

# MCQ Question Generation using Natural Language Processing Techniques

Mr.V Krishna Reddy<sup>1</sup>, B. Harshavardhan Reddy<sup>2</sup>, M Dheeraj<sup>3</sup>,  
G Vishal<sup>4</sup>

<sup>1,2,3,4</sup>Department of Information Technology, J.B. Institute of Engineering and Technology, Hyderabad.

## Abstract

Automated Multiple Choice Question Generation (MCQG) is an important and difficult task in Natural Language Processing (NLP). It is a job of creating correct and related questions from the input text data. Generally, it is a time-consuming process for teachers to create meaningful and relevant questions on a specific topic or concept. Here, in this paper we are introducing Natural Language Processing Techniques which help to generate MCQ questions for computer-based Testing Examination (CBTE). By using NLP Techniques, we first extract important keywords present in the input text. In addition to schools, colleges, and coaching institutions, students can use the system to generate question papers for self-assessment. A user-friendly environment can be used to present the results.

**Keywords:** CBTE, MCQ, Natural Language Processing, Question Generation, WordNet

## I. INTRODUCTION

Automated question generation is a process of automatically generating questions from a given set of text or data. It uses natural language processing (NLP) techniques to analyze the text and generate questions related to the content. The generated questions can then be used for educational or research purposes. Automated question generation systems are useful for creating educational content and testing students' knowledge.

Creating questions helps students understand and retain information better, actively involving them in the learning process. In addition, creating questions encourages students to think more deeply and helps teachers assess their students' understanding of a topic more comprehensively and accurately. Generating questions can help students develop creative and analytical thinking skills.

Suppose if a teacher in a school or a college is interested in conducting assessment for students then it will be a time-consuming process for a teacher to create a question from every topic. This work can be reduced by generating questions automatically. So, in this paper we are trying to create MCQ question automatically using Natural language processing.

Computer scientists, and specifically artificial intelligence, or AI, study natural language processing (NLP) as it pertains to giving computers the ability to understand spoken words and text in much the same way humans can. A rule-based model of human language incorporating computational linguistics is combined with statistical, machine learning, and deep learning techniques. By combining these technologies,

computers can process human language in the form of text or audio data, understanding the meaning as well as the intent and sentiment of the speaker or writer.

NLP is used to translate text from one language to another, to respond to spoken commands, and to summarize large volumes of text quickly. NLP has likely been incorporated into voice-operated GPS systems, digital assistants, speech-to-text dictation software, chat bots used in customer service, and other consumer conveniences. Moreover, NLP is increasingly being used in enterprise solutions to improve employee productivity, expedite mission-critical business procedures, and streamline business operations.

Natural language processing can be structured in a variety of ways using different machine learning methods, depending on what is being analyzed. It could be something as simple as frequency of use or mood, or something more complex. Regardless of the use cases, an algorithm must be formulated. The Natural Language Toolkit (NLTK) is a suite of libraries and programs that can be used for symbolic and statistical processing of English natural language, written in Python. It can help with all kinds of NLP tasks like tokenization (aka word segmentation), part of speech tagging, creating datasets for text classification, and more.

WordNet is a machine-readable dictionary. It is a lexical database for the English language. In WordNet, nouns, verbs, adjectives, and adverbs are grouped into synonym clauses. This set is known as synsets. Each word expresses a specific concept. These synsets are interlined. There are semantic relationships and linguistic relationships between the elements of synsets. It works like a thesaurus, but WordNet has an advantage because it groups words with a specific sense.

## II. EXISTING SYSTEM

- In the existing systems questions needs to prepare by the person manually or else generated by using rule-based approach.
- There is still a lot of work to be done in this area because the algorithms used in this technique can only create brief queries that begin with the terms “what”, “where”, “when” or “who”.
- These systems show good results for some questions and shows bad results for some questions.

### **Drawbacks:**

- ✓ Sometimes the questions are grammatically incorrect.
- ✓ The correct answers may or may not be provided.
- ✓ The questions generated only using some keywords

### **Advantages of proposed system:**

- ✓ In proposed system we are using T5 text transformers to generate Multiple Choice Questions.
- ✓ Questions are free from grammatical errors.
- ✓ Along with options, the correct answer should be provided.

