

Automated Question Generation

¹Saicharan Polishetty, ²Abhishek M Bhayya, ³Amadaguru Kundansai, ⁴Rashi

^{1, 2, 3, 4} Department of Computer Science and Engineering, PES University

¹saicharan19022002@gmail.com, ²abhishek30102001@gmail.com, ³saikundan2502@gmail.com,

⁴yadavrashi566@gmail.com

Abstract— The most important thing in learning is assessment and the question is crucial for assessment. Universities, colleges, and other educational organizations increasingly use online exams. In general, the assessment pattern is shifting towards objective evaluation, such as MCQs. Match the following, complete the blanks, and True/False based. Even while MCQs have several benefits, such as computerized evaluation and shorter testing times, manually preparing exam questions is a time-consuming task for teachers. This study provides an automated exam question generator to address this problem in the creation of objective evaluation. A real sentence from the book that serves as an instructional line is read, and then multiple false phrases are made by not only denying the true assertion while also substituting the basic vocabulary with an antonym. The test questions are made by randomly rearranging the True and False sentences. Students must be able to recollect and apply essential concepts from their course material in order to score well on exams. To achieve this, practice and self-evaluation through questions are crucial. As a result, we present an automated technique for creating Gap-Fill Questions. It is becoming more and more important to have an automated MCQ generation system that is efficient in both time and money. Wordnet and Sense2Vec are used to create distractions in order to produce query options.

Keywords—Objective Assessment, Distractors, Electronic Evaluation, Phrases, Gap-Fill questions, True and False Statements.

I. INTRODUCTION

In the fast-growing technology today, the user's accessibility of video is growing at an exponential rate. Users must view the entire video in order to glean any helpful information, which is the motivation for this. The editing and management of such huge data are different from those of text because of the temporal character of the video. As a result, it might be difficult and time-consuming to find anomalous, suspicious, or abnormal behavior in high-dimensional data.

Numerous organizations, including those in the education sector, were impacted by the Covid-19 pandemic. To make it simpler for students and to provide access, all educational institutions altered their curricula, and everything was made available online. Live videos of courses are now being posted to Google Drive. It can be challenging to follow the lessons in some locations where students are experiencing network connectivity issues. Due to the difference in setting between home and a college and the decorum there, students' focus will shift. Going through the live videos during revision is time-consuming; having a summary of the video instead would be more beneficial.

Exam-style questions are a fundamental teaching strategy that can be applied to several goals. Along with being a tool for assessment, questions have the power to affect how students learn. Students have access to learning resources and can learn from a few online sources. On the other hand, the

learner's assessment requires the manual questions from the course materials. To our knowledge, there hasn't been a general assessment method offered in the literature to determine learners' learning gaps from e-reading materials. Therefore, automated question generation and evaluation approaches can help with the automation of the assessment system.

In the objective test question, you must either insert a phrase or word to complete a sentence or choose the best response from a selection of possibilities. The creation of the handicraft evaluation item requires a lot of time and effort. However, the system may generate questions automatically using the active learning architecture. The best way to determine a learner's level of comprehension is through questions with alternatives, which also makes the assessment process quicker and easier. The suggested study focuses on developing fill-in-the-blank questions to evaluate a learner's knowledge gaps. Most of a text's sentences are poor candidates for insightful questions. As a result, the task of selecting informational sentences grabs our interest and makes us wonder. The chosen topic-word is dropped to form an inquiry or come from an instructive sentence.

Generating false statements from the ability to directly, readily, and efficiently generate valid evaluations from true sentences and other educational-purposes queries is useful for identifying whether students have any misunderstandings of the subject matter. In this essay, we will utilize a declarative phrase rather than an interrogative one to define the T/F question. The quality of the generated questions, however, is insufficient for evaluation due to the scarcity of training data, the challenge of selecting relevant testing points from a specific text, the high incidence of grammatical and semantic errors, and other factors.

Our methodology seeks to determine the most efficient method for extracting summaries from instructional texts and creating multiple-choice questions, i.e., MCQ's, Fill in the blanks, Match the following and True or False.

