CS685 : DATA MINING PROJECT PRESENTATION

TITLE : SIMILAR QUESTION RANKING FROM Q&A SITES.

SUPERVISED BY: Dr. ARNAB BHATTACHARYA

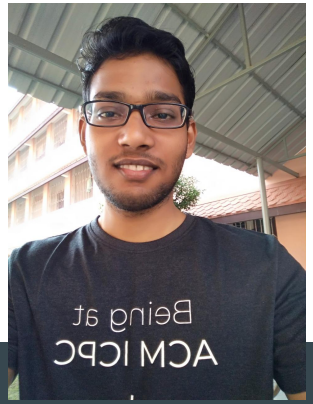
The Team : Group 23

AMEYA KARANJKAR

MTECH, CSE

20111011

ameya20@iitk.ac.in



TUSHAR GAUTAM

MTECH, CSE

20111071

tushark20@iitk.ac.in

JAY VORA

MTECH, CSE

20111023

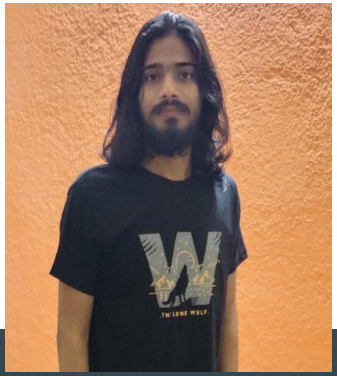
jvora20@iitk.ac.in

ANUP ROY

MS, CSE

20111403

puna20@iitk.ac.in



ABHAS KUMAR

MTECH, CSE

20111001

ababhas20@iitk.ac.in

PROBLEM STATEMENT AND DEFINITION

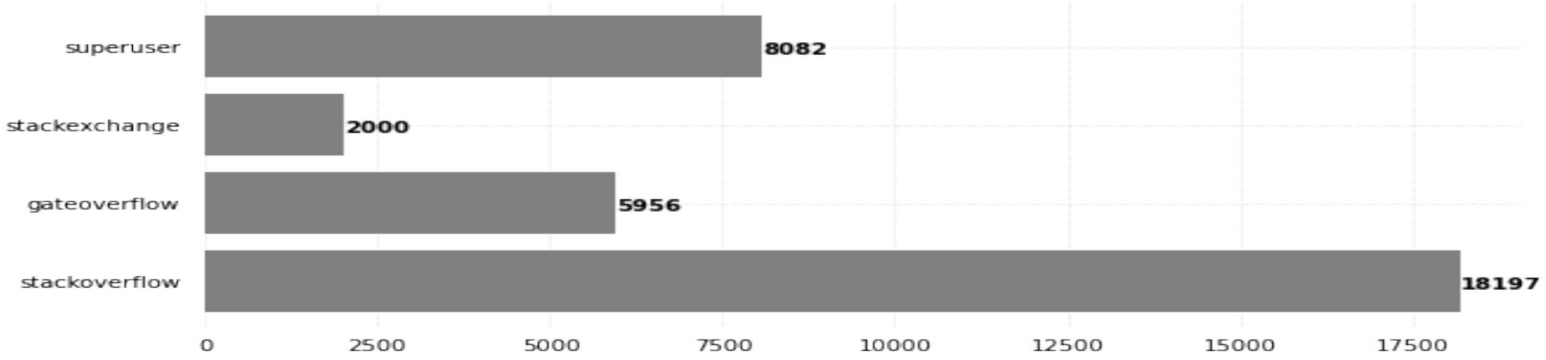
TO FIND SEMANTICALLY SIMILAR QUESTIONS TO GIVEN QUESTION AND RANK THEM.

Given a question q, Identify Questions in the corpus that are similar and related to this Question semantically.

DATASET USED

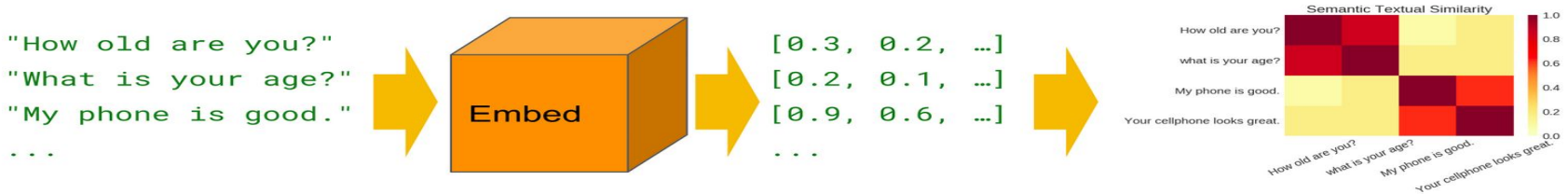
Source	Description
StackExchange	Network of various Q&A websites
StackOverflow	Features Q&A on topics in computer programming.
GateOverflow	Q&A on GATE Question
SuperUser	Q&A site for computer enthusiasts.

Table 1: Showing Sources of Datasets we have used.



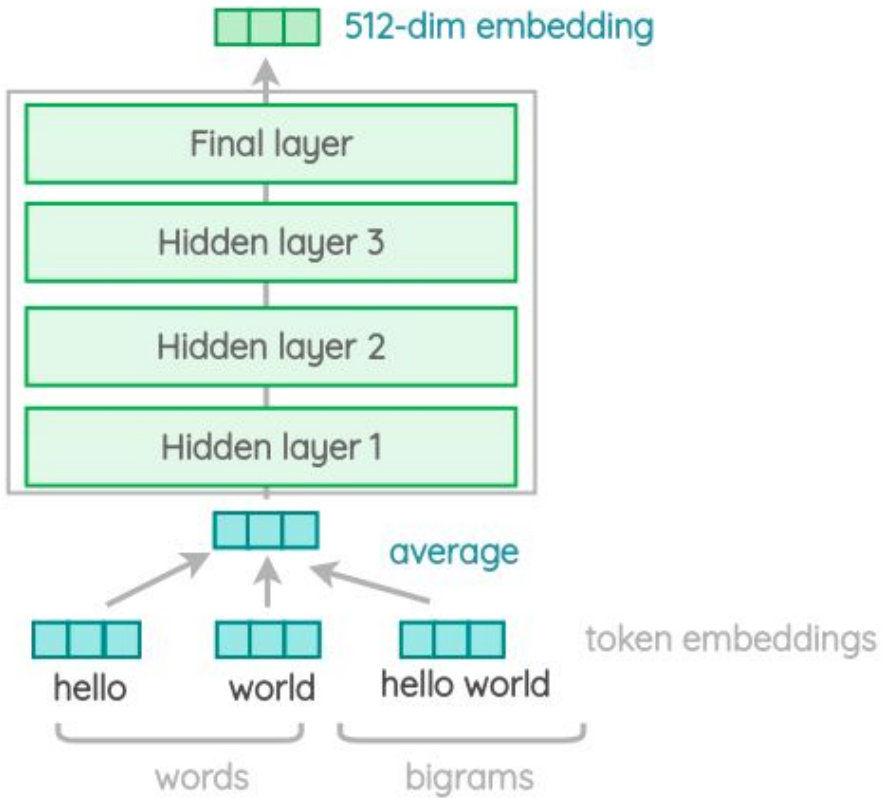
OVERALL IDEA ABOUT PROJECT

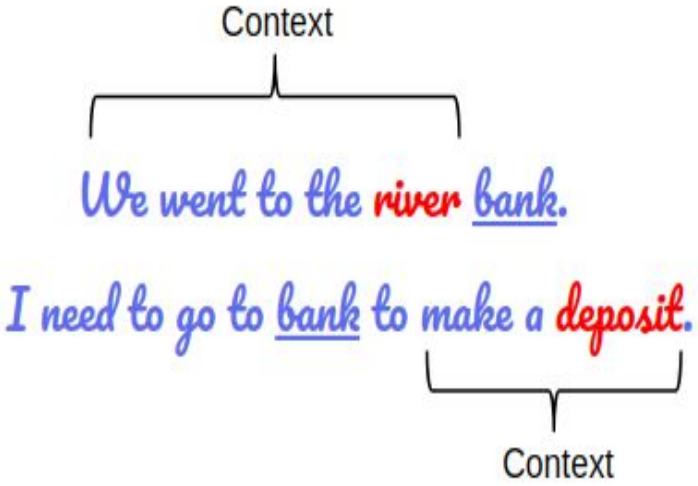
- WE WILL MEASURE THE SIMILARITY BETWEEN QUESTIONS BASED ON THE SEMANTIC.
- THIS TYPE OF SIMILARITY IS OFTEN COMPUTED BY FIRST EMBEDDING THE TWO TEXTS AND THEN COMPUTING THE COSINE SIMILARITY BETWEEN THEM.



