AGeES: AUTOMATIC MULTIPLE CHOICE QUESTION (MCQ) GENERATION FROM EXTRACTIVE SUMMARY OF VIDEO LECTURES USING BERTSUM

A PROJECT REPORT

Submitted by

TEAM 27

BARATH SRINIVASAN B (2019103008)

BHARATH M (2019103011)

LALIT ARVIND B (2019103612)

for the course

CS 6811 – PROJECT WORK

UNDER THE GUIDANCE OF

Ms. Karthika Devi M S, ASST. PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITY

CHENNAI 600 025

MAY 2023

ANNA UNIVERSITY : CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report titled "**AGeES: AUTOMATIC MULTIPLE CHOICE QUESTION (MCQ) GENERATION FROM EXTRACTIVE SUMMARY OF VIDEO LECTURES USING BERTSUM**" is the bonafide work of "**Barath Srinivasan B (2018103008), Bharath M (2018103011)** and **Lalit Arvind B (2018103612)**" who carried out the project work under my supervision, for the fulfillment of the requirements for the award of the degree of Bachelor of Engineering in Computer Science and Engineering. Certified further that to the best of my knowledge, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on these or any other candidates.

Place : Chennai

Date :

SIGNATURE

Dr. S. Valli

HEAD OF THE DEPARTMENT

Professor,
Computer Science and Engineering,
College of Engineering Guindy,
Anna University, Chennai - 600 025.

SIGNATURE

Ms. Karthika Devi M S

SUPERVISOR

Assistant Professor,
Computer Science and Engineering,
College of Engineering Guindy,
Anna University, Chennai - 600 025.

ABSTRACT

Multiple-choice questions (MCQs) from online lectures can benefit students in a number of ways. MCQs are a quick and accurate technique to evaluate knowledge and understanding since they can rapidly and accurately reinforce important topics discussed in the lectures. They are adaptable and time-effective, and they encourage students to study the content covered in class.

Existing research focuses on generating MCQs from textbook data which are typically well formatted and structural in the way the information is conveyed, unlike YouTube or any other educational videos where typically, the information is conveyed in an informal or casual manner without upholding the appropriate formal structures or grammar of the sentences. The present technologies for generating MCQs have a number of limitations which include the ability to create MCQs from only highly informative textual content, employing a singular approach for sentence selection, and inefficient keyword extraction for scientific data.

Our proposed work overcomes the challenges of generating efficient MCQs from less informative content retrieved from video lectures. To enhance the effectiveness of MCQ generation, our method proposes a dual approach for sentence selection and works well with generation for scientific data as well. In conclusion, our proposed work overcomes the structural differences that hinder the system’s ability to isolate factual information from the video. As a result, this methodology generates coherent MCQs by extracting transcripts from video lectures followed by a dual sentence selection approach.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BERT	Bidirectional Encoder Representations from Transformers
BERTSUM	BERT Extractive Summarizer
CBTE	Computer-Based Testing Examination
CFG	Context Free Grammar
FFmpeg	Fast Forward Moving Picture Experts Group
GLUE	General Language Understanding Evaluation
JSON	JavaScript Object Notation
LDA	Latent Dirichlet Allocation
MCQ	Multiple Choice Question
MCQG	Multiple Choice Question Generator
NLP	Natural Language Processing
NER	Named Entity Recognition
NLTK	Natural Language Tool Kit
OMR	Optical Mark Recognition
PMI	Pointwise Mutual Information
POS	Parts Of Speech
PTM	Parse Tree Matching
QNLI	Question-answering Natural Language Inference
QQP	Quick Queues Protocol
RAKE	Rapid Automatic Keyword Extraction

RDF	Resource Description Framework
RTE	Recognizing Textual Entailment
STS-B	Semantic Textual Similarity-Benchmark
TF-IDF	Term-Frequency Inverse Document Frequency
VQA	Visual Question Answering

CHAPTER 1 – INTRODUCTION

1.1 OVERALL OBJECTIVE

The overall objective of an automated MCQ generator system is to automatically generate a set of MCQs that test the knowledge of the user on a particular topic or subject. This system can be used in various educational and training settings to reduce the workload of educators and trainers and to provide students or trainees with a more interactive learning experience.

The generated MCQs can be used to test the knowledge of students or trainees in a variety of settings, such as online quizzes, practice exams, and assessments. The system can also be used to generate customized sets of questions based on the individual learning needs of students or trainees.

Overall, an automated MCQ generator system can save time and resources for educators and trainers, while providing students or trainees with a more engaging and interactive learning experience.

1.2 PROBLEM STATEMENT

Following the COVID-19-induced shift to virtual mode of class, educational institutes have started to shift towards online lectures. Although the offline mode of classes is in full flow presently, online lectures are still preferred due to the flexibility and the comfort that they provide to both teachers and students.

However, most of the students tend to procrastinate during the online lectures. This not just proves to be a waste of time for the faculties who invest their whole energy in teaching but also will bring down the quality of education over time.

Moreover, the faculties seldom get any kind of reception about the lectures, from the students. This deprives the faculties from the opportunity of improving their lectures.

1.3 ORGANIZATION OF THE THESIS

The outline of the thesis is as follows:

- **Chapter 2** iterates over the summary of other works that are similar in either purpose, approach or technology used and that have already been implemented.

- **Chapter 3** details the architecture of the system. This chapter also includes the input and expected output along with the module wise explanation and the pseudocode.

- **Chapter 4** discusses the evaluation metrics used as well as the results obtained. Various test cases are tabulated along with the expected and actual outputs. It also includes comparative analysis to determine the model’s robustness towards achieving the objective.

- **Chapter 5** concludes the thesis report by providing an overall gist of the project and giving suggestions for the future works.

CHAPTER 2 – SUMMARY OF RELATED WORK

The project aims to develop an automatic Multiple Choice Question generator based on video lectures. Many research papers have dealt with the topic of automated MCQ generation, sentence selection, distractor generation and question formation.

Chen Liang et al. [1] proposed an adversarial training framework for Distractor Generation consisting of two components: a generator G and a discriminator D. G is a generative model captures the conditional probability of generating distractors. D is a discriminative model that finds the probability of a distractor sample coming from the real training data. A cascaded learning framework is used to increase performance. The model, however, lacks an appropriate interface for users to interact with.

A pipeline for automatic generation of MCQs from textbooks of middle school level subjects was discussed by CH et al. [2], and the pipeline is partially subject-independent. The proposed pipeline comprises four core modules: preprocessing, sentence selection, key selection, and distractor generation. Techniques employed to implement include sentence simplification, syntactic and semantic processing of the sentences, entity recognition, semantic relationship extraction among entities, and WordNet. The proposed system is capable of generating quality questions that could be useful in a real examination. Despite this, the system faces issues with complex questions dealing with multi-line facts.

Ma et al. [3] proposed a general model of extractive and abstractive for text summarization, which is based on BERT's powerful architecture and additional topic embedding information to guide contextual information capture. The combination of token embedding, segment embedding, position embedding, and topic embedding can more abundantly embed the information that the original text should contain. The two-stage extractive–abstractive model shares information and generates salient summaries, which reduces a certain degree of redundancy. The analysis shows that the model can generate high-quality summaries with outstanding consistency for the original text. However, the model has limited processing power for long articles with multiple topics.

An NLP-based system for automatic MCQG for Computer-Based Testing Examination (CBTE) was presented by Nwafor et al. [4]. NLP technique is used to extract keywords that are important words in a given material. To validate that the system is not perverse, five lesson materials were used. The manually extracted keywords by the teachers were compared to the auto-generated keywords. The result shows that the system was capable of extracting keywords from lesson materials in setting examinable questions. However, MCQs about facts explicitly based on keywords did not prove to be efficient for exams.

