**SEMANTIC SIMILARITY USING BERT**

A PROJECT REPORT

***Submitted by***

      BL.EN.U4CSE19056 J V M S S N KAPARDHI

      BL.EN.U4CSE19070 K SYAM SAI

      BL.EN.U4CSE19044 G S S VIGNESH

***in partial fulfillment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

IN

COMPUTER SCIENCE AND ENGINEERING



AMRITA SCHOOL OF ENGINEERING, BANGALORE

 AMRITA VISHWA VIDYAPEETHAM

BANGALORE 560 035

MAY – 2022

**AMRITA VISHWA VIDYAPEETHAM**

**AMRITA SCHOOL OF ENGINEERING, BANGALORE, 560035**



**BONAFIDE CERTIFICATE**

This is to certify that the project report entitled **“Semantic Similarity Using BERT”**submitted by

      BL.EN.U4CSE19056 J V M S S N KAPARDHI

      BL.EN.U4CSE19070 K SYAM SAI

      BL.EN.U4CSE19044 G S S VIGNESH

in partial fulfillment of the requirements for the award of the **Degree Bachelor of Technology** in “**COMPUTER SCIENCE** **AND** **ENGINEERING**” is a bonafide record of the work carried out under my guidance and supervision at Amrita School of Engineering, Bangalore.

*<signature> <signature>*

Ms. Kavitha.C.R Dr. Amudha J Dr. Sriram Devanathan

Asst. Professor Associate Professor Professor and Chairperson

Dept. of CSE Dept. of CSE Dept. of CSE

This project report was evaluated by us on 19/05/2022.

**ACKNOWLEDGEMENT**

The satisfaction that accompanies successful completion of any task would be incomplete without mention of people who made it possible, and whose constant encouragement and guidance have been source of inspiration throughout the course of this project work.

We offer our sincere pranams at the lotus feet of **”AMMA”,** **MATA AMRITANANDAMAYI DEVI** who showered her blessing upon us throughout the course of this project work.

We owe our gratitude to **Br. Viswamrita Chaitanya Swamiji**, Director, Amrita School of Engineering, Bangalore.

We thank **Dr. Sriram Devanathan,** Principal and Chairperson-CSE, Amrita School of Engineering, Bangalore for his support and inspiration.

It is a great pleasure to express our gratitude and indebtedness to our project guide **Ms. Kavitha.C.R /Asst Professor**, Department of Computer Science and Engineering, Amrita School of Engineering, Bangalore for her/his valuable guidance, encouragement, moral support, and affection throughout the project work.

We would like to thank express our gratitude to project panel members for their suggestions, encouragement, and moral support during the process of project work and all faculty members for their academic support. Finally, we are forever grateful to our parents, who have loved, supported and encouraged us in all our endeavors.

**ABSTRACT**

Question generation is the process of automatically generating different types of questions from the statement and given Keyword. First, we will take the input of the sentence and the keyword. we will try to find the synsets of the given keyword and then we will find the correct synset of the word for the sentence using the Bert WSD model. now, we tried to generate the question using the T5 transformer model .finally, for the determined synset of the keyword we will try to find the distractors by determining the co-hyponyms from the wordnet. We will output the question generated and distractors as options

**TABLE OF CONTENTS**

**Page no**

ACKNOWLEDGEMENT i

ABSTRACT ii

LIST OF FIGURES iv or v

LIST OF TABLES vi

CHAPTER 1- INTRODUCTION 1

CHAPTER 2 – LITERATURE SURVEY 2

CHAPTER 3 –SYSTEM DESIGN 3

CHAPTER 4 – DESIGN

4.1 SOFTWARE REQUIREMENTS 4

4.2 HARDWARE REQUIREMENTS 4

CHAPTER 5 – IMPLEMENTATION

5.1 CUSTOM DATA GENERATOR 5 5.2 MODEL BUILDING 6

CHAPTER 6 – SYSTEM TESTING 7

CHAPTER 7 – RESULTS AND ANALYSIS 8

CHAPTER 8 – CONCLUSION AND FUTURE SCOPE 9

REFERENCES 10

**LIST OF FIGURES**

* Fig. 3.1 – System design
* Fig. 5.1 – Custom data generator
* Fig. 5.2 – Model building
* Fig. 6.1 – System testing
* Fig. 7.1 – Results and Analysis
* Fig. 7.2 – Results and Analysis