II. RELATED WORKS

The majority of machine learning methods need those inputs be represented by feature vectors of specified length. When the inputs are text sentences and paragraphs, this task is difficult. This issue has been covered in numerous research using both supervised and unsupervised methods. The issue of distractor generation or selection is still a focus of more modern techniques, but they apply it to various areas. It takes a lot of time to create assignments with varying degrees of difficulty. As a result, during the past 20 years, there has been a lot of research into the idea of questions being generated automatically suggesting a system that makes use of the

Naive Bayes algorithm with supervised learning. By distilling the text, selecting nouns from the summary, and creating questions from these nouns, they hoped to produce open, concise inquiries that were semantically sound.

Text summarization is an important aspect covered by all researchers as it can help to reduce lengthy passages of text into concise summaries, which can then be used as a basis for generating questions. Question stems, the elements of the questions that offer question to the reader, can be created from summarized material. It can be a valuable tool for automatically generating questions from large documents and can save time and effort in the question creation process. BERT (Bi-directional representation and Transformers) technique is used for text summarization in this paper [3]. It's the method for processing natural language based on neural networks. It's an open-source, pre-trained model which makes it easier for computer to comprehend language a little more than humans do. A lexical database for the English language called Wordnet is used to generate distractors which connects words through linguistic relationships. These relations include hypernym and hyponym. After completing all these procedures, questions in fill-in-the-blank variety are created by mapping the keyword to the appropriate sentence and Wordnet generates the necessary distractors.

In this paper [1], a new state of art approach for video text summarization based on audio, emotions and video content is proposed. The method takes the output text from the video and extracts it. The final product is a written summary of the frame description. Emotions extracts a second text, and the audio recording is converted into a third text through abstractive summarization. All the three retrieved texts are to be combined using NLP.

The majority of educational institutions and organizations already designate a board of staff members to review the curriculum or other materials and manually develop MCQ-based questions. As we can see, it takes a lot of time and requires a lot of employees. In this paper [7], the actual need of traditional approach to create multiple choice questions is proposed which will assist in resolving all those challenges in all conceivable ways.

When the input texts are sentences and paragraphs, it's challenging to represent them by fixed length vectors. In order to encode words with shared semantic information to vector representation, [2] demonstrated the usage of deep LSTM-based model in which data is prepared by tokenizing and padding the text, then the model architecture must be defined with an embedding layer and model evaluation is done using validation metrics and fine tuning. The given latent representation of sentences has proven effective for document summary and sentence paraphrase.

The most efficient and natural way for people to communicate is through audio. Although it is a simpler approach to understand the information, this type of communication has several drawbacks. The audio file that

serves as the primary input for text summarizing procedure described in this study [6] can either be recorded live or be extracted from previously recorded human speech by using a GUI. After that, the audio is transformed into a wav text file that serves as the input for the text summarizer processing.

The fundamental issue with creating Cloze Questions is that, without extensive experience, it might be difficult to come up with suitable distractors-incorrect options. The majority of currently used methods are built upon domain-specific templates that require professionals to elaborate. Recently, methods based on discriminative methods that use annotated training data have also been developed. This challenge of creating fill-in-blank-quizzes has been formalized in this study [8] utilizing two clear learning strategies: Sequence classification and sequence labelling and suggests concrete LSTM structures for both cases. These findings indicate that both suggested training plans appear to provide reasonable outcomes, with an Accuracy/F1- score of almost 90%.

Multiple True False (MTF) Questions are a sort of multiple-choice question that, among other question types, are an easy and effective approach to impartially test factual knowledge. The ability to discriminate between truthful and untrue statements is put to the test. Because a question stem can include both correct and wrong assertions, learners must assess each statement separately. The ability to discover learners' misconceptions and knowledge gaps is thus a benefit of MTF Questions as a machine gradable format. However, manually generating such questions by human writers is a time-consuming activity that is challenging to scale up. This article [9] explores a method for creating MTF questions automatically using resources from already-existing textbooks. Using summarization and GPT-2, the proposed MTF question generation system can produce MTF questions from academic study material. Thus, we were able to fit 33% of the created true and false statements. Additional raters could aid in evaluating a broader dataset encompassing more disciplines for future investigations. Additionally, extra subject-specific literature could be used to train the GPT-2 model in order to increase the believability of the created false assertions.

