

AUTOMATED QUESTION GENERATION: ENHANCING ASSESSMENT WITH TEXT SUMMARIZATION AND BLOOM'S TAXONOMY

Dhomala Aswani¹, Mokapwar Akansha Reddy², Yecharla Mounika³,
Vellenki Deekshitha Reddy⁴, Vallapureddy Shiva Charan Reddy⁵

¹Assistant Professor, Computer Science and Engineering, ACE Engineering College, India.

^{2,3,4,5}Computer Science and Engineering, ACE Engineering College, India.

ABSTRACT

This project presents a novel approach to automatically generating multiple-choice questions (MCQs) and subjective questions from input text or PDF files. Leveraging text summarization techniques, the system identifies key concepts within the text and replaces them with blanks to form MCQs. Distractor words are then utilized to create plausible answer options. For subjective question generation, the user selects the desired Bloom's taxonomy level, guiding the system in heavily summarizing the text and employing various transformer models to generate questions. The generated subjective questions are framed in alignment with Bloom's Taxonomy, enhancing the depth and complexity of the assessment. Through this innovative methodology, the project aims to streamline the process of question generation, facilitating efficient and comprehensive assessments in educational settings.

Keywords: Natural Language Processing, Text Processing, Sentence tokenization, Keyword Extraction, WordNet

1. INTRODUCTION

In the realm of education and assessment, the creation of high-quality questions for tests and exams is crucial for evaluating understanding and knowledge retention. However, crafting such questions manually can be time-consuming and labor-intensive. To address this challenge, our project aims to develop an automated system capable of generating both multiple-choice and subjective questions from given text inputs or PDF files.

The project leverages advancements in natural language processing (NLP), particularly transformer models, to enable efficient and effective question generation. By utilizing techniques such as text summarization and Important Noun Finder, our system can identify key concepts within the input text and formulate questions that assess comprehension and critical thinking skills. For multiple-choice question generation, the system employs text summarization and the Important Noun Finder to identify important words or phrases, which are then replaced with blanks to form the basis of questions. Distractor words are strategically selected to create plausible answer options, ensuring the validity and reliability of the assessment. In addition to multiple-choice questions, our project also facilitates the generation of subjective questions tailored to different Bloom's taxonomy levels. By allowing users to specify the desired cognitive complexity, the system heavily summarizes the input text and employs various transformed models to generate questions that align with Bloom's taxonomy. This approach ensures that subjective questions are appropriately framed to assess higher-order thinking skills, such as analysis, synthesis, and evaluation.

Through the development of this automated question generation system, we aim to streamline the process of assessment creation, saving educators valuable time and resources while enhancing the quality and diversity of questions used in educational settings. By harnessing the power of NLP and transformer models, our project seeks to contribute to the advancement of technology-enhanced learning and assessment practices.

2. METHODOLOGY

2.1 Data Collection and Preprocessing :

- The project begins with the collection of text inputs or PDF files containing the content from which questions will be generated. These documents are preprocessed to remove noise, such as special characters and formatting artifacts, and to tokenize the text into individual words or tokens.

2.2 Noun Extraction using YAKE:

- Noun extraction is a crucial step in identifying important concepts within the text for question generation. The project utilizes the YAKE (Yet Another Keyword Extractor) technique, specifically focusing on extracting nouns, which are typically the key entities in educational content.

YAKE employs a keyword extraction algorithm based on statistical features and domain-specific knowledge to identify relevant terms. By focusing on nouns, the technique ensures that the extracted keywords represent the most salient concepts in the text.

2.3 Text Summarization:

- After noun extraction, the project employs text summarization techniques to condense the input text into concise summaries. Text summarization helps to distill the main ideas and concepts from the text, making it easier to generate questions that focus on essential information. Extractive or abstractive summarization methods may be employed, depending on the specific requirements of the project and the nature of the input text.

1.4 Question Generation for Multiple-Choice Questions (MCQs):

- With the extracted nouns and summarized text, the system proceeds to generate multiple-choice questions.

Important nouns identified by YAKE are replaced with blanks to form the stem of the question. These blanks represent the missing concepts that students must identify or recall. Distractor words are selected based on context and domain knowledge to provide plausible but incorrect answer options. This ensures that the generated MCQs effectively assess students' understanding and eliminate guessing biases.

1.5 Subjective Question Generation:

- For subjective question generation, the system offers users the option to select the desired Bloom's taxonomy level to guide question complexity. The input text is heavily summarized to capture only the most critical information relevant to the selected cognitive level. Transformer models, fine-tuned on question generation tasks, are employed to generate subjective questions that align with Bloom's taxonomy. These questions aim to assess higher-order thinking skills such as analysis, synthesis, and evaluation.