**CHAPTER - 1**

**INTRODUCTION**

* 1. **INTRODUCTION TO QUESTION GENERATION**

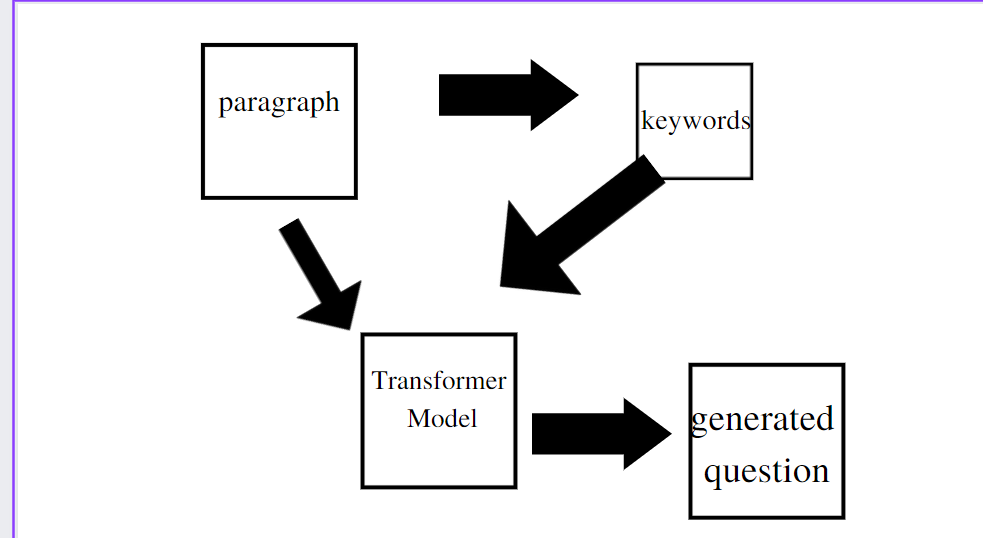
Question generation is the process of automatically generating different types of questions from statements. It involves using Bert, wordnet,t5transformer, etc

* 1. **MOTIVATION**

There are many potential applications we can make using NLP one such application is question generation, such as preparing exam papers without repeating questions and making work easy question generating. Ultimately, the goal is to provide a richer, more accessible experience for all users.

**CHAPTER – 2**

**SYSTEM DESIGN**

****

**Fig 3.1**

First a input is taken into model and keywords are generated and it is passed through t5 transformer model and question is generated and finally question is generated