The requirement for a model that generates pertinent questions faster is evident as a result of studying related work. The following section goes into more detail on how we went about creating such a model.

III. PROPOSED SOLUTION

In automated question generation systems, machine learning techniques are frequently applied. The purpose of this study is to develop an automated exam question generator that generates objective evaluation questions in order to address the time-consuming issue of manually creating test questions.

Currently automated question generation, like any other work using natural language processing (NLP) has its share

of difficulties. Automated question creation systems greatly benefit from machine learning, which makes it possible to generate questions from text, domain-specific data, datasets, and user preferences. Our approach is unique as it combines cutting-edge NLP methods with domain expertise and incremental enhancements based on testing and evaluation. In order to better grasp the context of the provided text or data, it may be helpful to use advanced natural language processing techniques, such as deep learning models or contextual embeddings. This can entail integrating domain-specific knowledge, syntactic and semantic links, and contextual information into the question creation process.

A. Overview

The methodologies general framework is to produce pertinent questions from literature to assess the user's comprehension. We created a variety of questions using middle school literature as the input, including Multiple Choice Questions, False from True assertions, Match the Following and Fill in the Blanks.

Data Preprocessing is crucial since the correctness of the project, which determines the ultimate result, is directly reflected in it. Therefore, it is crucial to give data pre-processing some time. The preparation of the data includes formatting the necessary features and removing the unnecessary information.

B. MCQ Generation

The two main steps that must be used to construct a MCQ question are creation of distractors and questions. Keyword extraction is done by Flash Text which is a Python library designed primarily for word replacement and searching in documents. Now, for Flash Text to function, it needs a word or list of words as well as a string. The string is then searched for or substituted with the terms that Flash Text refers to as keywords.

The complexity of the question increases as the relevance of the distractors to the correct solution increases. We employ a variety of methods to produce the Distractors for the multiple-choice questions, and we select two effective algorithms.

WordNet is a lexical database which contains the meaning relationships between words in more than 200 different languages. Synonyms, Hyponyms and Hypernyms are only a few of the semantic relationships that WordNet links word into. It labels the semantic relationship between the words and captures relationships between the words. Specific distractors are produced by utilizing semantic relations, which are hypernyms and hyponyms found in the wordnet. A hypernym and its hyponym have a certain kind of relationship. A hypernym is an all-encompassing term.

A free multilingual knowledge graph is ConceptNet. Both experts created materials and crowd sourced expertise are included. Word embeddings can also be made with ConceptNet, exactly like Word2Vec. ConceptNet marks the conceptual connections between words in a manner similar to WordNet. Compared to WordNet, it's more thorough. The

idea includes relationships between terms such as "is a", "made of", "similar to", "is used for" and others. Distractors for things like places, things etc. that have a "Part Of" relationship can be produced well using ConceptNet. A state like "California", for instance, might be a component of the US. There is no way to distinguish between various word senses in ConceptNet with a particular term, we must proceed with whatever sense it provides.

A Neural Network model called Sense2Vec creates vector space representations of words from significant corpora. This algorithm is an improvement on the infamous Word2Vec one. Instead of producing tokens of words, Sense2Vec produces embeddings for "senses". Information about the context is recorded. Since the outcomes were a lil better, we will use trained vectors from 2015 rather than those from 2019.

In contrast to some word vector algorithms, which are taught with only single words, noun phrases and named entities are annotated during training, allowing for the inclusion of multiword phrases like "Natural Language Processing" as well. Along with the similarity of already chosen keywords and key phrases, MMR (Maximal Marginal Relevance) takes the document's keywords and key phrases into account. As a result, the keywords chosen are as diverse as possible within the context of the content.

For questions we used Text to Text Transfer Transformer, often known as T5, is a transformer-based architecture that employs this method. Every task is framed as feeding the model text as input and training it to output some goal text, including translation, question answering and classification.

The T5 model will be trained to produce formal queries. We provide a context which may be a sentence or a paragraph- as well as a response, and the T5 model uses this information as input to create a question.