### III. LITERATURE SURVEY

Exams and the creation of relevant question papers seem difficult, ineffective, time consuming, and unneeded to teachers due to the expanding field of education. As a result, numerous programs, databases, and applications have been developed to address the issue. We have already investigated those applications, which include the following.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Automated question generation based on discourse connectives [1], with the question generation system's content selection and question construction modules divided into two parts. Finding the relevant part in the text from which to frame a question constitutes content selection, whereas the process of forming a question includes sense- disambiguating discourse connectives, identifying the question type, and applying syntactic transformations to the content. The researcher focuses on seven discourse connectives, including because, since, although, as a result, for instance, and for instance on that basis. Question type will be chosen based on whether the phrase includes since, in which case the question type will be Why. Two evaluators have examined the system for both the semantic and syntactic correctness of the question.

“A new methodology for streamlining and adapting the computational linguistics' practical component has been developed by Edward Loper and colleagues. A simple, expandable foundation for processing natural language is offered by the NLTK toolbox. Natural language processing is covered in both symbolic and statistical ways. Each module of the toolkit defines an alternate data structure or a job before it is executed. A set of core modules characterize various systems that are used all through the toolkit. Chunk Parsing and Probabilistic parsing are a few of many modules that are part of NLTK. Many of the toolkit's modules give us fantastic examples of what projects should look like, clean code, and comprehensive documentation” [2].

“A system that can generate numerous logical questions from the provided text input was built by experts. The system employs a three-step strategy, in which it chooses the best possible set of sentences from the input text from which it could generate questions, looks for the sentence's subject and context to determine its main idea (Gap Selection), and then examines the best type of question that could be generated from that sentence (Question Formation)” [3].

“The article by Aleena et al. focuses on the realization idea of the query generating system. The main idea is to understand the natural language of the system, only then can the machine process and manipulate the data. Data processing, key phrase extraction and NLP are the main concerns of the proposed system. In this way, a fast, secure and random system can be developed that is useful for many things, including education.” [4].

“Mrunal Fatang is working on a question paper generator that provides a solution to choose from various challenge boundaries and facilitates their creation in a very short time. It contains various modules that allow the system to quickly affect all systems. Modules such as admin module, user module and query entry and administration make this easy” [7].

“Noor Hasimah Ibrahim, who also worked on a system where text matching and question sorting were done within the system itself, but one of the major disadvantages of that system was that only a limited number of questions could be added to the system.” [9].