Sentence similarity is typically calculated by first embedding the sentences and then taking the cosine similarity between them. (Source)

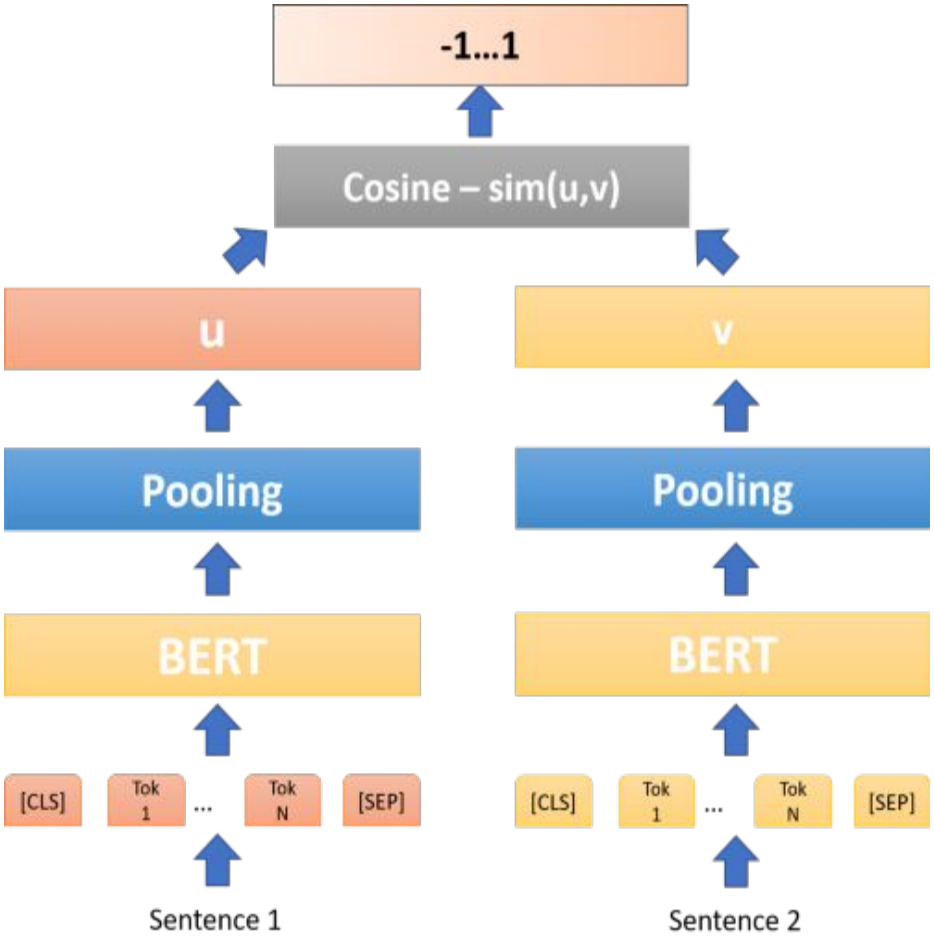
USE Model

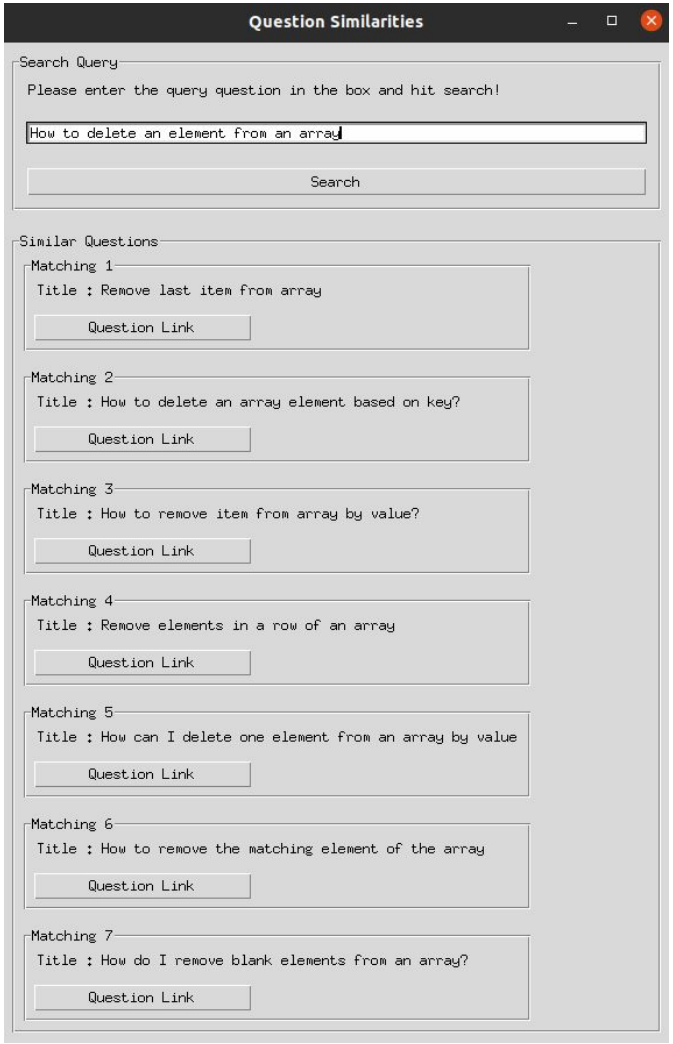

512-dim embedding
Final layer
Hidden layer 3
Hidden layer 2
Hidden layer 1
average
hello
world
hello world
token embeddings
words
bigrams


Context
We went to the river bank.
I need to go to bank to make a deposit.
Context

BERT Model

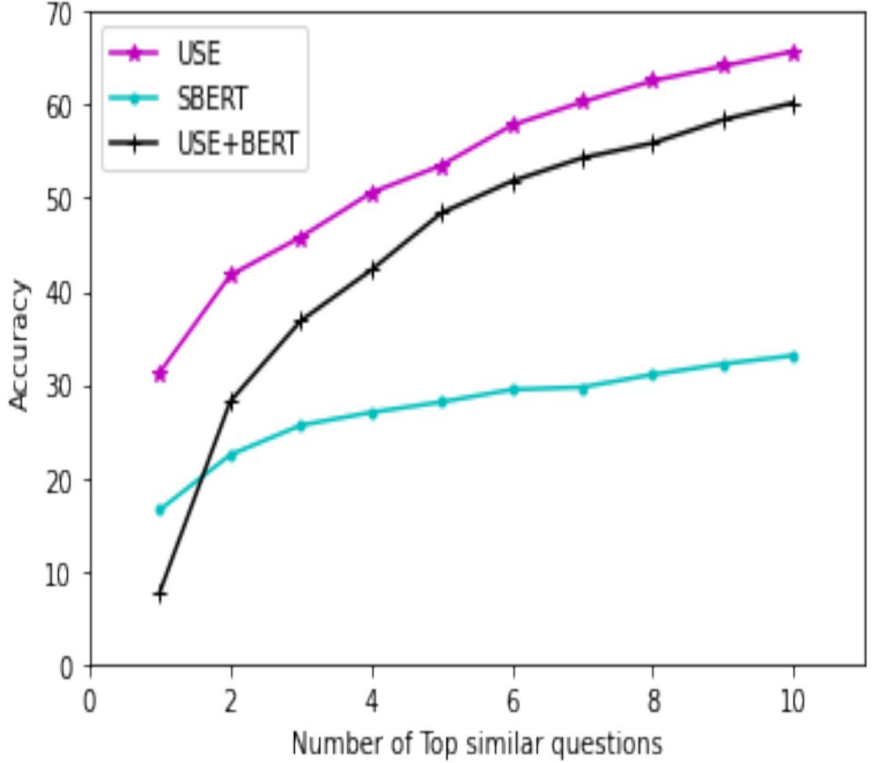
SBERT Model


-1...1
Cosine – sim(u,v)
u
v
Pooling
Pooling
BERT
BERT
[CLS]
Tok 1
...
Tok N
[SEP]
[CLS]
Tok 1
...
Tok N
[SEP]
Sentence 1
Sentence 2


Question Similarities
Search Query
Please enter the query question in the box and hit search!
How to delete an element from an array
Search
Similar Questions
Matching 1
Title : Remove last item from array
Question Link
Matching 2
Title : How to delete an array element based on key?
Question Link
Matching 3
Title : How to remove item from array by value?
Question Link
Matching 4
Title : Remove elements in a row of an array
Question Link
Matching 5
Title : How can I delete one element from an array by value
Question Link
Matching 6
Title : How to remove the matching element of the array
Question Link
Matching 7
Title : How do I remove blank elements from an array?
Question Link

RESULTS

ACCURACY


USE
SBERT
USE+BERT
Accuracy
Number of Top similar questions
0
10
20
30
40
50
60
70
0
2
4
6
8
10

CONCLUSIONS

- Predicted top 7 similar questions for any given user query with an accuracy of 60%.
- Learnt that there is a lot of data available freely which can be put to good use.
- Preservation of semantic meaning of the sentence is of vital importance.
- Trial and experimentation is important to get the technique that best fits the given task and dataset.
- Important to have a balance between performance and user experience.
- Appreciation for people who have developed state-of-the-art techniques.

THANK YOU!

ANY QUESTIONS?