2.6 Evaluation and Refinement:

- Generated questions are evaluated based on criteria such as relevance, clarity, and difficulty level. Feedback from educators and domain experts is solicited to refine the question generation process and improve the quality of the generated questions.

Continuous iteration and refinement of the methodology ensure that the automated question generation system produces high-quality questions that effectively assess students' knowledge and skills.

3. LITERATURE SURVEY

A literature survey for this project involves exploring existing research, publications, and advancements in the field of natural language processing (NLP), and question generation. The following summarizes key findings from relevant research papers and publications:

English Grammar Multiple-Choice Question Generation Using Text-To-Text Transfer Transformer - Peerawat Chomphoooyod , Atiwong Suchato , Nuengwong Tuaycharoen, Proadpran Punya bukkana (2023) [1]

In this paper, they automatically generated to reduce preparation time. In this system, the questions for MCQs are generated using a text generation model trained using the Text-to-Text Transfer Transformer (T5) architecture, a powerful deep learning model for performing text generation tasks, with a keyword and a part-of-speech (POS) template as the input for content and grammar topic control.

Multiple Choice Question Generation Using BERTXLNET - S. Adi Lakshmi, Rajesh Saturi, Anupriya Bharti, Meghana Avvari and Battu Bhavana-(2023) [2]

In this paper, The aim is to create accurate and pertinent questions using textual information. The sentences are summarized using the BERT XL NET algorithm after the data has been analyzed. Use the BERT XL NET summarizer to condense the cleaned-up text. The BERTSUM model, which is a fine-tuned version of BERT for extractive summarization, is used to summarize the input text. It displays the BERTSUM model's architecture.

Automated MCQ Generation Using Natural Language Processing - Pritam Kumar Mehta¹, Prachi Jain, Chetan Makwana, M Raut (2021) [3]

In Automated MCQ Generator, questions are generated automatically with the help of NLP. The text of any domain is provided as input to the system which is then summarized using the BERT algorithm. BERT is a deep learning-based technique for natural language processing, a pre-trained model from Google. Now the keywords are selected from the summarized text using the python keyword extractor (PKE) and accordingly mapping of a keyword is done with a sentence. The extraction of key word is done by python library RAKE .

Automatic Distractor Generation For Multiple-Choice English Vocabulary Questions - Yuni Susanti, Takenobu Tokunaga, Hitoshi Nishikawa, Hiroyuki Obari (2018) [4]

The use of automated systems in second-language learning could substantially reduce the workload of human teachers and test creators. study proposes a novel method for automatically generating distractors for multiple-choice English

vocabulary questions. The proposed method introduces new sources for collecting distractor candidates and utilises semantic similarity and collocation information when ranking the collected candidates.

Automatic Generation Of Multiple Choice Questions Using Wikipedia - Arjun Singh Bhatia, Manas Kirti & Sujun Kumar Saha (2019) [5]

Present a system for automatic generation of multiple choice test items using Wikipedia. A MCQ is composed of three elements: stem, target word and distractors. The system is using web information, specially Wikipedia, for generating questions and distractors. They have attempted to find the potential MCQ sentences with the help of existing questions of this domain.