Mukta Majumder et al. [5] form MCQs by performing Parse Tree Matching with test sentences, employing topic modeling to filter the sentences according to topics. NER is used to identify the keywords and gazetteer lists to generate distractors. The system has a tendency to exclude time and date-related information and selects incomplete sentences.

Xian Wu et al. [6] leverages a contextual encoder to generate semantic representations for all text materials. It then uses the attention mechanism to enrich the context of the question and the correct answer. Two modules are introduced by leveraging the gate layer to guarantee incorrectness, the Reforming Question Module and the Reforming Passage Module. Beam search is then used to generate several diverse distractors by controlling their distances. Talking about limitations, the model is incapable of generating distractors requiring multi-sentence/hop reasoning.

Dhanya et al. [7] proposed a system that involved four major processes - sentence selection, key selection, question formation and distractor generation. The system uses Google T5's fundamental principle to reframe all NLP tasks as sequence-to-sequence tasks. Text extraction is done using tesseract, followed by context recognition, then Google T5 model for question generation. Finally, Sense2Vec is used for distractor generation. This method provides quality work and reduces the intervene of humans in the process thereby automating it. However, the percentage of the relevance of auto-generated incorrect options were low.

Mehta et al. [8] proposed a system that creates automated questions with the help of NLP that reduces human intervention and it is a cost and time-effective system. The system is based on Google's BERT Model, the accuracy of the system will increase in the future as the performance of the model is improved. For the summarizer, BERTSUM model has been used. BERTSUM with Transformer has managed to achieve the best performance on all the three metrics that are considered. For generating distractors, the wordnet approach is used. The quality of wordnet is evaluated by examining the validity of its constituent synonyms and its hypernym-hyponym pairs. However, the failure to validate a synset pair is not a definitive indicator of erroneous construction and hence the method is not effective in all cases.

An approach that takes a text as an input, which will be paraphrased before the MCQs are generated from it, was proposed by Maniar et al. [9]. A question paper consisting of a set of random questions will be generated. For paraphrasing the input text, transformers are used. The tuner007 pegasus_paraphrase model which is an NLP Model implemented in the Transformer library, is used. The output of the paraphraser model will be given as input to the MCQ generation model. However, this model may generate multiple MCQs from the same line if the input text has a number of sentences less than the number of questions desired.

A mixed similarity strategy was presented by Ming Liu et al. [10] to generate Chinese multiple-choice distractors. A statistical regression model was used which considers the similarity between the distractor and the target character in appearance, pronunciation and semantic meanings. An annotation schema was established for rating the similarity of characters. It was found that the proposed strategy outperforms the three common distractor generation strategies in terms of distractor usefulness and power. Unfortunately, this system faces a problem with the extraction of semantic distance feature of the characters which are unavailable in the knowledge base.

Dmytro Kalpakchi et al. [11] have fine-tuned a pre-trained BERT2 for Swedish for distractor generation. Two linear layers with layer normalization and a softmax activation layer had been added on top of BERT2. They have also proposed an effective method to evaluate the generated MCQs. However, only half of the distractors generated using the model were plausible.

Animesh Srivastava et al. [12] proposed a state-of-the-art solution consisting of a pipeline that makes use of natural language processing and image captioning techniques. The proposed solution generates questions for both textual and visual inputs. In addition to this, their answers and the distractors are created. Although the system yields outstanding results, the captioning dataset can be further improved which in turn will improve the question generation model.

An ontology-based approach was introduced by Selvia Ferdiana Kusuma et al. [13] for the purpose of automating the process of question generation. A total of 11 categories of questions were automatically generated. All the information regarding the ontology is broken down into categories in the form of SPARQL queries which are further converted into questions. Unfortunately, the proposed method is heavily dependent upon the completeness of ontology information.

Jiaying Lu et al. [14] proposes a reinforcement learning based framework called GOBBET. The framework makes use of pre-trained Visual Question Answering models as an alternative knowledge base to guide the distractor generation process. The performance degradation of existing VQA models is utilized for detecting the quality of generated distractors. The utility of the distractors that are generated is exhibited through data augmentation experiments. The sparsity of training samples, however, proves to be a major challenge to the framework.

Ainuddin Faizan et al. [15]'s approach uses semantic annotation to find named entities in the slide content and utilizes property information of the entities to generate questions and find appropriate distractors using SPARQL queries. SPARQL is also used to retrieve further information about the entities in the form of RDF triples, which is then verbalized to form the question text. The model has instances where the resource is not identified and inaccurate distractors are produced.

CHAPTER 3 – SYSTEM DESIGN

3.1 ARCHITECTURE DIAGRAM

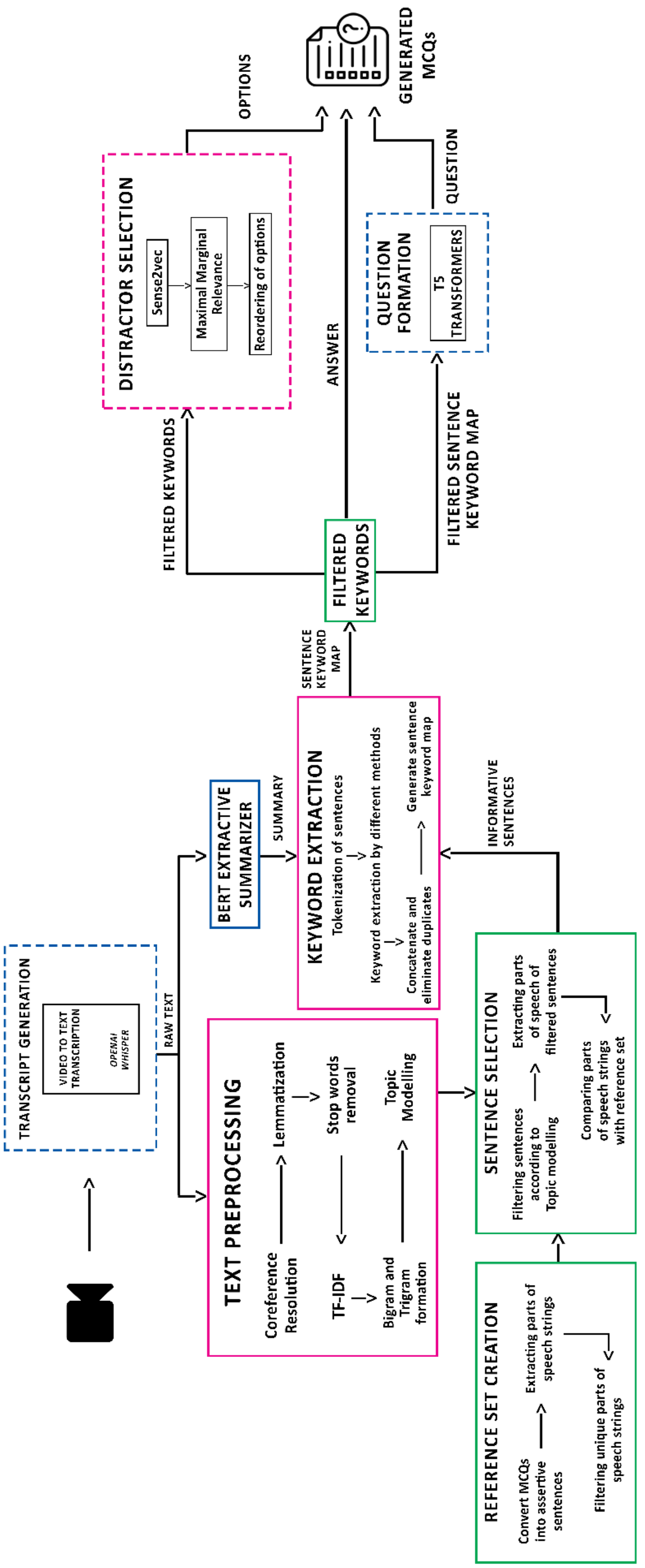


Fig 3.1. Architecture Diagram

3.2 DETAILS OF MODULE DESIGN

The list of modules involved in the entire process are as follows:

- Transcript Generation
- BERT Extractive Summarizer
- Text preprocessing
- Keyword Extraction
- Reference Set Creation
- Sentence Selection
- Distractor Selection
- Question Formation

3.2.1 TRANSCRIPT GENERATION

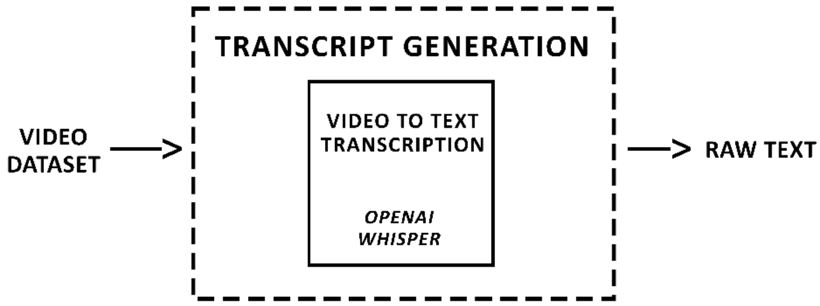


Fig 3.2.1. Transcript Generator

This module deals with extraction of audio from the video files using FFmpeg software, which is an open-source software project used for handling several multimedia files. Transcripts are generated for the extracted audio using OPENAI Whisper, which is an automated speech recognition model trained on a large dataset.

Input: Video Dataset

Output: Transcripts Generated

ALGORITHM

1. *Import all the necessary libraries.*
2. *Load video dataset from google drive.*

3. *Install and import OpenAI Whisper.*
4. *Define a function to generate a transcript of a video using Whisper.*
5. *For each video in the video dataset:*
    1. *Generate transcript for the video.*
    2. *Save the transcript as text file in the google drive*

3.2.2. BERT EXTRACTIVE SUMMARIZER

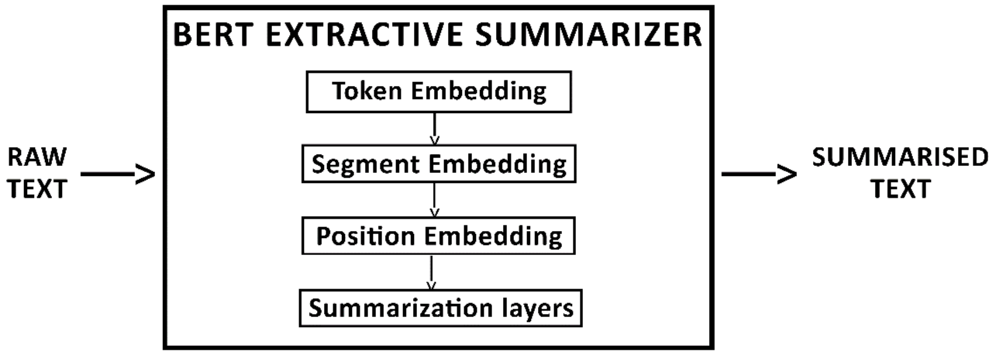


Fig 3.2.2 BERT Extractive Summarizer

In BertSum Model, the summarized text is generated for the text transcript. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained open-sourced model which helps computers to understand the language a bit more as humans do. The input text is summarized using the BERTSUM model, which is fine-tuned BERT for extractive summarization.

Input: Text Transcript

Output: Summarized Text

ALGORITHM

1. *Install and import the BERT extractive summarizer.*
2. *Create the summarizer model.*
3. *Load the generated transcript files.*
4. *Open each transcript file in read mode and save it in a temporary variable.*
5. *Using the summarizer model generates a summary for each text file.*

3.2.3. TEXT PREPROCESSING

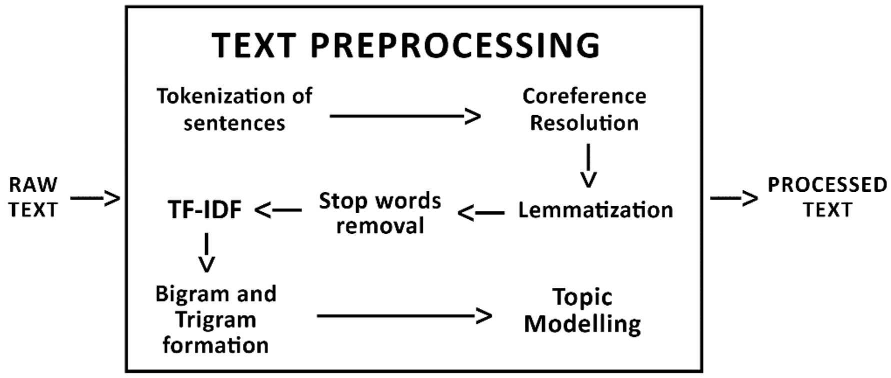


Fig 3.2.3. Text preprocessor

The text pre-processing step starts with the tokenization of the raw text. We then perform dependency parsing, which performs grammatical structure analysis of a sentence and detects related words along with the type of the relationship between them. This is followed by the process of coreference resolution, which is referred to as the process of finding all the expressions referring to the same entity in a text. The tokens are then tagged according to their Parts-Of-Speech and lemmatization is performed. After removal of stop words, bigrams and trigrams are formed. Finally, topic modelling which discovers the abstract topics inside a collection of documents, is performed.

Input: Raw text (Transcript of audio files)

Output: Processed text

ALGORITHM

1. *Import and install Allennlp Coreference model.*
2. *For each transcript file:*
    1. *Extract coreference clusters in transcript.*
    2. *For each cluster found:*
        1. *Remove redundant clusters.*
        2. *Check for cataphora problem.*
        3. *Resolve nested coherent mentions.*
3. *For each mention in the mapping:*

1. *Replace the span with its first mention in the transcript to finally form a resolved text.*

3.2.4. REFERENCE SET CREATION

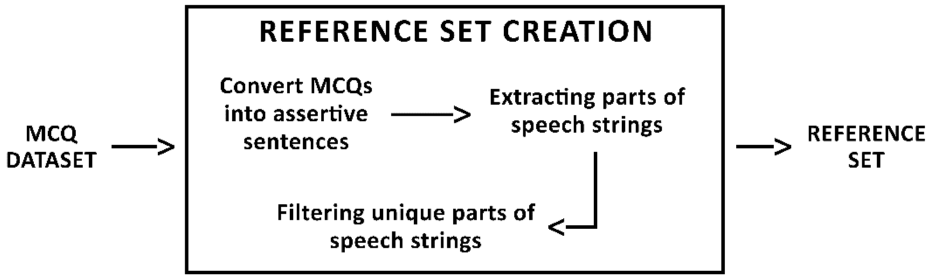


Fig 3.2.4. Reference Set Creator

This module deals with the creation of reference set. The multiple-choice questions present in a dataset are converted into assertive sentences. Parse trees are extracted for the assertive sentences after which unique parse trees are filtered to form the reference set.

