International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

TEXT SUMMARIZATION USING NLP

[1]Mr K.KOTESWARA CHARI, [2] AMBERI JAYANTH, [3] BOREDHA ANUSHA, [4]CHALLA MOUNIKA

[1](Assistant Professor) , CSE. Teegala Krishna Reddy Engineering College Hyderabad.

[2,3,4]B,tech , scholar , CSE. Teegala Krishna Reddy Engineering College Hyderabad.

ABSTRACT

The main objective of this project is to develop an application to generate the questions from a given passage or paragraphs. This project is an application of Natural Language Processing (NLP). The application generates different types of questions such as MCQ's. The current project helps the teachers to prepare questions to the examinations, quizzes etc. It also The main objective of this project is to develop an application to generate the questions helps the students to get summary from the text they provide.

The summary would be generated from given text then, the application identifies key concept in summary. It identifies key words from sentences and generates MCQ's. The other options other than correct option called as Distractors. The application also paraphrases the input text. It allows the extraction of summary from the text is done by T5 Transformer Technique. We generate distractors using WordNet. The key words can be extracted by using KeyBERT Technique. The questions are displayed or GUI is developed with the help of the Gradio (A user friendly Web-Interface). The application majorly targets every organization. The application develops Frequently Asked Questions (FAQ's) of the customers to the particular organizations, which in terms gives the customer more information of the organizations. This application overrides the traditional method of preparing questions, and the complexity involved in making questions.

KEY WORDS:- Natural Language Processing (NLP), T5 Transformer, KeyBERT, WordNet, Gradio

1 . INTRODUCTION

Objective and subjective question generation using Natural Language Processing (NLP) is a fascinating and complex field that involves using computational techniques to

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

automatically generate questions that test a person's knowledge or understanding of a particular topic. Objective questions are those that have a specific correct answer, such as multiple-choice questions, true/false questions, and fill-in-the-blank questions. In contrast, subjective questions are open-ended and require a subjective response, such as essay questions, short-answer questions, and opinion-based questions. NLP techniques can be used to generate both objective and subjective questions.

Objective questions can be generated by analyzing text and identifying important concepts or key facts that can be used as the basis for a question. On the other hand, subjective questions require a deeper understanding of the context and the ability to generate questions that elicit a personal response. There are several challenges involved in generating high-quality questions using NLP, including the need to understand the underlying context and the ability to generate questions that are appropriate for the intended audience. Nevertheless, with the rapid advancement of NLP technology, the field of question generation is poised to make significant strides in the years to come. Speech-to-Text with Natural Language Processing (NLP) is an exciting field that involves converting spoken words into written text using computational techniques.

This technology is widely used in various applications such as voice assistants, automated transcription, and closed captioning for video content. Speech-to-Text with NLP involves multiple stages of processing, including signal processing, feature extraction, and language modeling. The initial step is to convert the raw audio signal into a digital format that can be processed by a computer. Next, features such as frequency and amplitude are extracted from the audio signal to capture relevant information about the speech. Once the audio signal has been processed, NLP techniques are used to transcribe the spoken words into written text. This involves analyzing the structure of the language and using algorithms to determine the most likely words and phrases that were spoken. NLP is also used to improve the accuracy of speech-to-text systems by modeling the language context and using contextual information to make more accurate predictions about what was said. This is particularly useful in cases where the speech contains ambiguous or homophonic words that could be interpreted in different ways. Overall, Speech-to-Text with NLP is a powerful technology that has numerous applications in various industries. It enables us to interact with computers using our natural language, making it more accessible to a broader range of users. With continued advances in NLP technology, we can expect even more accurate and efficient Speech-to-Text systems in the future. Text summarization using Natural Language Processing (NLP) is a technique that involves automatically generating a concise and coherent summary of a longer text. The purpose of text summarization is to distill the most important information from a text and present it in a shorter format.

There are two main approaches to text summarization using NLP: extractive and abstractive. Extractive summarization involves selecting the most important sentences or phrases from the original text and combining them to form a summary. Abstractive summarization, on the other hand, involves generating new sentences that capture the essence of the original text. Extractive summarization using NLP involves several steps, including sentence segmentation, word tokenization, and sentence scoring. Sentence segmentation involves breaking the text into individual sentences, while word tokenization involves breaking each sentence into individual words. Sentence scoring involves assigning a score to each sentence based on its importance and relevance to the overall topic. Abstractive summarization using NLP is more complex and involves natural language generation techniques to generate new sentences that convey the meaning of the original text. This approach involves understanding the underlying meaning and context of the text and usingalgorithms to generate new sentences that capture the essence of the original text.

Text summarization using NLP has numerous applications in various fields, including news summarization, document summarization, and email summarization. It can help users quickly identify the most important information in a text and save time by avoiding the need to read lengthy documents. As NLP technology continues to advance, we can expect even more accurate and efficient text summarization systems in the future.

2 . LITERATURE SURVEY

Objective and subjective question generation is an active research area in the field of natural language processing. A variety of approaches have been proposed to generate both types of questions using NLP techniques. Some approaches use rule-based methods, while others use machine learning algorithms to generate questions. A study by Li et al. (2020) proposed a neural network-based approach for generating multiplechoice questions.

Their approach used a combination of convolutional and recurrent neural networks to encode the input text and generate candidate answers for each question. They evaluated their approach on a dataset of history questions and achieved promising results. Similarly, Wang et al. (2019) proposed a method for generating subjective questions that can be answered with a short sentence. Their approach used a sequence-to-sequence model with attention mechanisms to generate questions from a given text.

They evaluated their approach on a dataset of reading comprehension questions and achieved competitive results compared to existing methods. Text-to-summary is another active research area in natural language processing that aims to generate concise summaries of longer texts. A variety of approaches have been proposed to generate summaries using NLP techniques, including extractive and abstractive methods.

A study by Nallapati et al. (2017) proposed a sequence-to-sequence model with attention mechanisms for abstractive summarization. They evaluated their approach on the CNN/Daily Mail dataset and achieved state-of-the-art results. Another study by Zhang et al. (2018) proposed an extractive summarization approach based on graph neural networks.

They evaluated their approach on a dataset of scientific articles and achieved competitive results compared to existing methods. Speech-to-text is a well-established application of natural language processing that involves converting spoken language into written text. Several approaches have been proposed to perform speech-to-text using NLP techniques, including Hidden Markov Models (HMMs), Deep Neural Networks (DNNs), and Recurrent Neural Networks (RNNs).

A study by Hinton et al. (2012) proposed a deep neural network-based approach for speech recognition that outperformed traditional HMM-based approaches.

They used a Deep Belief Network (DBN) to pretrain a Deep Neural Network (DNN) for acoustic modeling, achieving stateof-the-art performance on a dataset of spoken digits. Similarly, Graves et al. (2014) proposed a recurrent neural network-based approach for speech recognition that used Connectionist Temporal Classification (CTC) to align the predicted text with the spoken input. They evaluated their approach on a dataset of spoken sentences and achieved stateof-the-art results compared to existing methods.

Overall, these studies demonstrate the effectiveness of various NLP techniques for objective and subjective question generation, text-to-summary, and speech-to-text applications. As NLP technology continues to advance, we can expect even more accurate and efficient systems in these areas.

3 . SYSTEM DESIGN

3.1 SOFTWARE DEVELOPMENT LIFE CYCLE (SDLC)

3.1.1 AGILE MODEL

The Agile model is a popular iterative and incremental approach to software development that emphasizes flexibility, collaboration, and customer satisfaction. The Agile model is based on the Agile Manifesto, which values individuals and interactions, working software, customer collaboration, and responding to change over processes and tools. In an Agile model, the development process is divided into short iterations or sprints, typically lasting 1-4 weeks, during which a working software increment is delivered. The process starts with gathering and prioritizing requirements, which are then broken down into smaller tasks and assigned to team members for implementation. During each sprint, the development team works closely together to design, implement, and test the software increment. Regular meetings are held, such as daily stand-ups, sprint planning meetings, and sprint reviews, to discuss progress, identify issues, and plan for the next sprint. The goal of each sprint is to deliver a working increment that meets the customer's requirements and can be tested and evaluated. The Agile model is particularly well-suited for NLP-based systems such as objective and subjective question generation, text to summary, and speech to text using NLP. This is because these systems often require a flexible approach to accommodate changing requirements or the need for rapid experimentation and prototyping. In an Agile model, the development team can work closely with the stakeholders, such as domain experts, data scientists, and end-users, to refine the requirements and functionality of the system over time. The iterative approach of Agile allows for continuous testing, feedback, and improvement, which can help to ensure that

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

the NLP system meets the user's needs and expectations. Overall, the Agile model offers a collaborative and flexible approach to software development that can be well-suited for the requirements of NLP-based systems.

3.2 UML DIAGRAMS

The purpose of a UML diagram is to enable customers and developers to view software systems from various angles and levels of abstraction. UML diagrams are frequently produced via visual modelling software. A use case, in its most basic form, can be defined as a particular way of utilizing the system from the User's (actor's) perspective. A use case could be described in more depth as:

• a pattern of behavior the system demonstrates

•a series of connected transactions carried out by an actor and the system;

• Providing the actor with something of value.

USE CASES OFFER A WAY TO

• record system requirements;

• interact with end users and subject-matter experts;

• test the system.

The best way to identify use cases is to look at the actors and specify what each one of them will be able to do with the system. A group of use cases is normal since one use case often cannot satisfy all of a system's requirements. Together, these use cases outline all possible system usage scenarios. A UML system is represented using five separate views, each of which offers a radically different viewpoint on the system. A set of diagrams that characterize each view is as follows.

UML MODEL VIEW

• This view depicts the system from the viewpoint of the user.

• The analytical representation gives a user's-eye view of a usage scenario.

STRUCTURAL MODEL VIEW

• In this model, the system itself provides the data and functionality.

• The static structures are modelled in this model view.

BEHAVIORAL MODEL VIEW

IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

• It depicts the interactions between various structural elements described in the user model and structural model view, representing the dynamic of behavioral as components of the system.

IMPLEMENTATION MODEL VIEW

• In this, the system's structural and behavioral components are depicted as they will be when completed.

ENVIRONMENTAL MODEL VIEW

• This represents the structural and behavioral aspects of the system will be used. UML is specifically constructed through two different domains they are:

• UML Analysis modeling, which concentrates on the system's user model and structural model views.

• UML design modeling, which emphasizes behavioral design.

Fig 4.2.1:Usecase diagram 4.2.2 Class diagram: It involve classes such as TextProcessor, Summarizer, QuestionGenerator, Document, and Sentence. These classes would interact to analyze input text, generate a condensed summary, and formulate relevant questions based on the content. Additionally, they would manage the representation and manipulation of documents and sentences within the system. Objective questions Subjective questions Question generation Summarized text or speech Summarization user system

3.2.1 USECASE DIAGRAM

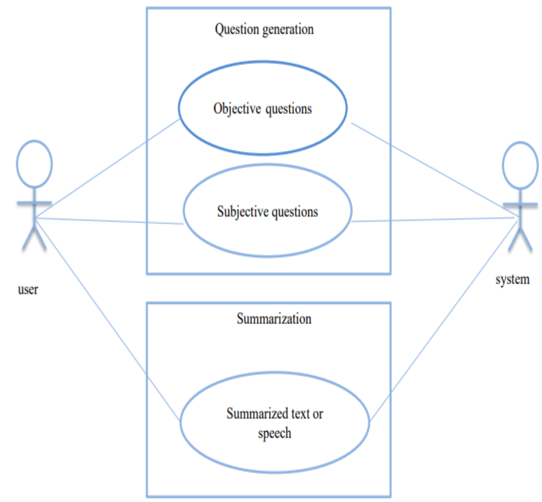

Fig 1: Usecase diagram

3.2.2 CLASS DIAGRAM:

It involve classes such as TextProcessor, Summarizer, QuestionGenerator, Document, and Sentence. These classes would interact to analyze input text, generate a condensed summary, and formulate relevant questions based on the content. Additionally, they would manage the representation and manipulation of documents and sentences within the system.

International Journal For Advanced Research
In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

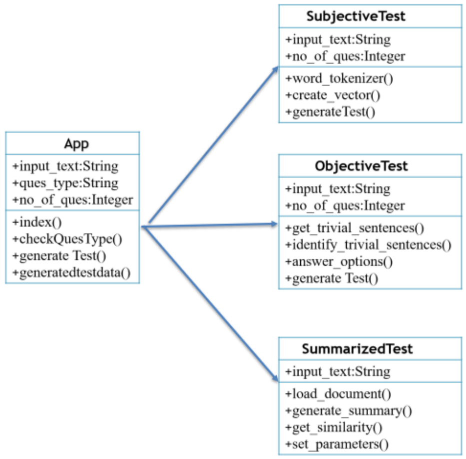

Fig: 2 :Class diagram

3.2.3 ACTIVITY DIAGRAM:

It illustrate the sequential processes involved, such as text preprocessing, summarization, question formulation, and output generation, depicting the flow of control between these activities. It would visually represent the steps and decision points in the system's workflow, demonstrating how input text is transformed into a summary and relevant questions.

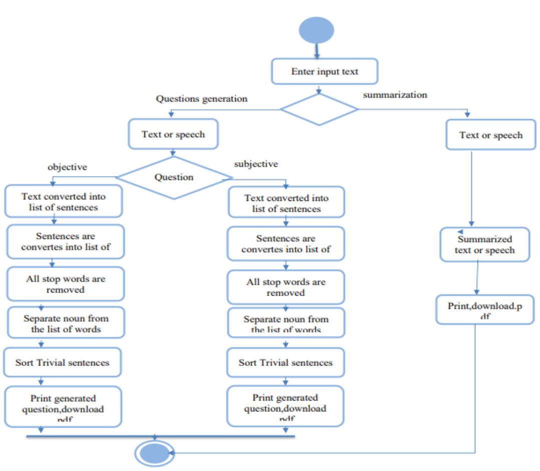

Fig 3:Activity diagram

3.2.4 SEQUENCE DIAGRAM:

The text summarization and question generation would depict the interactions between objects and classes, showcasing the order of method calls and messages exchanged during the process, from text input to output generation, illustrating the flow of control and data between components involved in summarization and question generation.

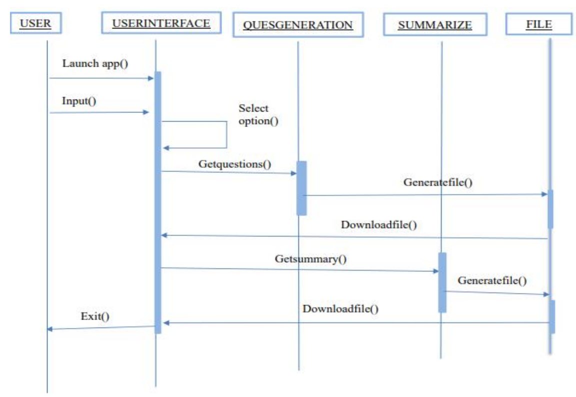

Fig 4: sequence diagram

4 . OUTPUT SCREENS

STARTING USER INTERFACE

The following output screen shows the starting page of the application which text copied in it and the choices of type of questions to generate.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
IJARST
www.ijarst.in

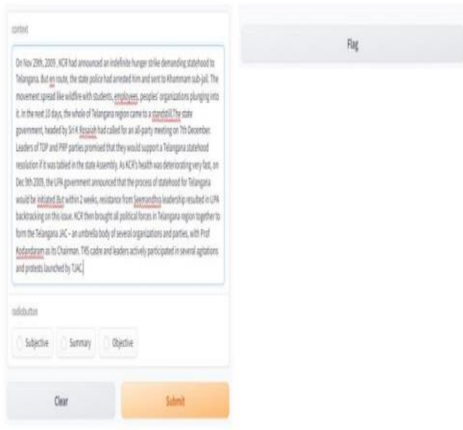

Fig 5: Starting User Interface

SUMMARY GENERATION

The following output screens shows that the summary is generated.

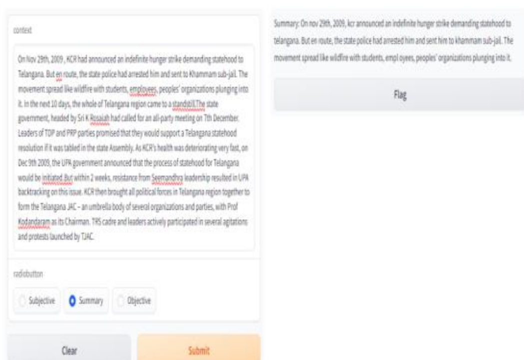

Fig 6: Summary Generation

SUBJECTIVE QUESTION GENERATION

The following output screen shows the subjective questions generated.

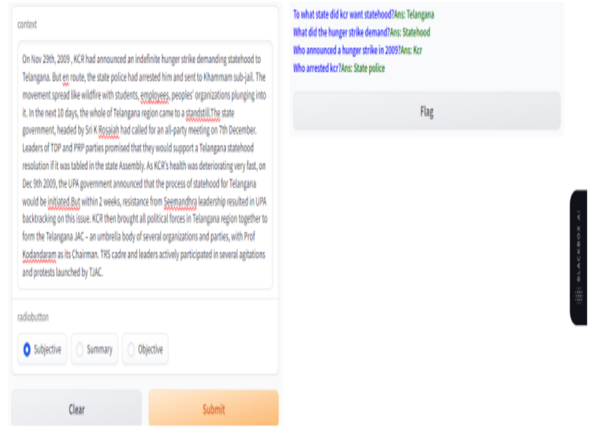

Fig 7: Subjective Question Generation

OBJECTIVE QUESTION GENERATION

The following output screen shows the objective questions being generated.

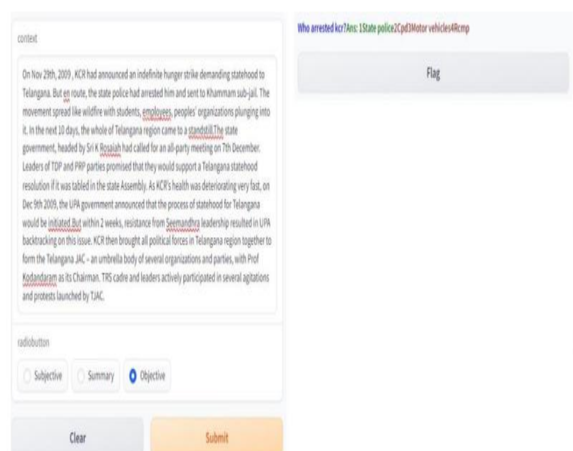

Fig 8: Objective Question Generation

5. CONCLUSION

The generation of questions from text is useful in numerous application domains, but manual generation is time-consuming and

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

costly. In addition to being a web-based anddesktop-based application system, it offers several features mainly for producing MCQs, Boolean type, and fill-inthe-blank questions. The goal of this application is to describe anautomatic question generator built on NLP. This system helps students to self- analyze their knowledge about the subject. The application can be used in the educational field, and it is helpful in generating test papers for assessment. The generated questions are displayed on the student side with questions and answers and like an assessment for practice test for students. It is helpful in competitive exams by generating questions and practicing them.

6 . FUTURE ENHANCEMENT

In conclusion, question generation using NLP is a promising area of research that has seen significant advances in recent years. Various approaches have been proposed for generating questions from text, including rule-based systems, neural network models, and unsupervised methods. These approaches have been applied in various domains, including educational systems, chatbots, and search engines, to enhance their functionality and improve user experience The studies reviewed in this literature survey demonstrate that NLP techniques can effectively capture key information from text and generate relevant and useful questions. However, there are still challenges to be addressed, such as the generation of questions that are diverse, non-redundant, and reflect a range of levels of complexity. Additionally, there is a need for further research to evaluate the performance of question generation systems across multiple domains and languages, and to explore their potential applications in new areas. Overall, the potential benefits of question generation using NLP are numerous and wide-ranging, and continued research in this area is likely to yield valuable insights and new applications.

7 . REFERENCES

[1] Zhou, Q., Yang, N., Wei, F., & Tan, C. L. (2017). Neural question generation from text: A preliminary study. arXiv preprint arXiv:1704.01792.

[2] Du, X., Shao, J., & Cardie, C. (2017). Learning to ask: Neural question generation for reading comprehension. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL).

[3] Alsaedi, M., Cetinic, E., Liu, H., & Yang, H. (2019). Automatic question

generation for literature review writing support. Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE).

[5] Gao, J., Li, W., Lin, Y., Zhang, M., Liu, Y., & Huang, Y. (2021). Question generation from knowledge graphs with neural machine translation. IEEE Transactions on Neural Networks and Learning Systems, 32(3), 1213-1223.

[6] Zhang, L., & Lee, W. S. (2003). Question classification using support vector machines. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, 26(1), 26-32.

[7] Zhang, X., Guo, D., Yu, Y., & Wang, X. (2021). A natural language processing approach for question generation. Journal of Ambient Intelligence and Humanized Computing, 12(5), 4555- 4566.

[8] See, A., Liu, P. J., & Manning, C. D. (2017). Get to the point: Summarization with pointergenerator networks. Proceedings of the 55th annual meeting of the association for computational linguistics, 1073-1083.

[9] Nallapati, R., Zhi, F., & Zhou, B. (2016). Summarunner: A recurrent neural network-based sequence model for extractive summarization of documents. Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, 747-756.

[10] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitley, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modelling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine, 29(6), 82-97.