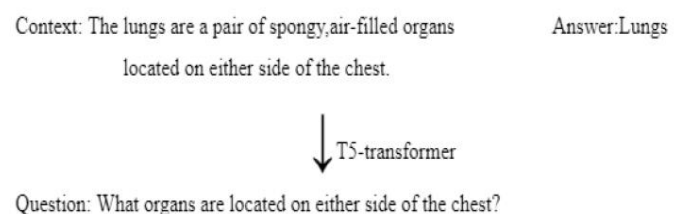


Fig. 1. Context to Question using T5

Thus, given a situation [1] in which a sentence or paragraph is provided along with its solution, our T5 model then generates the question/s you can see above. Transfer learning involves pre-training a model on a task with lots of data before fine-tuning it on a subsequent task, has become a potent method for Natural Language Processing. Models should be trained with general tasks first, then they should be fine-tuned with small amounts of text to perform particular tasks. T5 views every task processing issue as a "text-to-text" issue, where the text is the input and new text is produced as the output. A single T5 model is capable of carrying out numerous tasks, including summarization and translation. Simply alter the prefix and specify the action to take.

Utilizing Cross-Entropy loss, T5 is trained. T5 is a masked language modeling algorithm, or is trained on a denoising objective, where the model is trained to anticipate corrupted or missing tokens in the input. Input is a statement with a few words masked out. T5 employs a greedy algorithm during testing (i.e., choosing the highest probability logit at every timestamp).

C. Generating False statements

Generating false statements from true statements can be achieved using various techniques such as, adding or removing verb or noun phrases negation's, modifying name entities, altering adjectives or main verbs. Using Constituency Parsing and OpenAI GPT2, we will create True or False questions based on the input text.

Constituency Parsing is the process of separating a text into its constituent words or phrase. Types of phrases make up the parse tree's non terminals, while the sentence's words make up its terminals. To get rid of the final verb or noun phrase, we'll use the allennlp parser. Finding a noun phrase is the first step in using constituency parsing techniques to comprehend the sentence and identify the final noun or verb phrase [2].

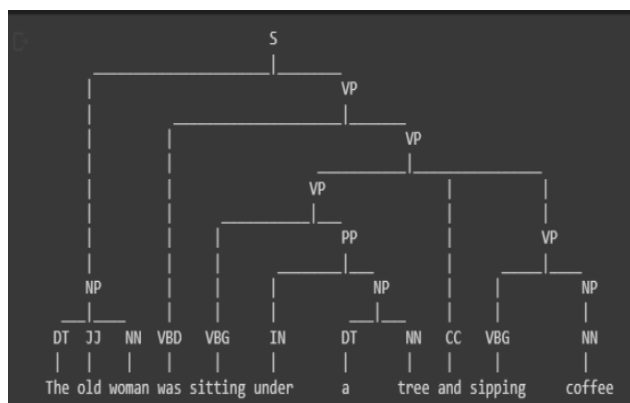
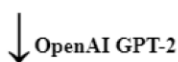


Fig. 2. Output of Constituency Parser

Split put the final word or verb to finish the sentence. This incomplete sentence is then fed to OpenAI Gpt-2 which is trained to anticipate words in transformed-based. The most well-known applications for transformer models are text production, summarization, and language translation. We choose the option that will provide the false statement the best out of all those produced by the OpenAI GPT-2 [3].

The old woman was sitting under the tree and



The old woman was sitting under the tree and looking at the bushes.
The old woman was sitting under the tree and reading a book.
The old woman was sitting under the tree and sipping coffee.

Fig. 3. Output of OpenAI GPT-2

D. Fill in the Blanks

Natural language processing (NLP) strategies and methods can be used to generate fill-in-the-blank questions from text theory components using machine learning.

Given the input text, extract key phrases using python keyword extraction package and return to the content and collect every sentence containing the specific phrase. The last step is to choose any sentence from the list, just change the keyword to "blank", and then instruct the reader to fill the blanks.

With the use of the Keywords Extraction text analysis technique, you may quickly learn crucial insights about a text. It should make it easier to extract pertinent keywords from any text and spare you the time-consuming task of reading the entire thing. We have numerous keyword extraction algorithms in the PKE (Python Keyword Extraction) library and using a multipartite graph technique, we will extract unsupervised key phrases from them. This unsupervised method allows it to work on any fresh document because there is no labeled data or prior training. We will extract elements from any supplied document, such as noun-phrases, verbs, or adjectives. As a result, we can tell the algorithm what to use as a keyword and presume that we are only providing noun-phrases. Therefore, we categorize all the collected nouns into subjects [4]. To organize these keywords by subjects, they apply hierarchical clustering in the algorithm.