Input: MCQ Dataset

Output: Reference set for sentence selection

ALGORITHM

1. *Import MCQ question dataset.*
2. *Extract questions and choices from json file.*
3. *For each question:*
    1. *Extract the parse tree for each MCQ question.*
    2. *Using the parse tree and answer, generate assertive sentences.*
    3. *Extract parts of speech string from each assertive sentence.*
4. *For each extracted parts of speech string:*
    1. *Compare it with other extracted parts of speech strings to find similarity.*
    2. *If the strings are similar:*

1. *Keep deleting the other parts of speech strings.*

5. *A set of unique parts of speech strings is obtained as a result.*

3.2.5. SENTENCE SELECTION

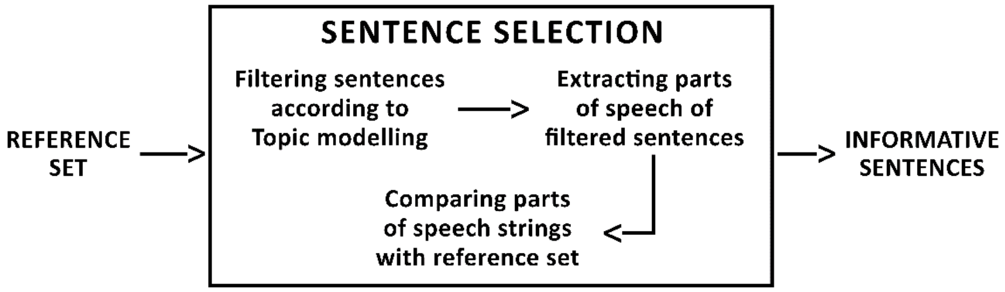


Fig 3.2.5. Sentence Selector

The Processed text might contain several sentences that are not informative (i.e., sentences that cannot be utilized for question formation). In this module, we filter out the informative sentences from the text by comparing the sentences to a reference set containing MCQ questions from various sources.

Sentences are first filtered out according to topic modelling and parts of speech strings for the filtered sentences are extracted. These parts of speech strings are then compared with the reference set to provide informative sentences as output.

Input: Processed text and Reference set

Output: Informative Sentences

ALGORITHM

1. *Filter the sentences containing the topic words extracted from LDA model.*
2. *For each filtered sentence:*
    1. *Generate Parts of speech string for each filtered sentence.*
    2. *For each parts of speech string in reference set:*
        1. *If POS of reference set is present as a substring in the POS of filtered sentence, then we consider the filtered sentence to be valid.*

2. *Extract and save the filtered sentence.*

3.2.6. KEYWORD EXTRACTION

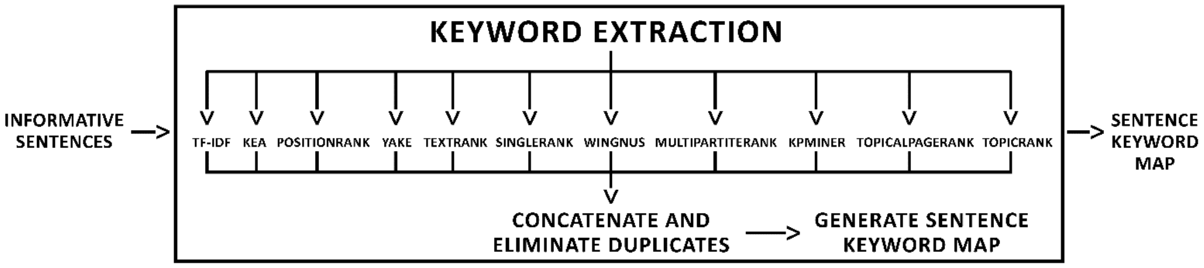


Fig 3.2.6. Keyword Extractor

A total of 11 keyword extraction methods are applied on the informative sentences. Following this, all the keywords are concatenated and the duplicate keywords are eliminated. The unique set of keywords is then used to generate the sentence keyword map.

Input: Informative Sentences

Output: Sentence Keyword Map

ALGORITHM

1. *Import all the necessary libraries.*
2. *Tokenize all the sentences present.*
3. *Perform keyword extraction by multiple methods.*
4. *Concatenate the keywords and then remove the duplicate keywords.*
5. *Generate a sentence keyword map.*

3.2.7. DISTRACTOR SELECTION

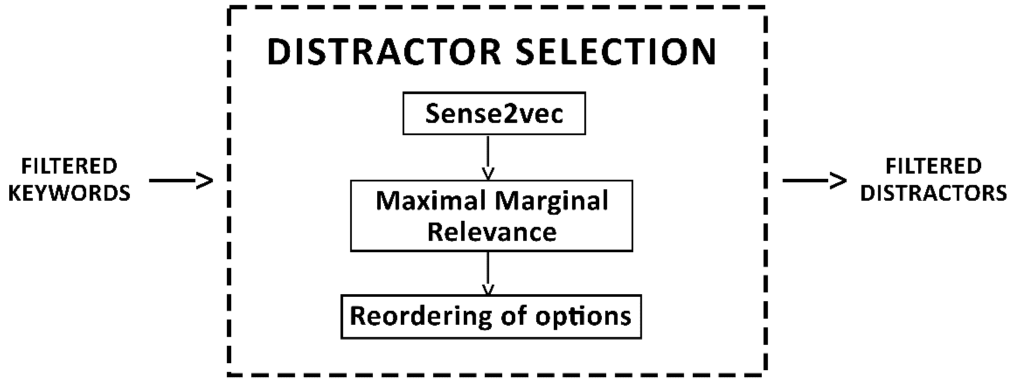


Fig 3.2.7. Distractor Selector

Distractors are basically the wrong answers that are provided in order to confuse the student. The distractors need to be as similar as possible to the keywords. In order to generate keywords, we use the sense2vec model. The sense2vec model provides a number of words which are good enough to become distractors. However, we make use of Maximal Marginal Relevance to narrow those words to 3 best words and thus get the distractors for a given question.

Input: Filtered Keywords

Output: Filtered Distractors

ALGORITHM

1. *Import and install sentence transformers.*
2. *Use Sense2vec get_best_sense function to find the meaning of the keyword.*
3. *Generate a list of words and phrases that are semantically similar to the keyword.*
4. *Use Maximal Marginal Relevance to rank the keywords and pick the top 3 values as our options.*
5. *Reorder the options.*

3.2.8. QUESTION FORMATION

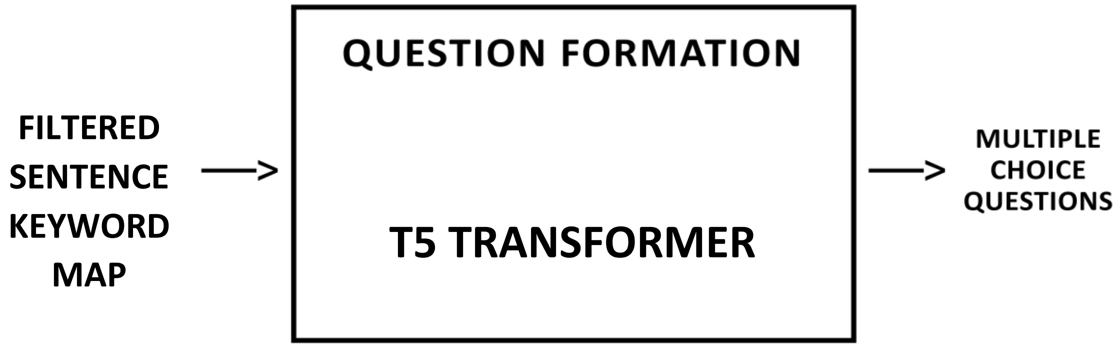


Fig 3.2.8. Question Former

The informative sentences are framed as questions using T5 transformer. Once the questions are formed, they are combined with the correct answer and distractors to produce multiple choice questions as output.

Input: Filtered Sentence Keyword Map

Output: Multiple Choice Questions

ALGORITHM

1. *Install T5 question generation model.*
2. *Load the sentence keyword map.*
3. *Sentence is used as the context and keyword as the answer for model.*
4. *Questions are generated.*

CHAPTER 4 – RESULTS AND DISCUSSION

4.1 PERFORMANCE METRICS

1. **GLUE (General Language Understanding Evaluation):** The GLUE benchmark is a collection of nine natural language processing tasks that are designed to measure the performance of models on a wide range of language understanding tasks.