**CHAPTER - 3**

**SYSTEM SPECIFICATIONS**

**4.1 Software requirements**

**SOFTWARE REQUIREMENTS & HARDWARE REQUIREMENTS:**

**4.1 Software Requirements:**

Python programming language,

Python Compiler,

Kaggle notebook

**Modules:**

pytorch

Numpy,

Pandas,

Bert

**4.2 Hardware Requirements:**

8 GB RAM,

GPU,

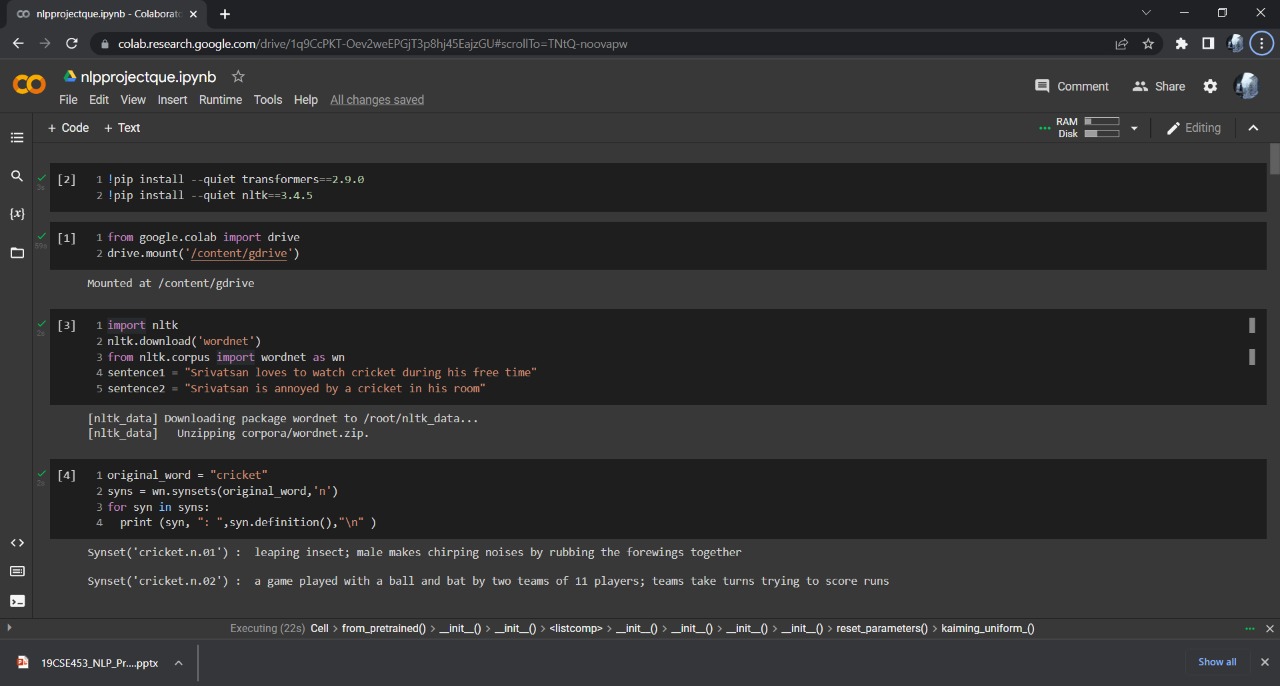
64-bit four-core,

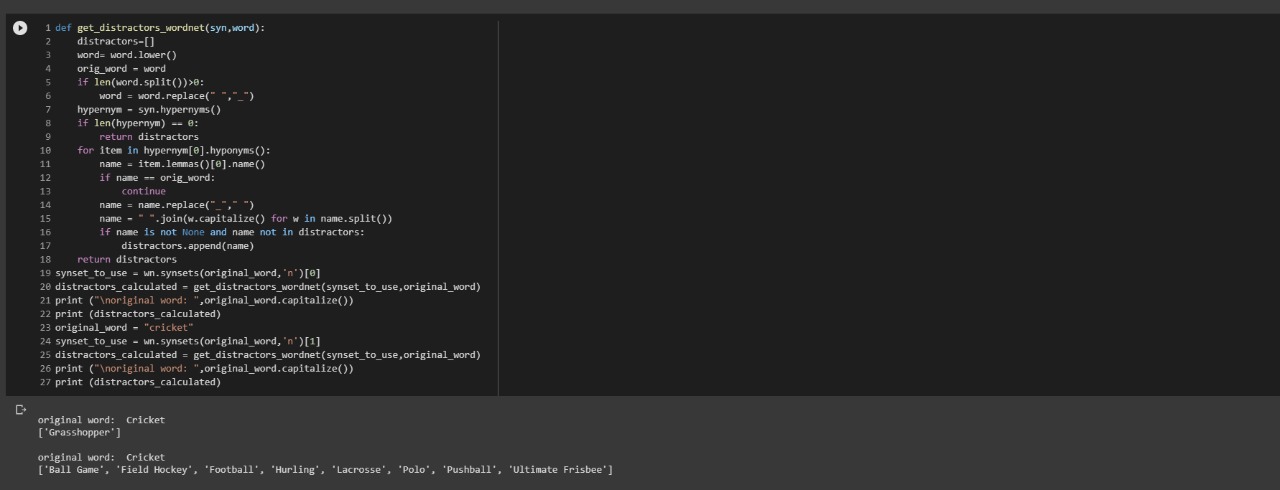
2.5 GHz core.

**CHAPTER – 4**

**SYSTEM IMPLEMENTATION**

**5.1:**

****

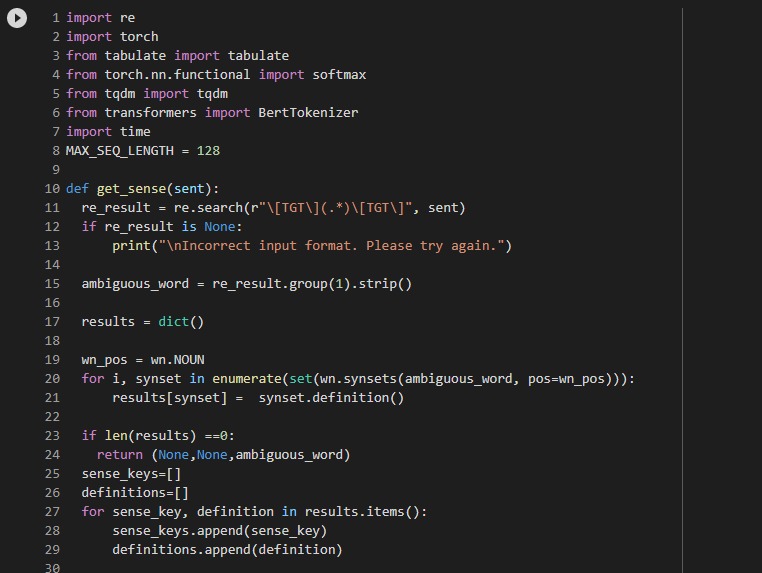
****

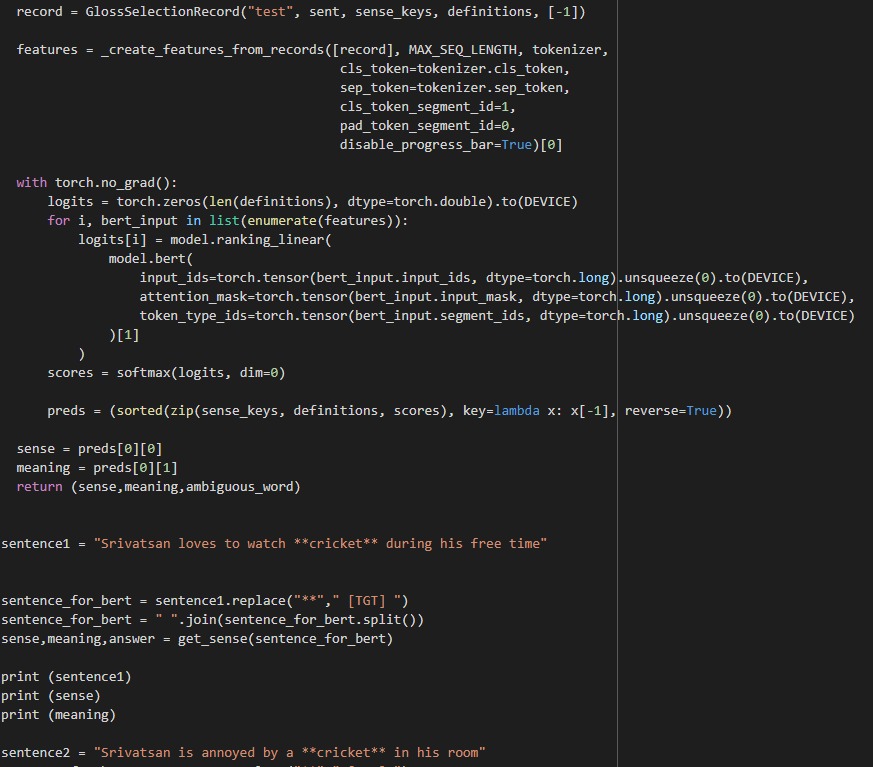
**Fig 5.1**

First we just imported the packages and taking a two different statements with a same word and different meaning we take the that word and we find the sys for that word.

And find the hypernms and hyponmes for the word for distracting the student.