4. COMPARISION ANALYSIS

S.No	Paper Title	Work done on paper	Future work	Drawback
1	Peerawat Chomphooyod , Atiwong Suchato , Nuengwong Tuaycharoen, Proadpran Punyabukkana- "English Grammar Multiple- Choice Question Generation Using Text-To-Text Transfer Transformer"-2023	Implemented T5 model for English grammar MCQs with keyword and POS templates, achieving good grammar control and diverse question types	Advanced fine-tuning, broadening question diversity, integrating contextual understanding, incorporating user feedback, and extending multilingual support for the T5 model.	Addressing overfitting, refining ambiguous case handling, exploring scalability for diverse rules, developing real-time interactivity, and ensuring ethical considerations for bias and inclusivity.
2	S.Adi Lakshmi, Rajesh Saturni, Anupriya Bharti, Meghana Avvari and Battu Bhavana- "Multiple Choice Question Generation Using BERT XLNET"-2023	summarization system using BERT XL NET and BERTSUM for concise question generation, complemented by keyword extraction through the YAKE Python library.	fine-tuning BERT XL NET and BERTSUM for enhanced summarization, exploring methods for diverse question generation, improving semantic understanding, and developing mechanisms to integrate user feedback for practical refinement.	potential over- reliance on extractive summarization, limited use of abstractive techniques
3	Pritam Kumar Mehta1, Prachi Jain, Chetan Makwana, MRaut- "Automated Mcq Generation Using Natural Language Processing."-2021	MCQ generator using NLP techniques, with BERT for text summarization, RAKE for keyword extraction, and a mapping mechanism to associate keywords with sentences for relevant question generation.	exploring BERT fine- tuning for domain- specific summarization, investigating advanced keyword extraction for improved precision, extending question types diversity.	potential limitations due to the dependency on pre- trained models like BERT, sensitivity of keyword extraction to text structure variations.
4	Yuni Susanti, Takenobu Tokunaga, Hitoshi Nishikawa, Hiroyuki Obari- "Automatic Distractor Generation For Multiple- Choice English Vocabulary Questions" - 2018.	Novel method for automated generation of distractors for multiple-choice English vocabulary question. The system introduces new sources for distractor candidates and utilizes semantic	the automated distractor generation method, exploration of additional linguistic features for enhanced accuracy.	potential drawbacks, including challenges in handling complex language nuances.

		similarity and collocation information during ranking.		
5	Arjun Singh Bhatia, Manas Kirti & Sujan Kumar Saha - "Automatic Generation Of Multiple Choice Questions Using Wikipedia" -2019	MCQ generation system utilizing Wikipedia for extracting potential MCQ sentences, consisting of stems, target words, and distractors.	exploration of additional knowledge sources beyond Wikipedia, and continuous improvement for better question variety and accuracy.	limitations in the system's ability to handle domain-specific nuances, potential biases from relying on existing questions.

5. FUTURE SCOPE

The System can be extended further in future with any of the following Ideas:

Adaptive Question Generation: Incorporate adaptive learning algorithms to tailor question generation based on individual student performance and learning preferences. This could involve dynamically adjusting the difficulty level of questions or customizing question topics to match each student's learning trajectory.

Integration with Learning Management Systems (LMS): Integrate the automated question generation system with popular learning management systems to seamlessly incorporate generated questions into course materials and assessments. This could streamline the workflow for educators and enhance the accessibility of generated questions for students.

Multimodal Question Generation: Extend the capabilities of the system to support multimodal inputs, such as images, audio, and video. This could involve generating questions based on visual or auditory content, thereby catering to diverse learning modalities and enhancing engagement.

By exploring these future scope areas and incorporating new ideas, the automated question generation project can continue to evolve and adapt to the changing needs of educators and learners, ultimately contributing to more effective and efficient educational practices.

6. CONCLUSION

In conclusion, the automated question generation project presents a promising approach to streamline the process of assessment creation in educational settings. By leveraging techniques such as noun extraction using YAKE, text summarization, and transformer models, the project successfully generates both multiple-choice and subjective questions based on selected Bloom's Taxonomy Level from input text or PDF files.

The developed system demonstrates the effectiveness of using natural language processing and machine learning techniques to automate the question generation process. By identifying important concepts within the text and formulating questions that assess comprehension and critical thinking skills, the system offers educators a valuable tool for creating diverse and engaging assessments.

Through evaluation and refinement, the project ensures the quality and validity of the generated questions, thereby enhancing the overall educational experience for students. By saving educators time and effort in question creation, the system enables them to focus more on instructional activities and personalized learning interventions.

7. REFERENCES

- [1] "English Grammar Multiple-Choice Question Generation Using Text-To-Text Transfer Transformer", an article by Peerawat Chomphoooyod, Atiwong Suchato, Nuengwong Tuay charoen, Proadpran Punyabukkana in 2023.
- [2] "Multiple Choice Question Generation Using BERT XLNET", an article by S.Adi Lakshmi, Rajesh Satari, Anupriya Bharti, Meghana Avvari and Battu Bhavana in 2023.
- [3] "Automated Mcq Generation Using Natural Language Processing.", an article by Pritam Kumar Mehta1, Prachi Jain, Chetan Makwana, MRaut in 2021.
- [4] "Automatic Distractor Generation For Multiple-Choice English Vocabulary Questions", an article by Yuni Susanti, Takenobu Tokunaga, Hitoshi Nishikawa, Hiroyuki Obari in 2018.
- [5] "Automatic Generation Of Multiple Choice Questions Using Wikipedia", an article by Arjun Singh Bhatia, Manas Kirti & Sujan Kumar in 2019.