   The GLUE evaluation metric for the BERTSUM model is the same as for other NLP models, and it is based on the average performance across all nine tasks in the benchmark. The GLUE score is a single number that ranges from 0 to 100, with higher scores indicating better performance.

   The formula for calculating the GLUE score is:

   GLUE score = (average task score - average baseline score) / (average best score - average - *Equation 1*

   where the average task score is the average score of the BERTSUM model on the nine tasks in the GLUE benchmark, the average baseline score is the average score of a set of baseline models on the same tasks, and the average best score is the average score of the best-performing models on the tasks.

Table 4.1. GLUE Test Results

MNLI	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE
82.2	88.5	89.2	91.3	51.3	85.8	87.5	59.9

2. **Distractor Closeness:** Distractor closeness is a measure of how similar the distractors (the incorrect answer options) are to the correct answer in a multiple-choice question. Its value ranges from 0 to 1. A higher value indicates that the

distractors are closer to the correct answer while a lower value indicates that the distractors have very little similarity to the correct answer.

$$\boxed{\text{Distractor Closeness} = 1 - (\text{Minimum Distance} / \text{Maximum Distance})} \quad - \quad \textit{Equation 2}$$

where minimum distance is the smallest distance between the correct answer and any of the distractors and maximum distance is the largest distance between the correct answer and any of the distractors.

3. **Coherence score:** The coherence score is a way of measuring the coherence or the degree of connectedness between topics generated by the model. This is useful for selecting the optimal number of topics. One popular coherence score for LDA models is the coherence score based on the Pointwise Mutual Information (PMI) measure.

   The formula for PMI coherence score of an LDA model is:

$$\boxed{\text{Coherence} = (1 / k) * \text{sum}(\text{PMI}(w\_i, w\_j))} \quad - \quad \textit{Equation 3}$$

   where k is the number of topics, PMI(w_i, w_j) is the PMI score for word pair (w_i, w_j) in the topic, and the sum is taken over the N most probable word pairs for each topic.

   As portrayed in Fig 27,

   Optimal number of topics ≈ 7, since the coherence score peaks at that value.

4. **Semantic Similarity:** Semantic similarity is a measure of how similar two pieces of text are in terms of meaning. The similarity between two documents can then be computed using the cosine similarity measure.

   The formula for cosine similarity between two vectors A and B is:

$$\boxed{\text{cosine\_similarity}(A, B) = (A . B) / (||A|| * ||B||)} \quad - \quad \textit{Equation 4}$$

   where A. B is the dot product of the two vectors A and B, and ||A|| and ||B|| are the lengths of the vectors A and B, respectively.

5. **Maximal Marginal Relevance:** Maximal Marginal Relevance (MMR) is a method which aims at retrieving a set of documents that are both relevant to the query and diverse in their content. In MMR, the relevance of a document is measured by its similarity to the query, while the diversity is measured by the dissimilarity between the selected documents.

$$\boxed{\mathbf{MMR\ (d) = \lambda * Sim(d, Q) - (1\text{-}\lambda) * max\ \{Sim(d, d')\}}} \quad \text{- } \textit{Equation 5}$$

where d → document being scored, Q → query, D_r → the set of already selected documents, Sim(d, Q) → the similarity between document d and query Q, and max {Sim(d, d')} is the maximum similarity between document d and any other document in the set of selected documents D_r. λ is a parameter that controls the trade-off between relevance and diversity.

4.2 COMPARATIVE ANALYSIS

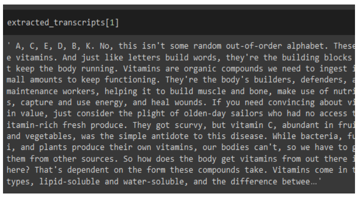
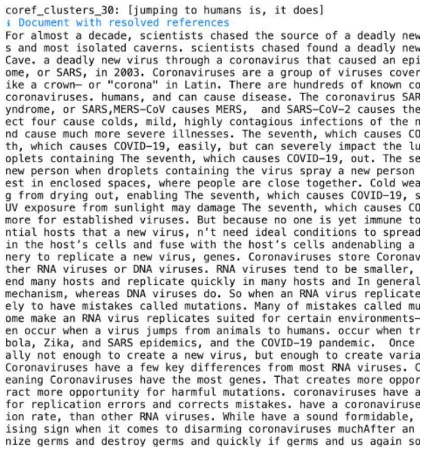
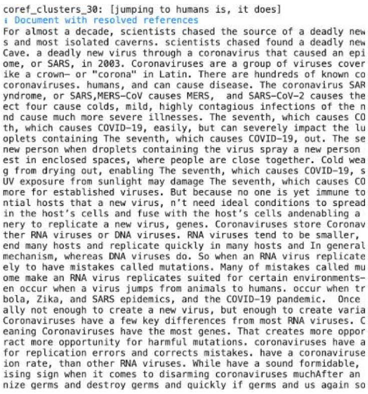
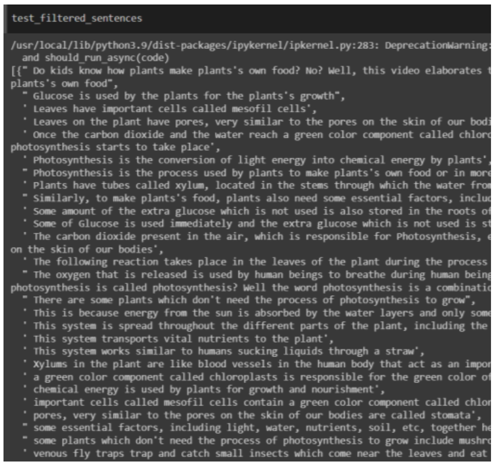
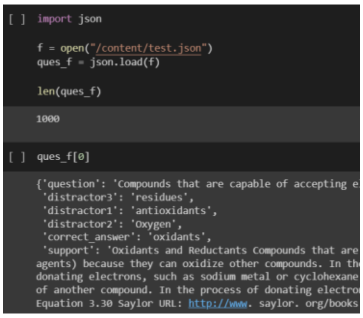
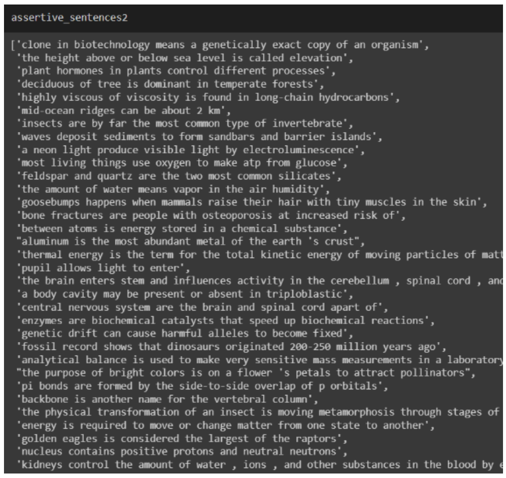
[2] has utilized the Parts of Speech tree comparison method which proves lacking in the case of video lecture input where the syntax and grammar are not perfect and the speech is informal. The parse tree method fails to select sentences that contain information within. AGeES uptakes a more flexible approach of creating the Parts of Speech strings of the reference set and the input transcript and checking if any POS string of the reference set is contained within the sentence from the transcript which are then selected for further processing.