**5.2 Model Building:**

****

****

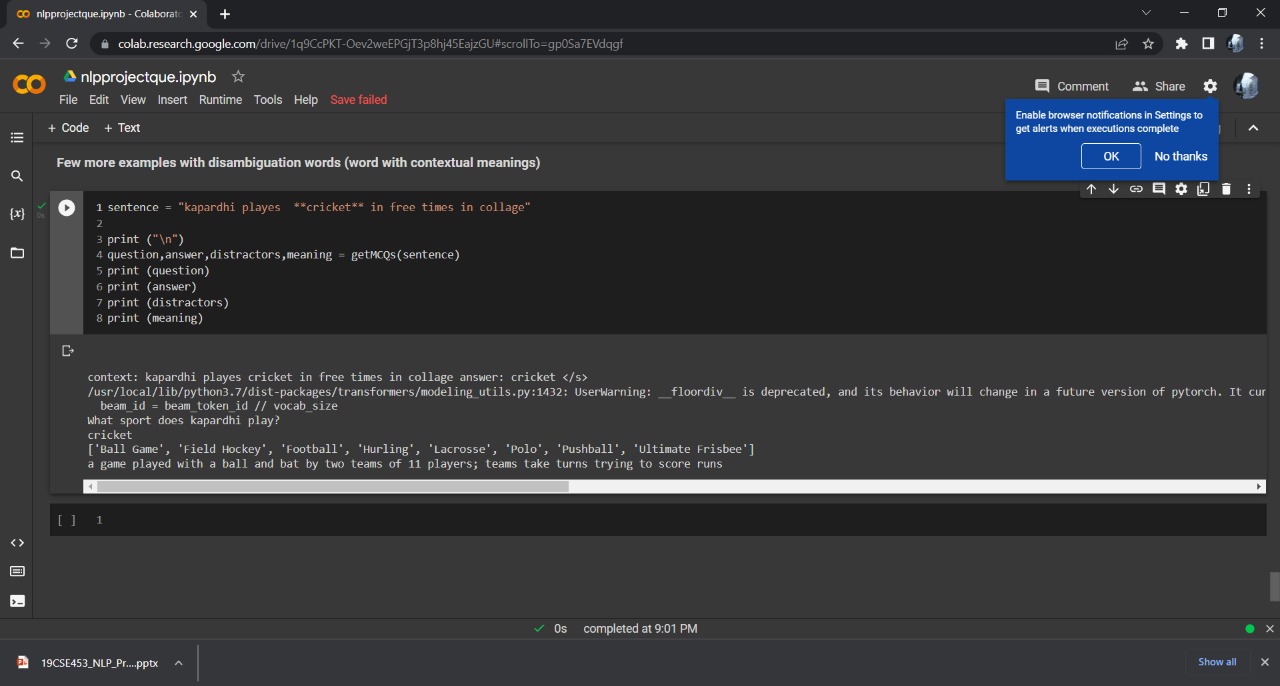
**Fig 5.2**

T5, or Text-to-Text Transfer Transformer, is a Transformer based architecture that uses a text-to-text approach. Every task – including translation, question answering, and classification – is cast as feeding the model text as input and training it to generate some target text.

BERT (Bidirectional Encoder Representations from Transformers) is a recent [paper](https://arxiv.org/pdf/1810.04805.pdf) published by researchers at Google AI Language. It has caused a stir in the Machine Learning community by presenting state-of-the-art results in a wide variety of NLP tasks, including Question Answering (SQuAD v1.1), Natural Language Inference (MNLI), and others.

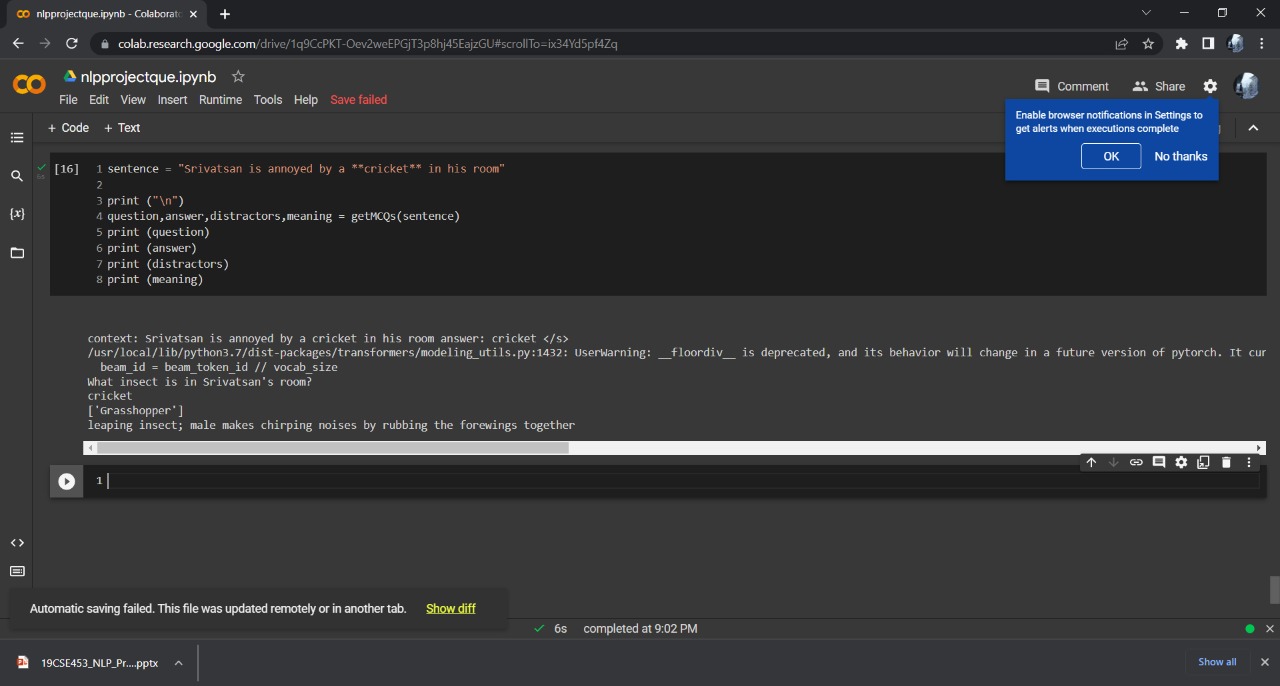
**CHAPTER - 5**

**SYSTEM TESTING**

****

**CHAPTER – 6**

**RESULTS AND ANALYSIS**

****

**CHAPTER – 7**

**CONCLUSION AND FUTURE SCOPE**

**8.1 Conclusion**

* Finally we can able to achieve question genation using bertwsd, wordnet,and some tokenixation techniques

**REFERENCES**

[1]https://www.irjet.net/archives/V7/i6/IRJET-V7I6848.pdf(priti Gumast1, Shreya Joshi, Srushtee Khadpekar, Shubhangi Mali)

[2] https://arxiv.org/abs/2111.06476 ,Cagatay Akyon