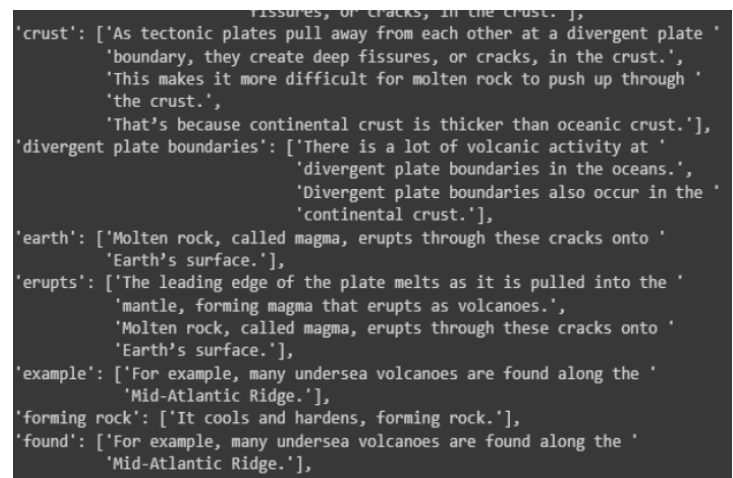


Fig. 4. Mapping of keywords with their sentences

Finally, we have a single topic with three or four keywords underneath it, and we also have themes and keywords on high-level pages. There is a graph in front of us, and each point or node has a varied weight while connecting to the others. Weights are determined by the co-occurrence of the key phrases. We can claim that a keyword is important if there are many incoming nodes to a certain node that is to a particular keyword or key phrase. Our objective is to identify the top-ranking nodes among all the nodes with associated edges and use them as the top 10 keywords or key phrases in the article. The top n nodes in the graph are calculated using the text rank method to determine which nodes are the most significant, and those top n words are returned as key phrases. To evaluate the importance of a certain key phrase.

Once the keywords have been obtained, we return to the context, map the phrases that include the keywords, and then store them in a dictionary [5,6].

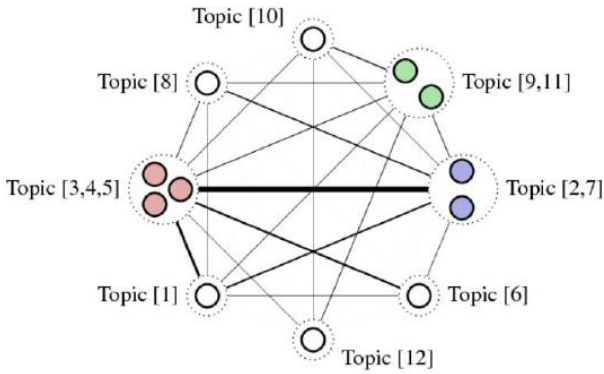


Fig. 5. Unsupervised Keyword Extraction- Topic rank graph

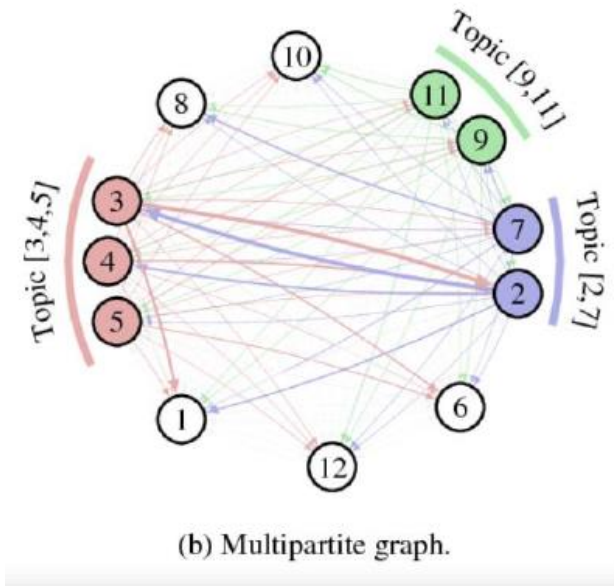


Fig. 6. Unsupervised Keyword Extraction- Multipartite Graph

E. Match the following

Any middle school textbook will have the following assessment sorts of questions, where we must match words with their meanings on the right after a list of terms is presented on the left. Usually, the words and their meanings are mixed up to make things more difficult to understand.

We'll utilize the Python Keyword Extraction package to accomplish this and extract the keywords from the context. Therefore, once the keywords have been extracted, those keywords can be used to replace the words on the left of the following match, and their meanings are provided on the right side. We've taken the keywords from a piece of material, but the task here is to take the sense of the word. We won't be able to discern the precise meaning of the word being used in our context without knowing the context's specifics. The term for this issue is Word Sense Disambiguation [7].

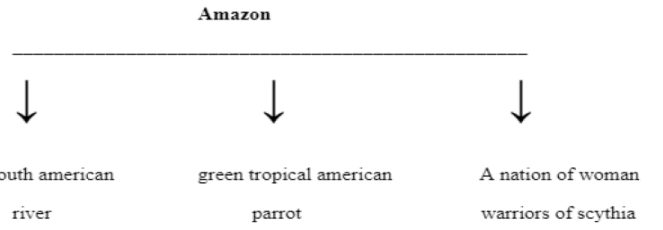


Fig. 7. WSD

Therefore, among all of these different contexts, we need to identify the proper context for the word “Amazon” that effectively clarifies and establishes the proper definition of the term. We gather all the sentences from the text that use the word “Amazon” as one of the methods for determining the context. To do this, we'll use the BERT algorithm, which is transfer based, much like the T5 transformer which is an encoder and decoder-based technique. With BERT, any input or block of text is encoded into a fixed encoding vector because it is merely an encoder-based algorithm. Given that BERT is only an encoder-based model, it is mostly use for text classification. BERT accepts a sentence or context as input, and its output is a fixed vector [8].

```
{ 'amazon': 'a major South American river; arises in the Andes and flows '
  'eastward into the South Atlantic; the world's 2nd longest river '
  '(4000 miles)',
  'big': 'above average in size or number or quantity or magnitude or extent',
  'free': 'able to act at will; not hampered; not under compulsion or restraint',
  'hunter': 'someone who hunts game',
  'lion': 'large gregarious predatory feline of Africa and India having a tawny '
  'coat with a shaggy mane in the male',
  'mouse': 'any of numerous small rodents typically resembling diminutive rats '
  'having pointed snouts and small ears on elongated bodies with '
  'slender usually hairless tails',
  'net': 'a trap made of netting to catch fish or birds or insects',
  'tiny': 'very small' }
```

Fig. 8. Mapping of Keywords and context

IV. RESULTS AND CONCLUSION

The findings of our study show how automated question generation system creates multiple choice questions (MCQs), true or false questions, fill in the blanks, and questions that match the following statements to the text.

The algorithm was able to produce questions that were pertinent to the text theory parts from which they were derived and covered a wide range of themes and concepts. The generated MCQs' level [9] of difficulty was suitable, offering a well-balanced set of questions for evaluation purposes.

Fig. 9. MCQ Generation

Similar to how it did well when creating fill-in-the-blank questions, our system had a good rate of success in completing the blanks with pertinent words or phrases gleaned from the text. The generated questions [10] were able to gauge students' retention and application of the knowledge contained in the text's theoretical sections, and the questions' level of difficulty was appropriate for the intended audience.



Fig. 10. Fill in the Blanks

Our system produced true or false questions[11] including accurate statements and responses. The technology was able to pick out pertinent passages from the text and turn them into true-false questions that efficiently assessed students' comprehension of the material's theoretical components.

