Total score is calculated from all these features. And only those sentences are selected which have a score value above 1.2 times the average score (threshold).

```
The Malaria Vaccine Implementation Programme in Africa: clarifying misperceptions


Extensive testing in clinical trials has confirmed that the malaria vaccine, RTS,S/AS01, reduces malaria significantly, including life-threatening severe
malaria. The Malaria Vaccine Implementation Programme (MVIP) is now underway to (i) support the pilot implementation of RTS,S/AS01 by routine health
systems, and (ii) evaluate the routine use of the vaccine.
```
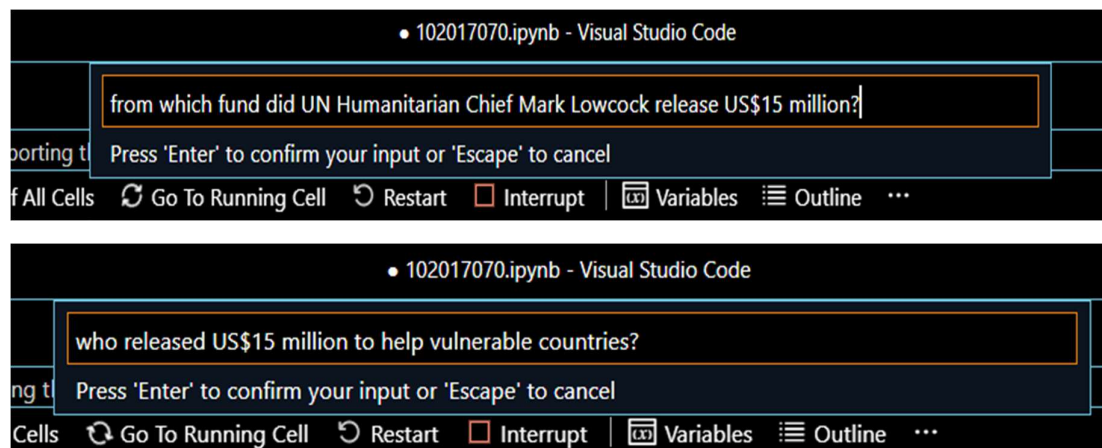
PART II

In this a question answering system is developed that can answer When/Where/Who type questions.

The user enters the question. The system uses sequencematcher to match pattern and calculate the ratio. It updates the max ratio and displays the sentence with the maximum ratio.

Questions:

Answers: