International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Comparative Review for Automatic Question Generation Using NLP

PROF. ARCHANA R. GHUGE[1],NISHANT ANIL BHABAD[2],SHRADDHA DIPAK BODAKE[2],ANUJ RAVINDRA DAWKHAR[4], ASHWINI PRALHAD GARJE[5]

[1]*Assistant Prof, Department of Information Technology, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India* [2345]*Department of Information Technology, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India*

[1] ARCHANA.UGALE@PRAVARA.IN
[2] NISHANTBHABAD29@GMAIL.COM
[3] SHRADDHABODKE55@GMAIL.COM
[4] ANUJDAWKHAR8055@GMAIL.COM
[5] GARJEASHWINI1031@GMAIL.COM

Abstract**—** This project introduces an innovative approach to advance the field of automatic question generation using natural language processing (NLP), with a specific focus on Bloom's Taxonomy. With the increasing availability of resources and online learning platforms there is a need for efficient methods to create diverse and contextually relevant questions. The main goal of this project isto develop a system that can automatically generate questions using Natural Language Processing (NLP) techniques aligned with first three cognitive levels of Bloom's Taxonomy: remembering, understanding, and applying. This project will make a contribution to the field of NLP by providing a framework for automatic question generation. The project follows stages; preprocessing the input text identifying concepts and information creating question rules and generating different versions of questions based on these rules. This project utilizes NLP techniques such as Named Entity Recognition (NER) Part of Speech tagging (POS), syntatic analysis and Discourse analysis. The overarching goal is to provide educators, content creators, and learners with an efficient and intelligent tool for generating questions that enhance comprehension and critical thinking. By automating this process, the project seeks to save time and effort while improving the overall learning and assessment experience.

Keywords: Automatic Question Generation, Bloom's Taxonomy, Natural Language Processing, Quality Assessment, Learning Assessment, Comprehension, NER, POS tagging.

1 INTRODUCTION

Automatic Question Generation (AQG) using Natural Language Processing (NLP). AQG is an emerging field that leverages advanced NLP techniques to automatically generate questions from given text or data. This innovative technology has the potential to revolutionize various sectors by enhancing educational tools, automating content creation, and improving user interaction in customer service. This system requires the input of a target topic and will generate questions whose answers will be found in the paragraph itself. This project seeks to save time and effort taken to generate relevant and diverse questions while improving the overall learning and assessment experience.

2 LITURATURE SURVEY

In 2021, Danny C. L. [1] – ICALT Conference The authors proposed an Automatic Question Generation (AQG) system intended for repeated testing to enhance student learning outcomes. The model utilized historical student interaction data to iteratively generate questions tailored for reinforcing learning and tracking progress. Their approach also integrated adaptive assessment principles to personalize difficulty levels, making it beneficial in e-learning platforms.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

In 2021, Shivali Joshi [2] – IJISRT This work presented a system that generates questions from text content using Natural Language Processing (NLP) techniques. The system aligns question generation with Bloom's Taxonomy, ensuring a pedagogical framework in categorizing the cognitive complexity of questions. It focuses on deriving questions from input paragraphs, applying POS tagging, parsing, and selecting appropriate question words for generation.

In 2019, Kalpana B. Khandale [3] – IOSR-JCE Though not directly about AQG, this paper compared POS taggers for identifying word similarities. The work is relevant to AQG as POS tagging is essential in syntactic analysis for extracting valid answer spans and formulating grammatically correct questions. The authors emphasize the importance of accurate linguistic preprocessing for downstream NLP tasks.

3 PROPOSED SYSTEM

This section clearly outlines architecture, methodology and innovation of proposed Automatic Question Generation (AQG) system.[8]

Develop a system that can automatically generate questions using Natural Language Processing (NLP) techniques aligned with first three cognitive levels of Bloom's Taxonomy: Knowledge, Comprehension and Application level.

Save time and effort taken to generate relevant and diverse questions while improving the overall learning and assessment experience.

Utilizes NLP techniques such as Named Entity Recognition (NER), Part of Speech tagging (POS), syntactic analysis, Discourse analysis and Context analysis to generate most relevant question to enhance learning process.

Use integrated approach to automate the process of question generation. The integrated approach will comprise of ML model and Rule Based mode.

Contextual Understanding: Enabled through transformer-based encoding

3.1 COMPARATIVE ANALYSIS OF EXISTING SYSTEM

Automatic Question Generation (AQG) typically involves transforming a given text into meaningful questions. Current methods often use pre-defined templates, neural networks, or rule-based systems, but these models tend to miss nuances in context and fail to adapt to various domains or question types. This theory proposes a **Context-Aware Adaptive Question Generation (CA-AQG)** model that combines deep contextual understanding, cognitive mimicry, and adaptive learning to enhance the quality and relevance of generated questions. The model aims to provide human-like questioning by considering the interplay between context, user knowledge, and question utility.

Automatic Question Generation (AQG) using NLP

	Earlier Approaches	Modern Approaches
NLP Techniques	Rule-based, shallow parsing	Deep learning, neural networks
Linguistic Resources	Dependency trees named entities	Pre-trained language models
Question Types	Mostly fact-based	Wide variety
Contextual Understanding	Limited	Better
Evaluation Methods	Manual evaluation	Automated metrics

Figure 1: Existing System Vs Proposed System

3.2 NLP Techniques:

Automatic Question Generation (AQG) has evolved from rule-based methods with shallow parsing to modern deep learning approaches using neural networks. Early systems relied on dependency trees and named entity recognition, while modern techniques use pre-trained models like BERT, T5, and GPT. Initially focused on generating fact-based questions, AQG now creates a variety of question types, such as why and how. Contextual understanding has improved significantly through attention mechanisms, enabling better comprehension of text. Evaluation has shifted from manual assessments to automated metrics, allowing for more scalable and efficient evaluation of generated questions. [3].

3.3 Linguistic Resources:

- Earlier: Relied on dependency trees and named entity recognition.
- Modern: Leverages pre-trained language models like BERT, T5, GPT. [4].

3.4 Contextual Understanding:

Earlier models had limited context understanding because they processed inputs sequentially, which made it difficult to capture long-range dependencies in data. Modern models like Transformers use Attention Mechanisms to dynamically focus on different parts of the input sequence, allowing them to better understand context by considering all parts of the input at once. This leads to improved performance in tasks like translation, summarization, and question answering. [5].

4 PROPOSED APPLICATION:

The proposed system aims to bridge the limitations of earlier models by offering a holistic suite of features:

Educational Assessments: Generates quizzes, tests, and exams for schools, colleges, and online platforms.

Interview Preparation: Produces mock interview questions for job aspirants.
Self-Assessment: Provides learners with practice questions to test their knowledge.
Medical Training: Supports the training of medical professionals by generating scenario-based questions for clinical case studies and patient simulations.
Content Recommendation Systems: Enhances content recommendation systems by suggesting questions that align with users' areas of weakness or interest.

5 SUMMARY

This comparative analysis build upon the strengths of transformer-based architectures by utilizing answer-aware input formatting to improve question relevance.

6 OBJECTIVE

The primary objective of this research is to design and develop an efficient system for **Automatic Question Generation (AQG)** using advanced Natural Language Processing (NLP) techniques. Specific objectives include:
To analyze existing AQG techniques and identify their strengths and limitations in terms of accuracy, contextual relevance, and scalability.
To develop an answer-aware question generation model using architectures , capable of generating high-quality questions from input text and answer spans.
To ensure diversity in the types of questions generated.
To improve overall learning and assessment experience.
To enhance the quality and variety of generated questions through data preprocessing, augmentation, and post-processing strategies.
To demonstrate the practical applicability of the system in domains such as education, intelligent tutoring, and QA system enhancement.

7 SYSTEM ARCHITECTURE

The diagram represents a flowchart for processing an input text paragraph and generating a question from it.
step-by-step explanation of the process:
1. Input Text Paragraph: The system starts with an input, which is a text paragraph.
2. Pre-Processor: This section breaks down the input paragraph into several preprocessing steps
• Tokenization: The text is broken into smaller units, typically words or tokens.
• Removing Stopwords: Common words (like "the", "is", etc.) that do not contribute much meaning are removed.
• POS (Part-of-Speech) Tagging:Each word in the sentence is tagged with its part of speech (e.g., noun, verb, adjective).

3. Further Processing: After preprocessing, the text undergoes more advanced linguistic analysis: Semantic Role Labeling: Identifies the role that words or phrases play in a sentence (e.g., who did what to whom).
4. Customized NER (Named Entity Recognition): This step identifies named entities in the text (such as people, organizations, or locations) but is customized to fit the specific requirements of the system.

5. Rules and Template: Based on the processed information, specific rules and templates are applied to frame a question from the text. This is guided by predefined patterns or rules.

6. Question Framing: Using the rules and templates, the system con- structs a question based on the understanding of the input text.

7. Display Generated Question: Finally, the generated question is displayed, completing the process.

8 RESULT AND FUTURE SCOPE

Currently, our modules do not achieve the best performance across all question categories.In addition, in order to improve the performance of paragraph based questions in all templates, we need to investigate how to better use the paragraph-level information
Introduce some kind of machine learning algorithms to improve the accuracy and to evaluate the results efficiently
Information conveyed from one sentence to other is a problematic issue that needs to be worked upon Implement method for sentences containing possessive and demonstrative pronouns as subject. Implement method for questions framed with Which, Whose and How.
Making system able to generate the question in various domains and languages.
Despite the successes, some challenges were identi- fied. For instance, variations in voice clarity and back- ground noise can affect the accuracy of the voice acti- vation feature. Future enhancements may include refin- ing the voice recognition algorithms to improve perfor- mance in various conditions, as well as considering the integration of additional safety features, such as auto- matic alerts based on unusual user behavior.

9 CONCLUSION

This Project represents a significant advancement in the field of automatic question generation, harnessing the power of Natural Language Processing techniques, specifically tailored to Bloom's Taxonomy. In the current educational landscape, characterized by rapid technological evolution and abundant online resources, there is a growing demand for efficient methods to create diverse and contextually relevant questions. This project addresses this need by devising a systematic approach aligned with the first three cognitive levels of Bloom's Taxonomy: remembering, understanding, and applying. Through meticulous stages, such as pre-processing the input text, identifying key concepts, and employing advanced NLP techniques like Customized Named Entity Recognition (NER), Part of Speech tagging (POS), syntactic analysis, and Discourse analysis, this project pioneers a comprehensive framework for automatic question generation. Its primary objective is to empower educators, content creators, and learners with an intelligent tool capable of generating questions that enhance comprehension and foster critical thinking skills.

10 REFERENCES:

[1] DannyC.L.Tsai, Anna Y.Q. Huang, OwenH.T.LuandStephenJ.H.Yang, "Au tomatic Question Generation for Repeated Testing to Improve Student Learning Outcome", In 2021 International Conference on Advanced Learning Technologies (ICALT) IEEE DOI10.1109/ICALT52272.2021.00108

[2] Shivali Joshi,Parin Shah and Sahil Shah,"Automatic Question Paper Generation, according to Bloom's Taxonomy, by generating questions from text using Natural Language Processing", In International Journal of Innovative Science and Research Technology—Vol.6 4, April–2021—ISSN No:-2456-2165.

[3] Deena Gnanasekaran, Raja Kothandaraman, and Kannan Kaliyan,"An Automatic Question Generation System Using Rule-Based Approach in Bloom's Taxonomy", In Recent Advances in Computer Science and Communications, 2021, 14, pp. 1477-1487.

[4] U. A. Mardand, and A. Gaikwad, "A survey onautomatic question paper generation system", Int.Adv. Res. J. Sci. Eng. Technol., Vol. 4, no. 4, pp. 18-20,2017

[5] Kalpana B. Khandale, Ajitkumar Pundage ,C. Namrata Mahender, Similarities in words Using Different Pos Taggers, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, PP 51-55 Published: 2019

[6] M. Blstak, and V. Rozinajova, "Machine learning approach to the process of question generation", In International Conference on Text, Speech, and Dialogue, 2017, pp. 102-110. [http://dx.doi.org/10.1007/978-3-319-64206-2 12]

[7] P. Garg, and C. S. Bedi, "Automatic question generation system from punjabi text using hybrid approach", Int. J. Comput. Trends Tech., Vol. 21, no. 3, pp. 130-133, 2015.[http://dx.doi.org/10.14445/22312803/IJCTT-V21P125]

[8] Ghuge Archana, Rajesh Gupta, Mohammed Usman, Pratibha V Kashid, Lekshmi Mohan, Vishakha A. Gaidhani , (2025). Artificial Intelligence and IoT in Retail Marketing: Innovations in Smart Stores and Personalized Shopping. Journal of Informatics Education and Research (JIER).