



MCQ GENERATOR USING TEXT SUMMARIZATION

Y. RAJALAKSHMI(Asst.prof)
Information Technology
(JNTUH)
G.Narayanamma Institute of
Technology and science(for women)
(JNTUH)
Hyderabad, India
yedlarajalakshmi@gmail.com

KONDAM ABHINAYA
Information Technology
(JNTUH)
G.Narayanamma Institute of
Technology and Science(for women)
(JNTUH)
Hyderabad, India
kondamabhinayareddy@gmail.com

[POTU](#) MANJULA
Information Technology
(JNTUH)
G.Narayanamma Institute of
Technology and Science(for women)
(JNTU)
Hyderabad, Indi
manjulareddypotu24@gmail.com

MEGAVATH KEERTHI
Information Technology
(JNTUH)
G.Narayanamma Institute of
Technology and Science(for women)
(JNTUH)
Hyderabad, India
megavathkeerthi123@gmail.com

MANYA DIVYA
Information Technology
(JNTUH)
G.Narayanamma Institute of
Technology and Science(for women)
(JNTUH)
Hyderabad,India
manyadivyaraj333@gmail.com

Abstract— Nowadays, Universities and other educational institutes are increasingly shifting towards online examinations. The pattern of assessment is majorly in the form of objective assessment (MCQs). To generate MCQs there is a lot of manual work involved in it which takes quite a long time to frame proper questions. So, to reduce the manual work and create MCQs in a simple way, we are building a model of generating MCQs using NLP. Here, we are creating a platform where the user can upload an input text related to basic sciences and the MCQs will be automatically generated by the system. In this process, text is first summarized using the BERT sum algorithm and then keywords are selected from the summarized text. Then, sentence mapping is done for generating MCQs. Wordnet, a lexical database for English, is used to generate distractors, which are options for the queries. We will also create a web interface wherein we can upload the text as input and the MCQs generated will be displayed as output. The generated MCQs can be copied or downloaded based on the user's requirements.

Keywords- BERT(Bidirectional Encoder Representation from Transformers), Wordnet, Natural Language Processing(NLP), MCQs, Summarization

I. INTRODUCTION

Online assessments like multiple choice questions are being very useful for conducting examinations nowadays. To create MCQs first a person needs to select a topic, go through

the entire context and then start framing questions with appropriate options. To reduce this lengthy process, we are engaged in the rapidly developing field of natural language processing. We intend to create a computer application system that can assist users in self-evaluation and eliminate any reliance on mentors. This computerised programme will enable us to eliminate human labour. We may also save a tonne of time if we are aware of the right questions to pose in light of the text input that has been provided. Students can input a basic sciences-related material here, and based on it, they will receive a series of questions and answers from which they can perform a self-analysis. With the aid of NLP, questions are automatically generated using the Automated MCQ Generator. The system receives the text as input, which is subsequently condensed using the BERT algorithm. A pre-trained model from Google called BERT (Bidirectional Encoder Representation from Transformers) uses deep learning to interpret natural language. After the text has been condensed, the keywords are chosen from it using the Python Keyword Extractor (PKE), and then a keyword is appropriately mapped with a sentence. A MCQ choice will include this keyword. The primary goal is to generate the pertinent distractors once the keywords have been established. The wordnet 2 method is used to produce distractors. An API called Wordnet is used to discover a word's correct meaning. So, the good and reliable distractors are generated. By eliminating the need for manual question development, this method saves time.



II. LITERATURE REVIEW

The major contributions by various authors are studied and analysed.

In the existing system, questions are generated from summarized text which is given as input. After summarization of text keywords are extracted from the summarized text then finds the co-occurrence matrix of the words using the RAKE algorithm and finds the score of the words and the one which is having the highest score will be taken as the keyword of the sentence and then generate distractors for that keyword using word net. The algorithm used in the existing system is RAKE algorithm, though the algorithm is efficient and accurate. The stop word list used in RAKE is not exhaustive, it would treat continuous exceedingly long phrases and extensive text as a phrase. Multi-word expressions that contain stop-words could be missed. For instance, if the word "good" is on the stop word list, it can be overlooked when a company called "Good Day" is mentioned.

III. ABBREVIATIONS AND ACRONYMS

BERT - Bidirectional Encoder Representation from Transformers

YAKE - Yet Another Keyword Extraction

NLP – Natural language processing

MCQs – Multiple Choice Questions

IV. DESIGN AND SIMULATION

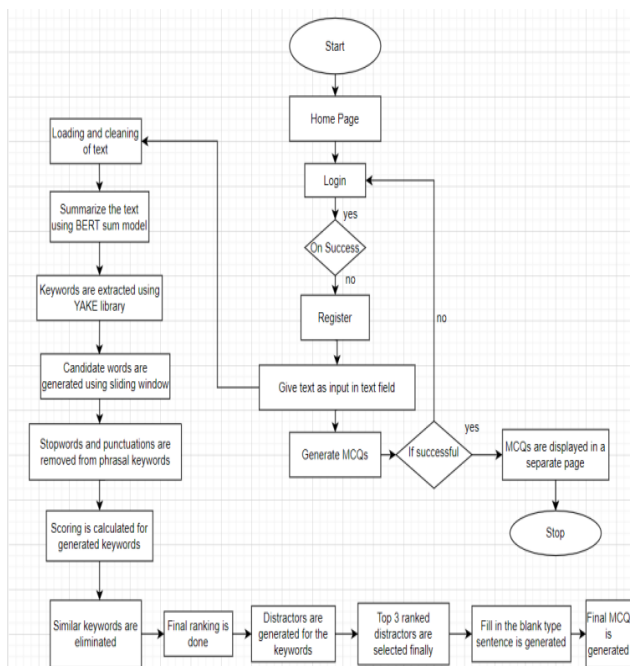


Fig-1 Block diagram for MCQ Generator

Loading and Cleaning

Raw text, or input material from a domain for which questions are to be generated, must be loaded before stopwords may be removed.

summarise the information

Using the BERT Algorithm, the material is summarised. BERT (Bidirectional Encoder Representations from Transformers) is a method for processing natural language based on neural networks. It is a Google open-sourced, pre-trained model. It makes it easier for computers to comprehend language a little more as humans do. The BERTSUM model, which is a refined version of Extractive summarization with BERT, is used to summarise the input text.

Extraction of keywords

The content is summarised, and then the sentence's keywords are chosen. Yet Another Keyword Extraction (YAKE), a common keyword extraction method, uses a list of stopwords and phrase delimiters to identify the most crucial words and phrases in a document.

Generation of Candidates and Preprocessing

Using spaces and special characters, the sentences are divided into words. We choose the keyword's maximum length before it is formed. If we choose a maximum length of three, a sliding window is used to generate candidate phrases of one, two, and three grammes. Then, we eliminate sentences that contain punctuation. Additionally, stop words at the start and conclusion of phrases are eliminated.

Candidate Scoring

YAKE uses few features to quantify how good each word is and to score them. They are Word Relatedness to Context, Word Position, Word Frequency, and Casing. Finally scoring is calculated. After calculating scoring final ranking is done. Based on the users requirements few Keywords are taken into account.

Generation of distractors

A good deterrent is one that closely resembles the key but is not the key. Therefore, the Wordnet technique is employed to generate distractions. The Princeton-developed WordNet lexical database for the English language is a component of the NLTK corpus. By means of linguistic linkages, words in the WordNet network are connected to one another. Meronym, holonym, hypernym, and hyponym are a few examples of these linguistic connections. Synsets, which are collections of cognate synonyms made up of nouns, verbs, adjectives, and adverbs, are how WordNet organises and saves synonyms. In a synset, every word has the same meaning. A Lesk algorithm is used to generate the MCQs.

Sentence Mapping

Following keyword selection, each keyword's sentence is mapped, i.e. The term from the condensed text is extracted from the sentences that relate to each keyword.

Question Generation

Finally the Questions are generated with the keyword and also required options. The user can also copy or download MCQs



Fig-2 Home Page

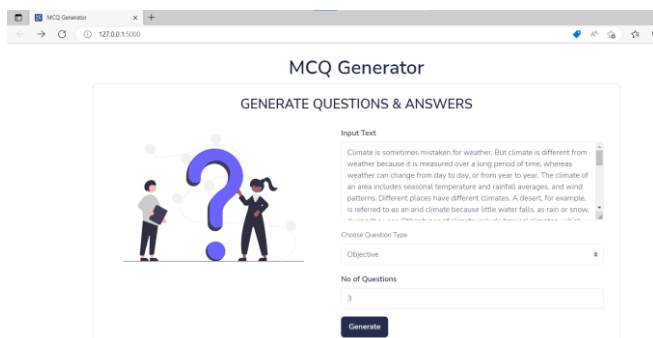
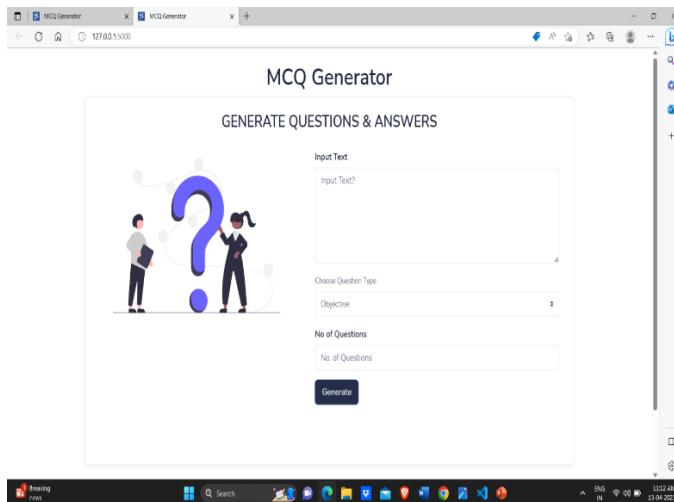


Fig-3 Text is given as input

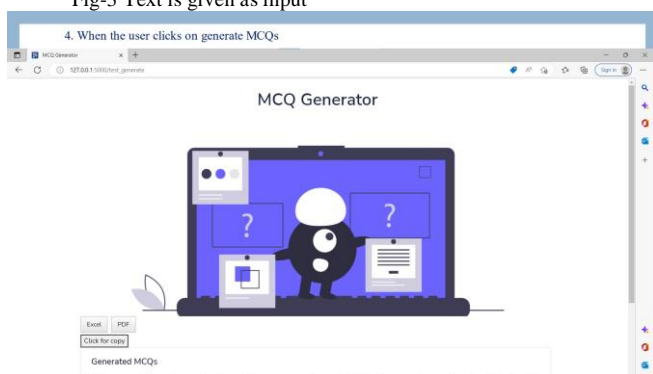


Fig-4 when the user clicks on generate MCQs

V. RESULTS

Using this project questions are produced with distractors. Hence it can be used for teachers to ease their work during the assessment. It is used as a time effective method .Easy to use and cost effective method This project is limited to only

general science field. In future we can make it applicable to other areas like technical fields as C, Java, Python.

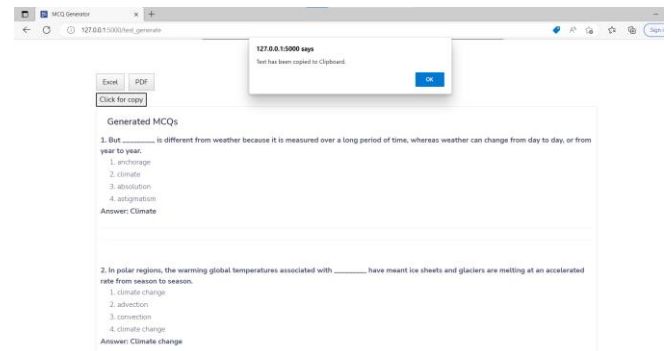


Fig-5 results for MCQ Generator Using Text Summarization

VI. CONCLUSION

The suggested solution guarantees an automated MCQ Generator; using NLP (Natural Language Processing), questions are generated automatically. First, it cleans a text input provided by the user. The BERT (Bidirectional Encoder Representation from Transformers) algorithm, a deep learning-based method for natural language processing, is used to condense the cleaned text. Now the keywords are selected from the summarized text using the YAKE algorithm and accordingly mapping of a keyword is done with a sentence. A MCQ choice will include this keyword. Distractors are created using a wordnet 2 technique to generate the remaining possibilities. To determine the correct meaning of a word, an API called Wordnet is employed. Thus, MCQs are generated.

ACKNOWLEDGMENT

It is our proud privilege to express the feelings of gratitude to several people who helped us in completing this work We express our heartfelt gratitude to Dr. I. Ravi Prakash Reddy, Head of the Department, IT department, GNITS for his constant guidance and moral support. We would like to express our sincere thanks to Dr. K. Ramesh Reddy, Principal, GNITS for providing working facilities in the college. Finally, we would like to thank all the faculty and staff of IT department who helped us during our project and also our parents and friends for their cooperation in completing the project.

REFERENCES

[1] "Automatic Multiple Choice Question Generation From Text: A Survey," IEEE Transactions on Learning Technologies, vol. 13, no. 1, pp. 14–25, 1 January–March



2020, doi: 10.1109/TLT.2018.2889100. D. R. CH and S. K. Saha.

[2] Int. J. Signal and Imaging Systems Engineering, Vol. 10, Nos. 1/2, pp. 54–62, Santhanavijayan, A., Balasundaram, S.R., Hari Narayanan, S., Vinod Kumar, S., and Vignesh Prasad, "Automatic generation of multiple-choice questions for e-assessment."

[3] A real-time multiple-choice question creation for language testing: A preliminary investigation, Ayako Hoshino and Hiroshi Nakagawa, EdAppsNLP 05: Proceedings of the Second Workshop on Building Educational Applications Using NLP, 2005.

[4] Deepshree S. Vibhandik and Rucha C. Samant "Automatic / Smart Question Generation System for Academic Purpose", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 4, Issue 4, July-August 2015.

[5] "Automatic distractor generation for multiple-choice English vocabulary questions", RPTEL 13, 15 (2018). Susanti, Y., Tokunaga, T., Nishikawa, H. et al. <https://doi.org/10.1186/s41039-018-0082-z>.