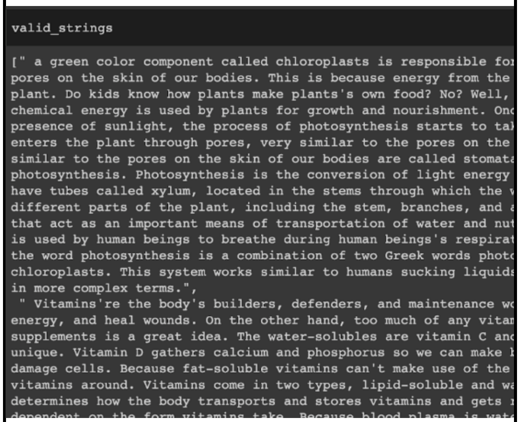
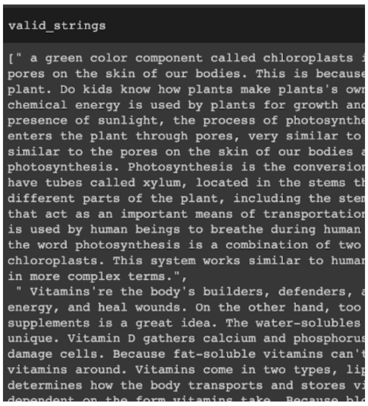
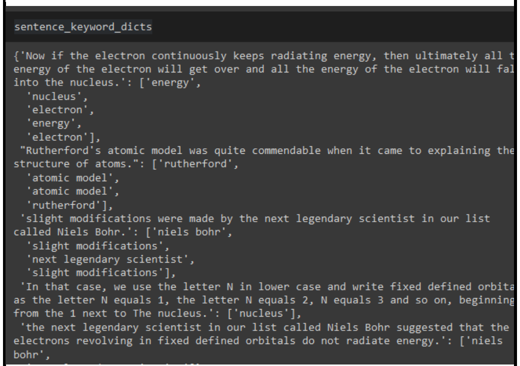
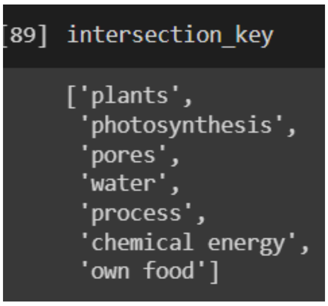
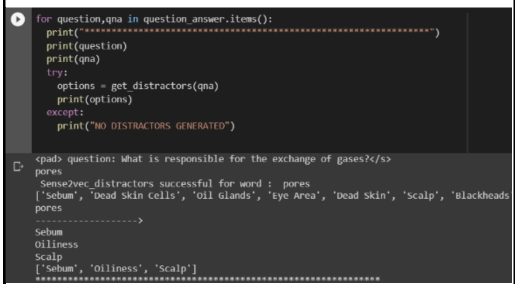
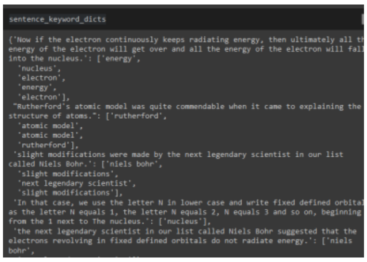
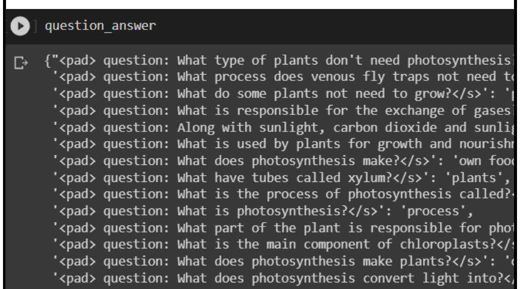
As stated in [2], straight forward strategies are used to perform keyword extraction. However, AGeES uses a total of 11 methods to perform keyword extraction. The given method performs better as compared to other keyword extraction methods like RAKE and NER. While NER fails to identify domain-specific words and RAKE gives long phrases as keywords on odd occasions, the proposed method ensures that a higher number of keywords is yielded without compromising on its quality.

Text pre-processing is an important step in MCQ generation systems, since this step will decide the efficiency and effectiveness of the whole system. Contrary to most of the MCQ generation systems, the proposed system intends to include techniques like coreference resolution (as seen in [5]) and topic modelling in text pre-processing. Coreference resolution helps in obtaining unambiguous sentences that are easily understood by computers and in converting complex sentences into simpler ones. Topic modelling facilitates in discarding unimportant sentences by detecting whether a sentence comes under any specific topic. With the above pre-processing steps, the processed text will be in such a state that the forthcoming steps can be carried out with more ease and accuracy.

4.3 TEST RESULTS

Test Case ID	**Test Case Description**	**Input Data**	**Expected Output**	**Actual Output**
1.	Generating transcripts for video lectures.	['1676963736', 'mp4'] ['1676963736(3)', 'mp4'] ['1676963736(1)', 'mp4'] ['1676963736(4)', 'mp4'] ['1676963736(5)', 'mp4'] ['1676963736(8)', 'mp4'] ['1676963736(2)', 'mp4'] ['1676963736(9)', 'mp4'] ['1676963736(6)', 'mp4'] ['1676963736(10)', 'mp4'] Video Dataset	Transcript Files	extracted_transcripts[1] ' A, C, E, D, B, K. No, this isn't some random out-of-order alphabet. These e vitamins. And just like letters build words, they're the building blocks t keep the body running. Vitamins are organic compounds we need to ingest i mall amounts to keep functioning. They're the body's builders, defenders, a maintenance workers, helping it to build muscle and bone, make use of nutri s, capture and use energy, and heal wounds. If you need convincing about vi in value, just consider the plight of olden-day sailors who had no access t itamin-rich fresh produce. They got scurvy, but vitamin C, abundant in frui and vegetables, was the simple antidote to this disease. While bacteria, fu i, and plants produce their own vitamins, our bodies can't, so we have to g them from other sources. So how does the body get vitamins from out there i here? That's dependent on the form these compounds take. Vitamins come in t types, lipid-soluble and water-soluble, and the difference betwee…'
2.	Generating summaries using BERT Extractive Summarizer.	extracted_transcripts[1] ' A, C, E, D, B, K. No, this isn't some random out-of-order alphabet. Thes e vitamins. And just like letters build words, they're the building blocks t keep the body running. Vitamins are organic compounds we need to ingest mall amounts to keep functioning. They're the body's builders, defenders, maintenance workers, helping it to build muscle and bone, make use of nutr s, capture and use energy, and heal wounds. If you need convincing about v in value, just consider the plight of olden-day sailors who had no access itamin-rich fresh produce. They got scurvy, but vitamin C, abundant in fru and vegetables, was the simple antidote to this disease. While bacteria, f i, and plants produce their own vitamins, our bodies can't, so we have to them from other sources. So how does the body get vitamins from out there here? That's dependent on the form these compounds take. Vitamins come in types, lipid-soluble and water-soluble, and the difference betwee…' Transcript Files	Summarized Text	0 1676963736.txt Hi, it's Mr. Andersen and this is AP Environmental Science Video 1. It's the fi Hi, it's Mr. Andersen and this is AP Environmental Science Video 1. We're pushi 1 1676963736(3).txt Hi, it's Mr. Andersen and this is environmental science video 5, it's on water Hi, it's Mr. Andersen and this is environmental science video 5, it's on water 2 1676963736(1).txt Hi, it's Mr. Andersen and this is environmental science video three, it's on ge Hi, it's Mr. Andersen and this is environmental science video three, it's on ge Number of Testing text: 3