#### IV. SYSTEM ARCHITECTURE

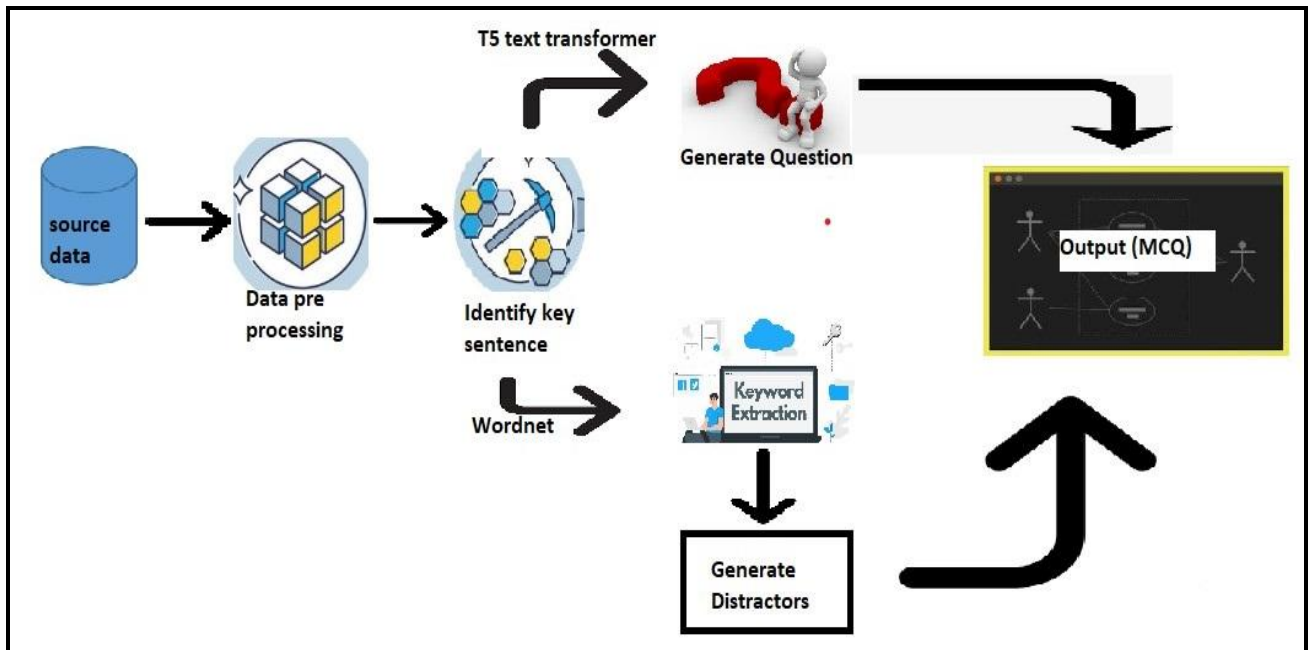


Fig. 1. System Architecture

The system architecture gives a brief explanation about the “Question generation using Natural Language Processing”. It represents the entities involved and the flow of generating a question.

#### V. ALGORITHMS

**T5 text transformer:** Text to text transformer model is mainly used to summarize the input text data. In this model it takes text as input and as a result it produces the summarized text as output. This model is pre trained on C4 dataset.

T5 is an encoder-decoder model that has already been trained on a variety of tasks that are both supervised and unsupervised and are each translated into a text-to-text format. The encoder-decoder is made up of 12 pair-blocks. Self-attention, a feed-forward network, and encoder-decoder attention are included in each block. T5 effectively handles a range of tasks right out of the box by appending a unique prefix to the input for each activity, e.g., for translation: translate English to German: ..., for summarization: summarize: .... Relative scalar embedding’s are used in T5. Both the left and the right padding can be done for encoder input. For NLP tasks.

A pre-trained T5 model that can perform a variety of tasks, including translation, summarization, question- answering, and classification, has been made available by Google.

**BERT:** BERT stands for Bidirectional Encoder Representations from Transformers. It is also mainly used to summarize text. There are two types of summarizations abstractive and extractive summarization. In abstract summarization, key points are rewritten, while in extractive summarization, the most important sentences/spans of a document are copied directly.

## VI. CONCLUSION

In conclusion, in this paper we learnt how to generate multiple choice questions from the given input text by using Natural Language Processing (NLP) Techniques. First, we identified the important sentences and keywords from the text, and we generated distractors using WordNet and Sense2vec. Here we have taken different sentences by varying lengths and observed the results. The output question consists of four options and an answer. Also, in future this work can be helpful in educational institutions for creating quizzes or assignments.

## REFERENCES

1. G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGr Automatic Question Generation using Discourse Cues by Agarwal, Shah and Prashanth.
2. Nltk: The Natural Language Toolkit by Edward Loper and Steven Bird.
3. Aditya S. (2018). Using Natural Language Processing for Smart Question Generation. <https://software.intel.com/en-us/articles/using-natural-language-processing-for-smart-question-generation>.
4. Aleena, Vidya – “Implementation of Automatic Question Paper Generator System”, International Research Journal of Engineering and Technology (IRJET).

- [1] Aldabe, I., De Lacalle, M. L., Maritxalar, M., Martinez, E., and Uria, L. (2006). Arikiturri: An Automatic question generator based on corpora and nlp techniques. In International Conference on Intelligent Tutoring Systems (pp. 584-594). Springer, Berlin, Heidelberg.
- [2] Gall MD. (1970). The use of questions in teaching. *Review of Educational Research*, 40: 707-720.
- [3] “Android Based Exam Paper Generator” Proceeding of the Second International Conference on Inventive Systems and Control (ICISC 2018) by Mrunal, Rushikesh, Shreyas Dorle, Uday Biradar, Kaustubh Kale.
- [4] Susanti, Y, Tokunaga, Nishikawa, H., and Obari, 2018. Automatic distracter generation for multiple Choice English vocabulary questions. *Research and Practice in Technology Enhanced Learning*.
- [5] “Representing Examination Question Knowledge into Genetic Algorithm”, IEEE Global Engineering Education Conference (EDUCON), 2014 by Noor Hasimah Ibrahim Teo, Nordin Abu Bakar and Moamed Abd Rashid.