AI Based Chatbot To Answer FAQs

Dr.Rashmi Amardeep[1], Arfa Tazmeen[2], Dinoj Kumar V[3], Kalyan Kumar R[4], Kandula Thanmayee[5]

[12345] **dr.rashmi-is@dsatm.edu.in,tazmeenarfa@gmail.com, vdinojkumar@gmail.com, kalyanreddy0503@gmail.com, thanmayeemittu@gmail.com,**

[1]*Faculty in Department of Information Science and Engineering, Dayananda Sagar Academy of Technology And Management, Bangalore, India*
[2,3,4,5] *Department of Information Science and engineering, Dayananda Sagar Academy of Technology And Management, Bangalore, India*

Abstract

In this digital age, the smart application is essential. Ought to continuously generate a huge amount of data into existence. These data can be used to develop a big volume of knowledge that deploys various uses. Education is one of the primary industries that create massive amounts of data. However, because of the speed and volume of data generated by numerous online educational resources, it is difficult to obtain only the necessary information. One type of technology that can be used to extract pertinent information from textual data is text summarization and analysis tools. Although a lot of text summarization tools are being created, the majority of them are primarily concerned with summarizing a single document well. The objective of this project is to create a text summarizing tool that uses NLP methods to extract pertinent and useful information from a variety of texts, enabling users to learn more efficiently. A chatbot interface can convey this information to users in an interactive and effective manner. In order to govern the conversational flow and personalize the user experience, the technology also runs different analytics on the user responses provided to the chatbot.

Keywords - Text Summarizer; chatbot; ROUGE; BERT; Online Education; Natural Language Processing (NLP)

I. INTRODUCTION

One of the most significant recent changes in the education profession is the tremendous increase in student use of online education. Students have begun to take a huge number of online courses to supplement their regular formal education. Students also commonly utilize the internet to find answers to their inquiries, clear their doubts, and in other situations. However, due to the large number of available sources and documents, there is information overload, making it increasingly difficult for students to obtain the right information to meet their learning requirements.

Online courses are currently being developed with the goal of providing a big volume of content while also enhancing the quality of the content. However, there is no emphasis on the user in this process, and each user is presented with the same content in the same manner. When people take such a course, they are frequently overwhelmed by the content, which is delivered in the form of massive text documents or videos. Users find it difficult to maintain concentrate throughout the duration of their consumption of this content. As a result, consumers frequently overlook part of the information supplied. The evaluation methods utilized in these courses are quizzes and tests administered at the end of each module. These methods of evaluation are ineffective and do not contribute to the learning process.

As a result, the present online courses lack personalization and require more user participation as well as enhanced user rating. Users are always engaged and do not lose concentration by producing an excellent summary of the numerous papers linked to the course and delivering this information to them in an interactive and user friendly interface via a chatbot. Users are also continually assessed as their responses are analyzed in real-time to track user learning and promptly rectify any misconceptions or missing information. As a result, this technology can help users increase the overall efficacy of online education.

II. RELATED WORK

A. Text Summarization:

A substantial amount of research is being conducted on various text summarizing approaches. Text summary can be categorized in several ways, as seen in Fig.1. The most prevalent approach of categorizing summarizing techniques is based on the output type, which is classified as -Extractive and Abstractive summarizing. [1]z

Fig. 1. Types of Text Summarization

1. Selecting the most significant sentences from a document and including them exactly as they are in the summary is known as extractive summarization. The sentences are directly copied from their parent document.

2. To understand a document's contents, abstractive summarization includes analysing and interpreting it as a whole. The generated summary is composed of freshly created words that are not from the parent document, but which are grammatically sound and convey the tone and meaning of the parent document's most crucial passages. [2][3].

3. Extractive summarization is a less complex technique than abstractive summarization, but because the summary's content is restricted to sentences from the parent papers, it is more linguistically restrictive. A summarising tool that uses a variety of articles as input must make sure that all of the sources are considered in the final summary. Therefore, the best method is to choose a few sentences from each paper using an extractive approach, then combine these sentences into a succinct, coherent summary using an abstractive approach.

The practise of text summary is growing in acceptance in both academic and applied contexts. As a result, summarising tools and approaches have undergone a number of important developments. Many different industries use text summarization software, including:

Media monitoring - summarization have been developed to help with the issues of information overload and content shock.

In shorts - news delivery app that chooses the most popular domestic and international news stories. For each story, the app compiles information on the subject from the internet and other news sources and condenses it into a succinct 60-word essay.

Financial research - To help them make judgements, investment banks sift through enormous amounts of market data. As a result, text summarizers for financial documents like earnings reports and financial news have been developed. As a result, analysts may be able to identify market patterns more quickly.

B. Chatbots:

Chatbots are practical tools because they enhance user interaction with websites and applications and can be quickly implemented on already-existing websites and apps. As a result, the use of chatbots for numerous use cases in a number of industries has skyrocketed, including lead generation in online sales, customer service for various organisations, in-app assistance to improve the user experience, and so forth. During user interactions, chatbots can be used to read user comments and respond correctly in order to obtain more useful information, such as feedback. [11]. Some of the advancements in chatbot technology include:
Endurance: A Dementia Patient's Companion - This is an open-source chatbot system that was created to connect with dementia patients. Because these patients frequently suffer from short-term memory loss, they find

it difficult to interact with others. The chatbot can be used to connect with such patients as well as track their health and detect memory lapses. Conversation logs are also kept for doctors to review.

Quiz chatbots - These are chatbots that can be used to create quizzes that can be used to instruct or entertain users. These are tests with fixed question lists and multiple-choice answers. These chatbots simply track user selections and do not perform much analysis, and user responses are very limited because they can only select answers from the given options.

III. PROBLEM STATEMENT

This project is aimed at developing an AI based Chatbot to answer FAQ’s.

IV. PROPOSED SYSTEM

The suggested system is a general-purpose tool for summarizing and analyzing numerous paragraphs of instructional content and presenting it to the user as an engaging and intelligent discussion via a chatbot interface. With minor changes, it may be adapted to diverse use cases and effectively applied to different types of people.

The system is organized into three major subsystems:
1. Chatbot user interface
2. The backend services - RESTful exposed via an API gateway.
3. Database layer

Through HTTP, a chatbot communicates to an API gateway. Backend services, each of which is covered by a separate database, are selfcontained. This separation of interests is in line with the microservices architecture envisaged for this tool, as shown in Fig 2.

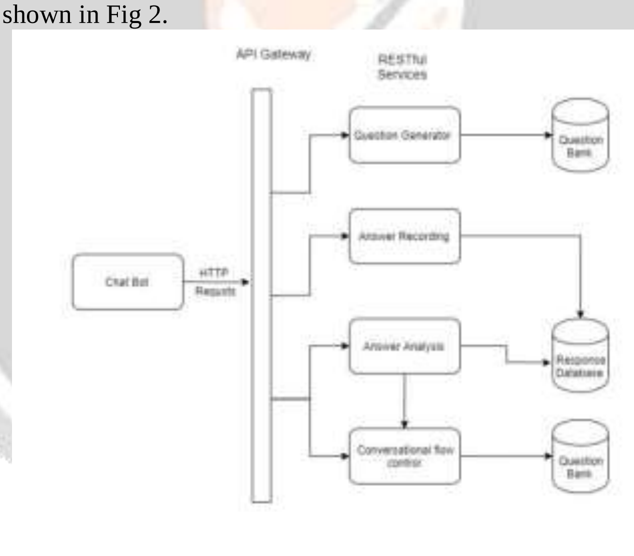

Fig. 2. Microservices Architecture

The main functionalities provided by this tool are:

1. Topic Modelling - Multiple text documents on a specific topic are used as input. All of these texts are presumptively divided into a given number of subjects. Some words are more closely associated with one topic than others. Based on the terms in the sentence, each sentence is categorized into one of the following subjects. This allows us to break down the content so that we can present it in an engaging manner, as well as automatically produce the questions that need be asked to ensure that the user is interested and has correctly digested the topic. Users can learn more successfully since they are always engaged.

2. Semantic Relationship Modelling - The content of a topic is made up of numerous sentences. These phrases have a number of underlying links. For example, one sentence may introduce a notion to which another sentence may provide an illustration. If multiple sentences are listing points about the same concept, they have a part whole relationship.

For each topic, a map depicting the relationship between sentences that influence each other is constructed and saved. This map, together with the numerous sentences, represents a path and proposes how the topic should be handled when the content is provided to users via the chatbot.

3 .Question Generation - The corresponding content is analyzed for each of the subjects specified by topic modelling. Questions are produced for users based on this content. The significance of each question, as well as the relationship between them, is determined. This is used to establish a conversational flow that allows users to cover all topics through inquiries. The content is also used to produce model answers for each of the questions. Identifying the key phrases that will constitute the answers and developing the corresponding questions [9] and automatic question generation by converting declarative statements and textual passages into questions relevant to their content [8] are two developments in the field of question generation.

The questions are generated through extractive summarization of the content to collect the important sentences. These statements are then transformed into questions. BERT: Bidirectional Encoder Representations from Transformers [12] can be used to do this summarization.

BERT is a Google open-source tool that specializes in NLP activities such as question answering and natural language inferences. In contrast to all prior techniques, this tool sees a text in two directions: left to right and right to left. The accuracy of this new technique has improved significantly. Instead of the typical sequential technique, BERT is bidirectional in that it analyses a whole phrase at once. As a result, it is also known as nondirectional. Attention processes are used by BERT. In order to find the elements of a single sentence connected with which parts of it, the attention mechanism is comparing two sentences and comparing portions of both sentences against each other. It is a useful technique for machine translation, however when the phrases are connected to each other, i.e. if correlation between parts of one sentence occurs by considering every single part of that sentence in combination with another word it will be called bidirectional model. The BERT model has the highest parameters and is trained on a large corpus, compared with its competitors. As more training steps are taken, the BERT model shows a significant improvement in accuracy. The convergence time of this model is longer than that of its competitors, but it beats them with some basic training techniques. Database for analyzing the user's progress over time.

4. Answer Analysis - The stored responses from each user are analyzed in order to track the users' learning. This is done to personalize their learning experience and to assist them in filling any gaps or ambiguities in their learning by offering appropriate knowledge. Furthermore, user interaction has been analyzed to determine user learning styles. We may categories various topics based on how difficult it is for people to grasp that issue.

This tool may then be customized to match the needs of the user in a variety of ways, including -

1. Quick learning tool for busy professionals: In today's fast-paced world, many, especially professionals, find it challenging to satisfy their hunger for information. Learning new skills and concepts takes a significant amount of time. People can use this tool to catch up on their learning at any time and from any location for as long as they want. This would be an excellent learning tool.

2. Exam preparation: This tool may be customized to assist students prepare for examinations fast and effectively. The test subjects can be covered by the questions, and the replies of the users are analyzed to assess their knowledge. This is used to identify the subjects on which the user should concentrate in order to increase performance.

3. Training requirements analysis - In order to enhance productivity, organisations are often required to evaluate the capabilities of their employees and areas in which they must be trained. To accomplish this, a questionnaire will be created and interviews conducted with employees in order to collect data which are then analysed for the purpose of evaluating employee training requirements at various levels within an organisation.By utilising the summarising tool and chatbot, this process may be completed with a great deal less time and effort. It is possible to analyse and summarise the replies of workers to surveys and interviews. It is possible to conduct employee interviews using the chatbot. You can utilise this to complete the final training requirements.

V. METHODS OF EVALUATION

The relevance of summarization assessment methods has grown in tandem with the adoption of text summary approaches. The techniques of evaluation are intended to determine the quality of summary. The many forms of quality measures are as follows:

1. Extrinsic Evaluation: This statistic is used to assess the value of the summary output for subsequent tasks including question-answering, relevance assessment, and sentiment analysis.

2. Intrinsic Evaluation: Internal quality evaluations are conducted throughout the summary process. The quality and informativeness of the summary may be assessed via intrinsic evaluation.

3. The most recent and extensively used method for judging how useful a summary is is recall oriented understudy of gisting evaluation (ROUGE)[5].The length of typical unigrams, bigrams, or n-grams between the output and a reference output or the original text is determined by ROUGE-N. Based on the longest common subsequence method, the weighted longest common subsequence (ROUGE-W) method gives sequences that are close together more weight. A comparison between a perfect summary and the output of the summarization yields the F-score. It is the process used to assess the paraphrases produced by the summary.

VI . FUTURE SCOPE

Any college or university can take the initiative to integrate in their website. The chatbot can provide educational data in addition to being expanded by including additional data sources. Then it will be both beneficial to students and other visitors. The chatbot therefore can offer a variety of information based on the developer's requirements and configuration. NLP, or natural language processing, can be added to the chatbot to improve it.

REFERENCES:

[1]. Tas, O., Kiyani, F.: A Survey Automatic Text Summarization. In: 2nd World Conference on Technology, Innovation and Entrepreneurship, Istanbul, Turkey. (May, 2017)

[2]. Yeasmin, S., Tumpa, P. B., Nitu, A. M., Uddin, M.P., Ali, E., Afjal, M.I.: Study of Abstractive Text Summarization Techniques. American Journal of Engineering Research Volume-6, Issue-8, 253-260. (2017) [3]. Nayeem, M. T., Fuad, T. A., Chali, Y.: Abstractive Unsupervised Multi-Document Summarization using Paraphrastic Sentence Fusion.

[4]. Banerjee, S., Mitra, P., Sugiyama, K.: Multi-Document Summarization Using ILP Based Multi-Sentence Compression. In: Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI) (2015).

5]. ROUGE - Tool to Evaluate Summarization :http://kavitaganesan.com/what-is-rouge-and-how-it-works-for-evaluationof-summaries/#.Xcz0IlczbIU

[6]. Topic Modelling Tong, Zhou & Zhang, Haiyi. (2016). A Text Mining Research Based on LDA Topic Modelling. Computer Science & Information Technology. 6. 201-210. 10.5121/csit.2016.60616.

[7].Guifen Zhao, Yanjun Liu, Wei Zhang, Yiou Wang : TFIDF based Feature Words Extraction and Topic Modeling for Short Text, ICMSS 2018: Proceedings of the 2018 2nd International Conference on Management Engineering, Software Engineering and Service Sciences

[8]. Husain, Mohd Shahid & Jethwani, Himanshu & Akbar, Mohd. (2015). Automatic Question Generation from Text. International Journal for Innovations in Engineering, Science and Management. 3. 27-31.

[9]. Angelica Willis, Glenn Davis, Sherry Ruan, Lakshmi Manoharan, James Landay, Emma Brunskill: Key Phrase Extraction for Generating Educational Question-Answer Pairs. Proceedings of the Sixth Annual ACM Conference on Learning at Scale, 2019

[10]. Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L. Murnane, Emma Brunskill, James A. Landay: QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge. ACM Conference on Human Factors in Computing Systems, 2019

[11]. Y. Alex Kolchinski*, Sherry Ruan*, Dan Schwartz, Emma Brunskill (* equal contribution): Adaptive Natural-Language Targeting for Student Feedback. Proceedings of the Fifth Annual ACM Conference on Learning at Scale, 2018 [12]. Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805v2 [cs.CL] 24 May 2019