3.	Simplifying transcripts using coreference resolution.	 extracted_transcripts[1] ' A, C, E, D, B, K. No, this isn't some random out-of-order alphabet. These e vitamins. And just like letters build words, they're the building blocks t keep the body running. Vitamins are organic compounds we need to ingest i mall amounts to keep functioning. They're the body's builders, defenders, a maintenance workers, helping it to build muscle and bone, make use of nutri s, capture and use energy, and heal wounds. If you need convincing about vi in value, just consider the plight of olden-day sailors who had no access t itamin-rich fresh produce. They got scurvy, but vitamin C, abundant in frui and vegetables, was the simple antidote to this disease. While bacteria, fu i, and plants produce their own vitamins, our bodies can't, so we have to g them from other sources. So how does the body get vitamins from out there i here? That's dependent on the form these compounds take. Vitamins come in t types, lipid-soluble and water-soluble, and the difference betwee…' Transcript Files	Simplified Text	 coref_clusters_30: [jumping to humans is, it does] ℹ Document with resolved references For almost a decade, scientists chased the source of a deadly new s and most isolated caverns. scientists chased found a deadly new Cave. a deadly new virus through a coronavirus that caused an epi ome, or SARS, in 2003. Coronaviruses are a group of viruses cover ike a crown- or "corona" in Latin. There are hundreds of known co coronaviruses. humans, and can cause disease. The coronavirus SAR yndrome, or SARS,MERS-CoV causes MERS, and SARS-CoV-2 causes the ect four cause colds, mild, highly contagious infections of the n nd cause much more severe illnesses. The seventh, which causes CO th, which causes COVID-19, easily, but can severely impact the lu oplets containing The seventh, which causes COVID-19, out. The se new person when droplets containing the virus spray a new person est in enclosed spaces, where people are close together. Cold wea g from drying out, enabling The seventh, which causes COVID-19, s UV exposure from sunlight may damage The seventh, which causes CO more for established viruses. But because no one is yet immune to ntial hosts that a new virus, n't need ideal conditions to spread in the host's cells and fuse with the host's cells andenabling a nery to replicate a new virus, genes. Coronaviruses store Coronav ther RNA viruses or DNA viruses. RNA viruses tend to be smaller, end many hosts and replicate quickly in many hosts and In general mechanism, whereas DNA viruses do. So when an RNA virus replicate ely to have mistakes called mutations. Many of mistakes called mu ome make an RNA virus replicates suited for certain environments- en occur when a virus jumps from animals to humans. occur when tr bola, Zika, and SARS epidemics, and the COVID-19 pandemic. Once ally not enough to create a new virus, but enough to create varia Coronaviruses have a few key differences from most RNA viruses. C eaning Coronaviruses have the most genes. That creates more oppor ract more opportunity for harmful mutations. coronaviruses have a for replication errors and corrects mistakes. have a coronaviruse ion rate, than other RNA viruses. While have a sound formidable, ising sign when it comes to disarming coronaviruses muchAfter an nize germs and destroy germs and quickly if germs and us again so
4.	Evaluating performance of Latent Dirichlet Allocation topic modelling algorithm.	 coref_clusters_30: [jumping to humans is, it does] ℹ Document with resolved references For almost a decade, scientists chased the source of a deadly new s and most isolated caverns. scientists chased found a deadly new Cave. a deadly new virus through a coronavirus that caused an epi ome, or SARS, in 2003. Coronaviruses are a group of viruses cover ike a crown- or "corona" in Latin. There are hundreds of known co coronaviruses. humans, and can cause disease. The coronavirus SAR yndrome, or SARS,MERS-CoV causes MERS, and SARS-CoV-2 causes the ect four cause colds, mild, highly contagious infections of the n nd cause much more severe illnesses. The seventh, which causes CO th, which causes COVID-19, easily, but can severely impact the lu oplets containing The seventh, which causes COVID-19, out. The se new person when droplets containing the virus spray a new person est in enclosed spaces, where people are close together. Cold wea g from drying out, enabling The seventh, which causes COVID-19, s UV exposure from sunlight may damage The seventh, which causes CO more for established viruses. But because no one is yet immune to ntial hosts that a new virus, n't need ideal conditions to spread in the host's cells and fuse with the host's cells andenabling a nery to replicate a new virus, genes. Coronaviruses store Coronav ther RNA viruses or DNA viruses. RNA viruses tend to be smaller, end many hosts and replicate quickly in many hosts and In general mechanism, whereas DNA viruses do. So when an RNA virus replicate ely to have mistakes called mutations. Many of mistakes called mu ome make an RNA virus replicates suited for certain environments- en occur when a virus jumps from animals to humans. occur when tr bola, Zika, and SARS epidemics, and the COVID-19 pandemic. Once ally not enough to create a new virus, but enough to create varia Coronaviruses have a few key differences from most RNA viruses. C eaning Coronaviruses have the most genes. That creates more oppor ract more opportunity for harmful mutations. coronaviruses have a for replication errors and corrects mistakes. have a coronaviruse ion rate, than other RNA viruses. While have a sound formidable, ising sign when it comes to disarming coronaviruses muchAfter an nize germs and destroy germs and quickly if germs and us again so Simplified Text	Top keywords and weights associated with keywords	 test_filtered_sentences /usr/local/lib/python3.9/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: and should_run_async(code) [{" Do kids know how plants make plants's own food? No? Well, this video elaborates t plants's own food", " Glucose is used by the plants for the plants's growth", ' Leaves have important cells called mesofil cells', ' Leaves on the plant have pores, very similar to the pores on the skin of our bodi ' Once the carbon dioxide and the water reach a green color component called chloro photosynthesis starts to take place', ' Photosynthesis is the conversion of light energy into chemical energy by plants', " Photosynthesis is the process used by plants to make plants's own food or in more ' Plants have tubes called xylum, located in the stems through which the water from " Similarly, to make plants's food, plants also need some essential factors, includ ' Some amount of the extra glucose which is not used is also stored in the roots of ' Some of Glucose is used immediately and the extra glucose which is not used is st ' The carbon dioxide present in the air, which is responsible for Photosynthesis, e on the skin of our bodies', ' The following reaction takes place in the leaves of the plant during the process " The oxygen that is released is used by human beings to breathe during human being photosynthesis is called photosynthesis? Well the word photosynthesis is a combinatio " There are some plants which don't need the process of photosynthesis to grow", ' This is because energy from the sun is absorbed by the water layers and only some ' This system is spread throughout the different parts of the plant, including the ' This system transports vital nutrients to the plant', ' This system works similar to humans sucking liquids through a straw', ' Xylums in the plant are like blood vessels in the human body that act as an impor ' a green color component called chloroplasts is responsible for the green color of ' chemical energy is used by plants for growth and nourishment', ' important cells called mesofil cells contain a green color component called chlor ' pores, very similar to the pores on the skin of our bodies are called stomata', " some essential factors, including light, water, nutrients, soil, etc, together he " some plants which don't need the process of photosynthesis to grow include mushro ' venous fly traps trap and catch small insects which come near the leaves and eat
5.	Creating a dataset by converting existing MCQs to assertive sentences.	 [ ] import json f = open("/content/test.json") ques_f = json.load(f) len(ques_f) 1000 [ ] ques_f[0] {'question': 'Compounds that are capable of accepting e 'distractor3': 'residues', 'distractor1': 'antioxidants', 'distractor2': 'Oxygen', 'correct_answer': 'oxidants', 'support': 'Oxidants and Reductants Compounds that are agents) because they can oxidize other compounds. In th donating electrons, such as sodium metal or cyclohexane of another compound. In the process of donating electro Equation 3.30 Saylor URL: http://www. saylor. org/books MCQ Dataset	Reference Set	 assertive_sentences2 ['clone in biotechnology means a genetically exact copy of an organism', 'the height above or below sea level is called elevation', 'plant hormones in plants control different processes', 'deciduous of tree is dominant in temperate forests', 'highly viscous of viscosity is found in long-chain hydrocarbons', 'mid-ocean ridges can be about 2 km', 'insects are by far the most common type of invertebrate', 'waves deposit sediments to form sandbars and barrier islands', 'a neon light produce visible light by electroluminescence', 'most living things use oxygen to make atp from glucose', 'feldspar and quartz are the two most common silicates', 'the amount of water means vapor in the air humidity', 'goosebumps happens when mammals raise their hair with tiny muscles in the skin', 'bone fractures are people with osteoporosis at increased risk of', 'between atoms is energy stored in a chemical substance', "aluminum is the most abundant metal of the earth 's crust", 'thermal energy is the term for the total kinetic energy of moving particles of matt 'pupil allows light to enter', 'the brain enters stem and influences activity in the cerebellum , spinal cord , and 'a body cavity may be present or absent in triploblastic', 'central nervous system are the brain and spinal cord apart of', 'enzymes are biochemical catalysts that speed up biochemical reactions', 'genetic drift can cause harmful alleles to become fixed', 'fossil record shows that dinosaurs originated 200-250 million years ago', 'analytical balance is used to make very sensitive mass measurements in a laboratory "the purpose of bright colors is on a flower 's petals to attract pollinators", 'pi bonds are formed by the side-to-side overlap of p orbitals', 'backbone is another name for the vertebral column', 'the physical transformation of an insect is moving metamorphosis through stages of 'energy is required to move or change matter from one state to another', 'golden eagles is considered the largest of the raptors', 'nucleus contains positive protons and neutral neutrons', 'kidneys control the amount of water , ions , and other substances in the blood by e