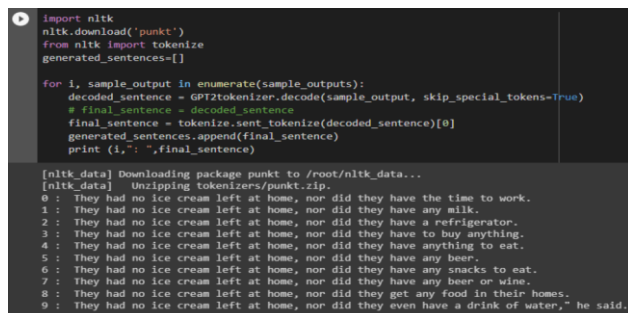


Fig. 11. True or False questions

According to the findings of our work, our system was highly accurate at producing Match the Following questions from text [12]. The algorithm efficiently constructed pairs of related items for matching by extracting keywords or phrases from the text passages. The created questions evaluated the students' capacity to establish relationships and linkages between various concepts offered in the text and were pertinent to the passages of the text.

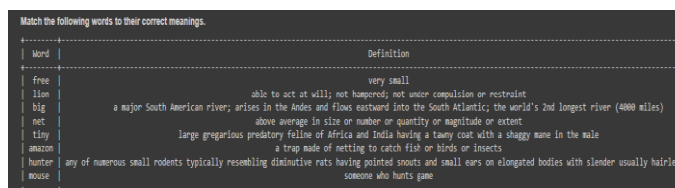


Fig. 12. Matching the Following

In conclusion, our work proposes an automated question creation system that efficiently generates multiple-choice questions (MCQs), true or false questions, and questions that fit the following statements from text theory parts using machine learning approaches and pretrained keyword extraction models.

Future studies might concentrate on enhancing the system's performance by incorporating more sophisticated natural language processing techniques, investigating various methods for question generation, and carrying out extensive user evaluations to confirm the system's efficacy in actual educational settings. Additionally, attempts could be made to guarantee the system's equity, dependability, and security in producing questions that are free from biases and errors, as well as safeguarding the integrity of the assessment procedure.

V. ACKNOWLEDGEMENT

This research work is funded technically and financially by PES University, Bangalore.

VI. REFERENCES

- [1] Ahmed Emad, Fady Bassel, Mark Refaat, Mohamed Abdelhamed, Nada Shorim, Ashraf Abdelraouf, "Automatic Video Summarization with Timestamps using Natural Language Processing text fusion", Misr International University, Cairo, Egyptahmed1703326, Fady1710742, Mark1711712, Mohamed1701989, nada.ayman, ashraf.raouff@miuegypt.edu.eg, 2021
- [2] Chi zhang, Shagan Sah, Thang Nguyen, Dheeraj Peri, Alexander Louiy, Carl Salvaggio, Raymond Ptucha, "Semantic Sentence Embeddings for Paraphrasing and Text summarization", Rochester Institute of Technology, Rochester, NY 14623, USA, Kodak Alaris Imaging Science R&D, Rochester, NY14615, USA, 2017
- [3] Pritam Kumar Mehta, Prachi jain, Chetan Makwana, M Raut, "Automated MCQ Generator using NLP", Dept. of Computer Engineering, Datta Meghe College of Engineering, Navi Mumbai, India, 2021
- [4] Zygmunt Pizlo, Arnon Amir, Dulce B. Ponceleon, Edward J.Delp, "Automated Video Summarization using speech transcripts", article in IEEE transactionson Multimedia, September 2006
- [5] Marco Furini, Vittorio Ghini, "An audio-video summarization based on audio and video analysis", Conference paper, February 2006
- [6] Pravin Khandare, Sanket Gaikwad, Aditya Kukade, Rohit Panicker, Swaraj Thamke, "Audio Data Summarizatoion system using NLP", Assistant Professor, Dept. of Computer Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 2,3,4,5UG Student, Vishwakarma Institute of Technology, Pune, Maharashtra (India)
- [7] Dr. S. Muthusundari, Vishwa A, Srinivasa kiruthik K S, Sukesh Raj R, "Automatic MCQ Generator", Associate Professor, Computer Science & Engineering Department R.M.D Engineering College Kavaraipettai, Chennai, India 2Zoho Corporation, Chennai
- [8] Edison Marrese-Taylor, Ai Nakajima, Yutaka Matsuo, "Learning to Automatically Generate Fill-In-The-Blank Quizzes", Graduate School of Engineering The University of Tokyo
- [9] Regina Kasakowskij , Thomas Kasakowskij, Niels Seidel, "Generation of Multiple True False Questions"