6.	Comparing processed sentences with the sentence present in reference set.	 Reference Set and Processed Sentences	Informative Sentences	
7.	Extracting plausible answer words from the selected sentences.	 Informative Sentences	Sentence Keyword Map	
8.	Extracting plausible wrong options which are closely related yet diverse.	 Filtered Keywords	Distractors	
9.	Converting selected sentences into questions.	 Filtered Sentence Keyword Map	Questions	

CHAPTER 5 – CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

In conclusion, AGeES can be a potent tool for not just improving the quality of education but also enhancing the processes of learning and assessment. Moreover, the system saves a large amount of time and energy of the teaching community in framing questions. Such a system has the capability to efficiently extract important material from video lectures or tutorials, produce relevant questions, and assess the responses of students in real-time. However, there are still certain issues that need to be resolved, such as assuring the accuracy and reliability of the questions, reducing bias and mistakes, and accommodating various learning preferences and styles.

5.2 FUTURE WORK

As an attempt to enhance the use of the system in future, the question generation model can be upgraded to form multiple kinds of questions other than just 'wh' questions. In fact, the current distractor generation model face setbacks when it comes to generating for subject specific words, which can be improved with addition of input context.

Moreover, the current model is restricted to create MCQs for science-based videos. This can be improved by expanding the reference set.

APPENDICES

A. Fault analysis

```
****************************************************************
<pad> question: What is responsible for the exchange of gases?</s>
pores
 Sense2vec_distractors successful for word :  pores
['Sebum', 'Dead Skin Cells', 'Oil Glands', 'Eye Area', 'Dead Skin', 'Scalp', 'Blackheads', 'Dry Flakes', 'Nose Area', 'Oiliness', 'Dryness']
pores
------------------>
Sebum
Oiliness
Scalp
['Sebum', 'Oiliness', 'Scalp']
****************************************************************
<pad> question: Along with sunlight, carbon dioxide and sunlight, what is one of the three essential factors for photosynthesis?</s>
water
[]
NO DISTRACTORS GENERATED
****************************************************************
```

Fig A.1. Distractor fault case

The sent2vec module, which is used for distractors generation, often face trouble when input is a jargon. Such inputs lead to no distractors being generated.

```
{'<pad> question: What does the electron radiate?</s>': 'energy',
 '<pad> question: What will all the energy of the electron fall into?</s>': 'nucleus',
 '<pad> question: What is the main source of energy that will eventually get over and fall into the nucleus?</s>': 'electron',
 '<pad> question: Whose atomic model was laudable?</s>': 'rutherford',
 '<pad> question: What model did Rutherford use to explain the structure of atoms?</s>': 'atomic model',
 '<pad> question: Who was the next scientist to be on our list?</s>': 'niels bohr',
 '<pad> question: What was made by Niels Bohr?</s>': 'slight modifications',
 '<pad> question: Who was Niels Bohr?</s>': 'next legendary scientist',
 '<pad> question: What is the first letter next to?</s>': 'nucleus',
 '<pad> question: Who suggested that electrons revolving in fixed defined orbitals do not radiate energy?</s>': 'niels bohr',
 "<pad> question: What is Niels Bohr's title?</s>": 'next legendary scientist'}
```

Fig A.2. Question formation fault case

The T5 transformer model used for question generation is limited to only 'what' question and not 'how' or 'which' questions which makes them redundant.

```
[ ]  print(keywords)

     ['charged', 'energy', 'nucleus', 'structure', 'points', 'revolve', 'sun', 'electron', 'atom', 'rutherford', 'atomic model',
```

Fig A.3. Keyword extraction fault case

AGeES only considers answers with one-two words so all keywords with more than two words are not considered.

REFERENCES

1. Chen Liang, Xiao Yang, Neisarg Dave, Drew Wham, Bart Pursel, and C. Lee Giles. (2018). Distractor Generation for Multiple Choice Questions Using Learning to Rank. In Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications, 284–290.

2. CH, D. R., & Saha, S. K. (2022). Generation of Multiple-Choice Questions from Textbook Contents of School-Level Subjects. IEEE Transactions on Learning Technologies.

3. Ma, T., Pan, Q., Rong, H., Qian, Y., Tian, Y., & Al-Nabhan, N. (2021). T-bertsum: Topic-aware text summarization based on bert. IEEE Transactions on Computational Social Systems, 9(3), 879-890.

4. Nwafor, C. A., & Onyenwe, I. E. (2021). An Automated Multiple-Choice Question Generation Using Natural Language Processing Techniques. arXiv preprint arXiv:2103.14757.

5. Mukta Majumder and Sujan Kumar Saha. (2015). A System for Generating Multiple Choice Questions: With a Novel Approach for Sentence Selection. Conference: Proceedings of the 2nd Workshop on Natural Language Processing Techniques for Educational Applications.

6. Zhaopeng Qiu, Xian Wu, Wei Fan. (2020). Automatic Distractor Generation for Multiple Choice Questions in Standard Tests. Conference: Proceedings of the 28th International Conference on Computational Linguistics.

7. Dhanya, N. M., Balaji, R. K., & Akash, S. (2022, February). AiXAM-AI assisted Online MCQ Generation Platform using Google T5 and Sense2Vec. In 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 38-44). IEEE.

8. Mehta, P. K., Jain, P., Makwana, C., & Raut, C. M. (2021). Automated MCQ Generator using Natural Language Processing.

9. Maniar, S., Parmani, J., Bodke, M., & Saxena, K. (2021, July). Generation and grading of arduous MCQs using NLP and OMR detection using OpenCV. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

10. Ming Liu, Vasile Rus and Li Liu. (2018). Automatic Chinese Multiple Choice Question Generation Using Mixed Similarity Strategy. IEEE Transactions on Learning Technologies.

11. Dmytro Kalpakchi & Johan Boye. (2021). BERT-based distractor generation for Swedish reading comprehension questions using a small-scale dataset. 14th International Conference on Natural Language Generation.

12. Animesh Srivastava, Naeem Patel, Shantanu Shinde, Siddhesh Despande, Anuj Dalvi, Shweta Tripathi. (2020). Questionator - Automated Question Generation using Deep

Learning. International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE).

13. Selvia Ferdiana Kusuma, Daniel O Siahaan & Chastine Fatichah. (2020). Automatic Question Generation In Education Domain Based On Ontology. International Conference on Computer Engineering, Network and Intelligent Multimedia.

14. Jiaying Lu, Xin Ye, Yi Ren, & Yezhou Yang. (2022) Good, Better, Best: Textual Distractors Generation for Multiple-Choice Visual Question Answering via Reinforcement Learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

15. Ainuddin Faizan & Steffen Lohmann. (2018). Automatic Generation of Multiple Choice Questions from Slide Content using Linked Data. Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics.