International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 78-85
Published Online November 2022 in IJEAST (http://www.ijeast.com)

VOICE ENBALED Q & A SYSTEM

Prasad Kashish, Shaik Mohammed Arshad, Supriya N
Department of Electronics and Communication BNMIT
Bangalore, Karnataka State, India

Abstract— **The aim of question answering is to respond to queries that are expressed in natural language (QA). Question answering systems offer automated ways to locate solutions to queries posed in natural language. One important development in information can be seen in the concept of question-answering systems particularly in its capacity to retrieve information resources in a natural way, retrieval technologies method through effective word-for-word querying and retrieval of the appropriate responses.**

The development of machine learning algorithms like text summarizing, which can automatically shorten lengthier texts and extract summaries of individual sections of text without losing the meaning, is another result of advances in NLP. Text summary is a cutting-edge method of information processing in a time-constrained and efficiency-obsessed culture. It shortens the amount of time needed to read and makes finding, analyzing, and assimilation of relevant information simpler.

In order to answer the questions that users have asked and to provide a concise overview of the same context, a novel method for extracting text from a picture and converting it into text files has been developed. In order to validate the proposed methodologies, a manual evaluation of the quality of the responses was also carried out.

Keywords— **QA System, Text Summarization, BERT, NLP, Tessaract OCR, DeepSpeech, flask framework.**

I. INTRODUCTION

A group of technologies known as conversational AI enable human-like interactions between people and machines via speech and natural language, which are our most instinctive interfaces. Conversational AI-based systems can comprehend commands by detecting speech and text, translating between languages on the fly, comprehending our intentions, and reacting in a manner that resembles human interaction.

Within machine learning and artificial intelligence, the discipline of natural language processing (NLP) is quickly expanding. Simply said, machine learning is the act of teaching computers to read, interpret, and process human languages. It has a wide range of uses, including spell checking, auto-correction, chatbots, product suggestions, and many more. In this project, a voice-enabled question-answering assistant has been presented which can read text from an image or any text file inserted and train the text file thus generated using machine learning models to answer any questions related to that context through voice inputs.

An attempt to make this project user-friendly was made by creating a website where the users will have their own account and can feed images, or any text file like in order to ask questions from the fed input and also generate summary for the same if required by the user. It can also help to store the frequently asked questions in the database and suggest users with relevant questions. The methodology implemented is approached by giving a comprehensive study of different machine learning models like BERT, RoBERTa, DistilBERT as such and their underlying algorithms and then implementing the best out of them.

Motivation

One of the most important tools for daily work is search, which has recently undergone a rapid evolution. We change our behavior from word matching to in-depth understanding of questions, and we start inputting questions instead of obvious terms in search fields. When you view a specific web page from the search results, Google and YouTube now highlight the answers when they are present on the page. These services already provide immediate responses to questions. Blind people are unable to do visual tasks. For instance, using a braille reading system or a digital speech synthesizer when reading text (if the text is available in digital format). There are still fewer published printed works available in braille, audio, and digital formats, respectively. Therefore, there is a lot of promise and value in creating a smartphone application that can convert text that has been written on a wall, a piece of paper, or similar support into speech. Text can be recognized from image data thanks to optical character recognition (OCR) technology.

The Question Answering (QA) systems, which go beyond standard search by immediately responding to questions rather than looking for information that matches the query, often aid in finding information more quickly.

Machine learning techniques such as text summarizing, which can automatically shorten lengthier texts and extract summaries of parts of text without losing the message, have been developed as a result of advances in NLP. Text summary is a cutting-edge approach to process information in a culture confined by time and preoccupied with efficiency. It cuts down on reading time while also making finding, processing, and digesting relevant information easier.

The continuous expansion of text that is available on the Internet in a variety of formats necessitates in-depth research for text summarization software. The essential content is communicated in the condensed version created from one or more papers, which is much shorter than the original text. At the same time, there are a number of difficulties in summarizing the extensive text collection. One of the main problems with text summarization, aside from the time complexity, is the degree of semantic similarity. The user can quickly and easily understand the enormous corpus thanks to the condensed text. In this project, several categories of text summarization were reviewed.

In addition to web search, there are numerous fields where people deal with papers that are specific to a certain domain and require effective solutions to manage business, medical, and legal records. Recently, the COVID-19 event considerably increased curiosity about QA systems. Due to the daily publication of hundreds of research articles, reports, and other medical materials, efficient information retrieval methods are required. In this case, QA systems are an excellent addition to conventional search engines. For information retrieval systems that will be widely integrated in the near future, QA is thought to be just the next sensible step.

Objective

The goal of this project is to develop a website that can swiftly sift through data from any text file, including data contained in books, documents, images, and other text files, and then present users with the information they need in an approachable manner.

This website can find its application in various fields starting from aiding the blind, helping a student by summarizing the context when required in their preferable language and also in the medical field. It would make it easier for drivers to find the information they require in a lengthy car manual, saving them time and effort in tense or challenging circumstances.

The application might also make it easier for experts like physicians, lawyers, and other professionals to scan through lengthy documents rapidly in search of particular medical results or specialized legal precedent. The use of technology would facilitate their work and free up more time for them to apply their knowledge to patient care or the creation of legal opinions.

II. LITERATIRE SURVEY

This section is divided into different sections covering survey on QA systems and text summarization. A brief introduction to each one of these problem statements is presented, conducting a literature survey on QA systems and text summarization using BERT model and reviewed the methodologies that has been used to implement each of these sections.

Bert Model

• In paper [1] titled “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, by creating auxiliary sentences to change the task into a sentence-pair one, a BERT-based text classification model is suggested in an effort to include additional task-specific knowledge and address the task-awareness difficulty. Also mentioned is a post-training strategy that makes use of domain-related corpora to address domain challenges. The disadvantage of the suggested strategy is the additional computer resources needed to train, fine-tune, and derive conclusions.

• In paper [2] titled “RoBERTa: A Robustly Optimized BERT Pretraining Approach”, Robustly Optimized BERT Pre-training Approach is known as RoBERTa. It was given by researchers from Washington University and Facebook. The goal of this study was to accelerate pre-training for the BERT architecture through training process optimization. Although RoBERTa's architecture is almost exactly the same as that of BERT, the authors made a few small changes to the training procedure and architecture to enhance the results for BERT. These changes are:
o Elimination of the Next Sentence Prediction (NSP) goal.
o Larger sample sizes and longer sequences during training.
o Adjusting the masking pattern dynamically.

• In paper [3] titled “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter”, the DistilBERT general purpose language representation model, which may be adjusted to perform well on a range of tasks like its larger counterparts, is advised to be pre-trained. The study also shows that a BERT model can be made 40% smaller while still retaining 97% of its language understanding abilities and running 60% faster.

QA Systems

Machine learning or deep learning models that can answer questions with or without context are referred to as question-answering models. They can choose one option from a list of available options, generatively paraphrase the answer, extract answer phrases from paragraphs, and more. The training dataset (such as SQuAD, CoQA, etc.), the goal for which it was trained, and to some extent the neural network design all have a role.

• In paper [4] titled “A literature review on question answering techniques, paradigms and systems”, in addition to QA approaches, systems, tools, metrics, and indicators, this study attempts to investigate the connection between question answering and natural language processing. It also seeks to establish the relationship between these two fields. A thorough literature assessment of studies released between 2000 and 2017 was used as the technique. There were 130 out of 1842 papers that described a QA approach that was developed and evaluated using different techniques. The conclusion of this work is that knowledge bases, information retrieval paradigms, and natural language processing have been the primary research areas for Question Answering researchers.

The majority of studies concentrated on free domain. Precision and Recall are the most discussed measures used to assess the techniques.

• In paper [5] titled "Research and reviews in question answering system", the diverse quality assurance systems that have been created so far are divided into three categories based on the general method that each system takes while processing data: linguistic, statistical, and pattern matching. The advantages and disadvantages of each strategy are explained, and a comparison of these approaches is provided. Finally, it is seen that selecting a technique is very problem-specific. However, it is understood that methodologies for answering questions based on linguistic, statistical, and pattern-based approaches will continue to be of great interest to QAS researchers and stay in sharp focus.

• In paper [6] titled "Question Answering Systems: A Systematic Literature Review", Deep learning, artificial intelligence, machine learning, word encoding and knowledge systems, syntax and context, and deep learning are the main techniques of the systematic literature review of QAS that are covered here. The use of ineffective and inefficient models, the development of highly specialized QAS systems, the decreased practicality of QAS due to the requirement for standard datasets or question formats, and the inability to appropriately evaluate QAS are just a few of the significant flaws that have been found.

The systematic review of the literature for the suggested method was restricted to research that were published in English, which is one of its shortcomings. It's probable that some useful researches were dropped as a result of this condition, even if it was vital to make sure the author could understand the studies that were chosen. The poll was limited to QAS studies released after January 1, 2018. Despite being relatively older, this criterion may also have restricted the inclusion of potentially useful studies.

Text Summarizer

Text summarization is the act of computationally condensing a set of data to produce a subset (a summary) that captures the key ideas or information from the original text. Below mentioned are the few methodologies proposed for text summarization collected from the following research papers.

• In paper [7] titled "An Entity-Driven Framework for abstractive Summarization", the paper presents an entity driven summarization framework. SENECA, utilizes a two-step process:

o a module for entity-aware content selection

o an abstract generation module

The drawback of the proposed is that it is complex in implementation.

• In paper [8] titled "Summarization of emergency news articles driven by relevance feedback", An itemset-based method for summarizing collections of news articles is presented in this work. The suggested method, called Feedback-driven News Summarizer (FeedNewsSum), summarises news articles based on item sets and is motivated by relevance feedback. In order to remove irrelevant lines that were inadvertently included in the summary or to recognize important sentences that were overlooked during the first summation process, feedback scores on summary sentences are used. The proposed method's disadvantage is that it requires feedback from a human (a domain expert).

• In paper [9] titled "News Text Summarization based on Multi Feature and Fuzzy Logic", For news material, a new automatic text summarizing model based on fuzzy logic principles, multi-feature analysis, and evolutionary algorithms is proposed. The word feature, which considers extracted words with scores over a threshold, is the most crucial feature. Direct keyword extraction is possible for time, place, and person. The suggested model's disadvantage is that weights must be manually assigned to each attribute in various contexts.

III. PROPOSED METHODOLOGY

Machine Learning Models

• Tesseract OCR

Tesseract is an Optical Character Recognition (OCR) engine that supports Unicode and has built-in support for more than 100 different languages. Other languages can be taught to be recognized by it. Tesseract is used to detect text, video, and image spam in Gmail on mobile devices. Text can be extracted from images directly by programmers or through the use of an API. Many different languages are supported. Tesseract doesn't come with a built-in graphical user interface, however, there are plenty on the third-party website. Through the use of wrappers, which may be obtained here, Tesseract works with a variety of frameworks and programming languages. It can be used along with the layout analysis now being used to locate text inside a huge document, or it can be used along with an outside text detector to locate text in an image of a single text line.

• Deep Speech/ Speech Recognition

An open-source embedded speech-to-text engine called DeepSpeech can operate in real time on a variety of devices, including a Raspberry Pi 4 and powerful GPU computers. The most prevalent form of human communication, speech is crucial for understanding behavior and cognition. Speech recognition is a technique used in artificial intelligence

that enables computer systems to understand spoken words. For machines to ingest this information, it must be stored as digital signals that software can then decode. For the purpose of making audio machine-readable, they are modifying the frequency. The user must clean up the data during this stage of the data preprocessing process in order for the computer to process it. The DeepSpeech model must be reorganized if there are a lot of audio and text files required for training. The filenames and transcripts must all be arranged according to the specifications. Installing and setting the training environment can start as soon as the data is prepared. Once the system is operational, several prerequisites are installed for deep learning the DeepSpeech model. This open-source speech-to-text engine called DeepSpeech makes use of a standard machine learning model that was developed for Baidu's Deep Speech research paper.

• BERT versions
Bidirectional Encoder Representations from Transformers, a machine learning technique built on transformers, was developed by Google for pre-training natural language processing. BERT is a broad machine learning framework for addressing natural language (NLP). BERT takes advantage of the context provided by the surrounding text to help computers understand confusing words in text. With question-and-answer datasets, the BERT framework can be improved. It was pre-trained on Wikipedia text. BERT is a deep learning model based on the Transformers (Bidirectional Encoder Representations from Transformers). In a transformer, each input and output element are connected, and weightings between them are dynamically created based on this coupling (In NLP, this process is referred to as attention). Language models could only process text input sequentially, either from the right to the left or from the left to the right. BERT is unique in that it can simultaneously read in both directions. Transformer technology enabled this capacity, also known as bidirectionality. Using these bidirectional capabilities, BERT is pre-trained on the two separate but related NLP tasks of Masked Language Modeling and Next Sentence Prediction.

In Masked Language Model (MLM) training, a word is hidden within a sentence, and the computer is given instructions to guess the hidden word based on the context of the hidden word. The goal of the training for next sentence prediction is to educate the software to identify whether two presented sentences connect logically and sequentially or whether their relationship is purely random.

Every NLP technique aims to comprehend spoken human language within its context. For BERT, this typically necessitates selecting a phrase from a list. Models must typically be trained using a sizable collection of specialized, labelled training data in order to accomplish this. As a result, groups of linguists will need to laboriously label data by hand. The only plain text corpora employed for BERT's pre-training, however, were the Brown Corpus and the full English Wikipedia. While being employed in actual applications, it continues to learn unsupervised from the unlabeled text, improving (i.e., Google search). The "knowledge" on which development is based is provided through pre-training. From there, BERT can be tailored based on the tastes of the user and the steadily expanding body of searchable content. Transfer learning is the name for this procedure.
As was already mentioned, Google's examination of Transformers led to the development of BERT. The model's transformer is what provides BERT its enhanced capacity for understanding verbal ambiguity and context. Instead of processing each word independently, the transformer achieves this by processing each word in relation to every other word in the sentence. With the help of the Transformer, the BERT model is able to interpret a word's entire context and, as a result, better understand the searcher's purpose by examining all the phrases around it. Earlier approaches like GloVe and word2vec would map every word to a vector, which only captures a small fraction of its meaning in one dimension. This is in contrast to the traditional approach to language processing, known as word embedding. The table that follows provides an examination of the various BERT-based text classification models, and it includes a brief summary of the models' size, performance, data, and methodologies. These word embedding models require a significant amount of tagged data. However, because to the fact that every word is connected to a vector or meaning in some way, they have difficulty dealing with the context-dependent, predictive aspect of question responding. To prevent the word in focus from "seeing itself" or having a fixed meaning regardless of its context, BERT employs a technique called masked language modelling. BERT therefore states that context alone must be used to decipher the concealed term. In BERT, words don't have a fixed identity; instead, they are defined by their environment. Therefore, when all pertinent factors are taken into account, BERT proves to be more useful.

Flowchart
Figure 1 represents the flowchart for the methodology implemented. The website will have two types of logins, one for the admin access and one for the user access. Each user will be provided with a username and password through which they will access the website and maintain their individual activities.

• Firstly, the home page displays relevant information about the Q&A System and Text Summarization. The sidebar contains all the necessary actions that the user wants to
indulge.
• For Question- Answering, the user can either input the texts in the form of type format or can upload an image which contains information about the context from which the questions need to be asked. The uploaded image content is converted to text using Tesseract OCR and fed to BERT model for further training. The question is then asked which the system tends to answer for.
• Followed by this is the Text Summarizer which is used to provide a brief information of the content provided. It can also provide summarization for the content present in an image by

converting image text to text file using Tesseract OCR as done in Question Answering.

• Output answers as well as the summarized content can both be available as audio on click of the microphone, also termed as voice- enabled output.

• It also supports the functionality of storing the context along with the question and answer for Question Answering and the paragraph with its respected summary for Text Summarization in their respective history sections. This helps user to access their materials as per their requirement without asking the same question over the same content repeatedly. Also, one user's history won't affect other users' history.

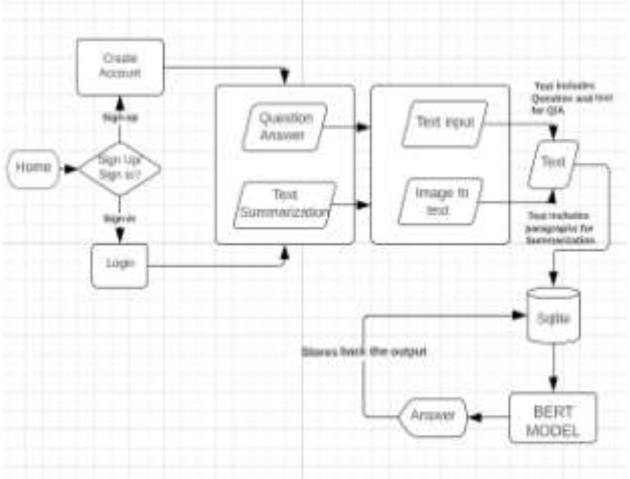


Fig. 1. Flowchart for the implemented methodology

IV. RESULT AND DISCUSSION

The proposed approach is ordinarily tested utilizing a variety of input data in text and graphic formats. It may be inferred from the simulation of the experiment results that this strategy is resilient to many various sorts of queries asked while providing the answers or the summary that the user requests.

Representation of the web interface of the Sign-up page is shown in Figure 2 where users get to register on this portal in order to create an account as a new user. The user must sign up and register after providing the necessary information, including their username, email address, and password, in order to log in using those same details the next time around as a returning user.

A web framework called Flask offers libraries for creating simple web applications in Python which was used for the backend.

Fig 2. Sign up page of the web interface

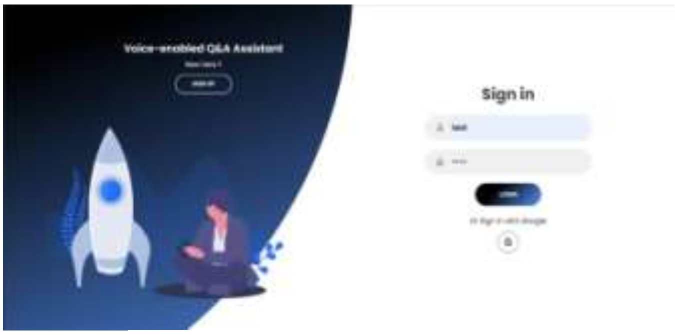


Fig 3. Sign in page of the web interface

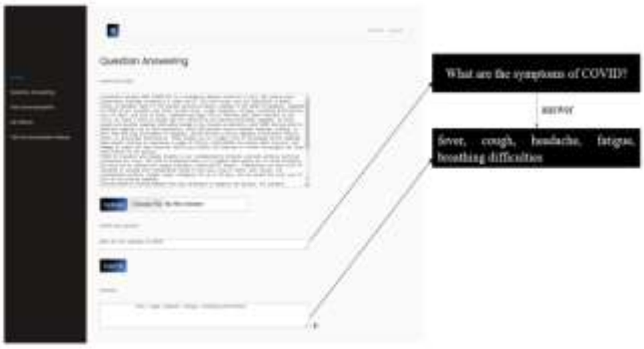
Representation of the Sign-in page of the web interface is shown in Figure 3. Those users who have already registered and whose names are on the list of registered can sign in. Their respective login information would already be preserved with the account, so they would only need to verify their identity as a returning user. One must register before they can successfully log in to anything they can access or view for the first time.

Question Answering

Text Summarization

Fig 4. Profile page for each user

Representation of the outline of the profile page of each authenticated user after their successful log in is shown in Figure 4. The profile page makes the templates of question answering, text summarization options and their history of the inputs provided in the past and obtained outputs according to each profile accessible to each of the users.



What are the symptoms of COVID?

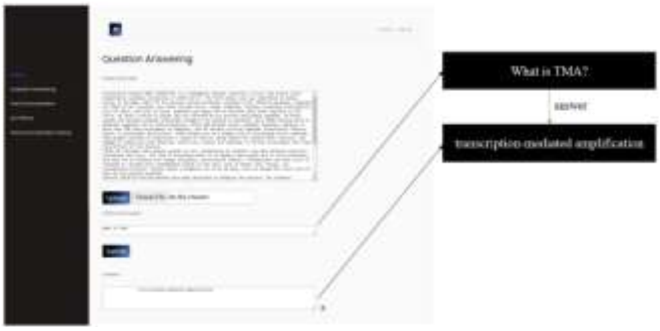
fever, cough, headache, fatigue, breathing difficulties

Fig 5. Question Answering section for the type format context with the answer 1

An illustration of the outcomes obtained following the successful compilation of the algorithms from the adopted methodology of the software part is shown in Figure 5. A text file related to COVID-19 has been inserted and a query on the same has also been provided in the query section of the web page. After successful processing of the implemented algorithms, the BERT model analyses the inserted text file, and the application gives the user the most accurate answer.

Figure 6 shows another such example of the obtained results where a text file related to COVID-19 has been inserted and a query on the same has also been provided in the query section of the web page. The fed context is analyzed and then the best solution is shown as an output.

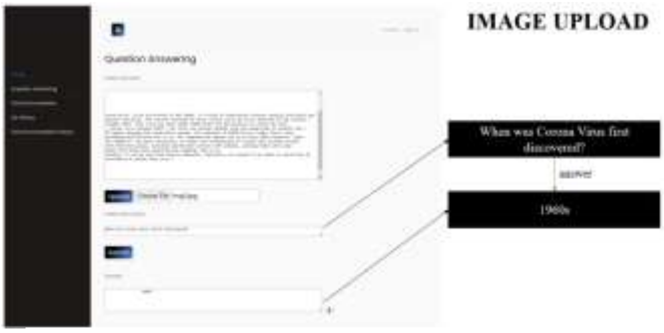
Figure 7 shows the result section of another option that has been provided with the webpage in which users can choose to upload images from their file section. Once the upload is complete, the text files present in the image will be extracted for the BERT model to analyses the same. The user can thus insert queries relating to the context of the image uploaded into the respective windows and obtain the necessary answers. The output has also been enabled with an audio option which is going to activate upon click to recite the answers which are thus obtained.



What is TMA?

transcription-mediated amplification

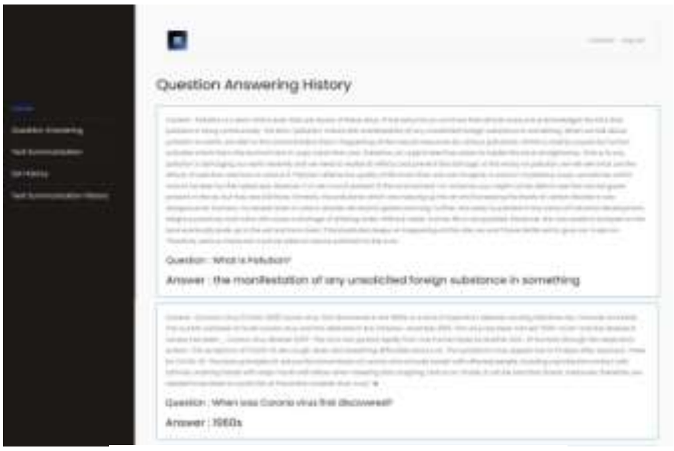
Fig 6. Question Answering section for the type format context with the answer 2



When was Corona Virus first discovered?

1960s

Fig 7. Question Answering section for the uploaded image context with the answer


Question Answering History

Question : What is Pollution?

Answer : the manifestation of any unsolicited foreign substance in something

Question : When was Corona virus first discovered?

Answer : 1960s

Fig 8. Question Answering History section

Figure 8 and 11 represent the history sections of the question answering and text summarization options of each individual user's profile. This section of the webpage contains the past records; the history of the information contained in text files fed into the portal; necessary answers; and summaries obtained after the successful compilation of the implemented algorithms.

Figure 9 and 10 depict the project's web interface for text summarization, with examples demonstrating the necessary summaries obtained as the result for the inserted text files in text format and for the texts extracted from uploaded images, respectively. Extracting texts from an image involves the use of Tesseract OCR as the python module which has the ability to fetch texts from any document such as pdf, or an image and then the BERT model can be implemented on the fetched texts to answer the questions so asked with the most appropriate answers.

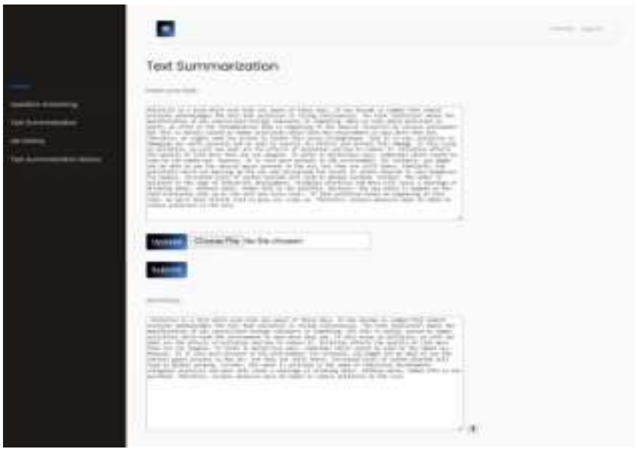


Fig 9. Text Summarization section for the type format context with the summary

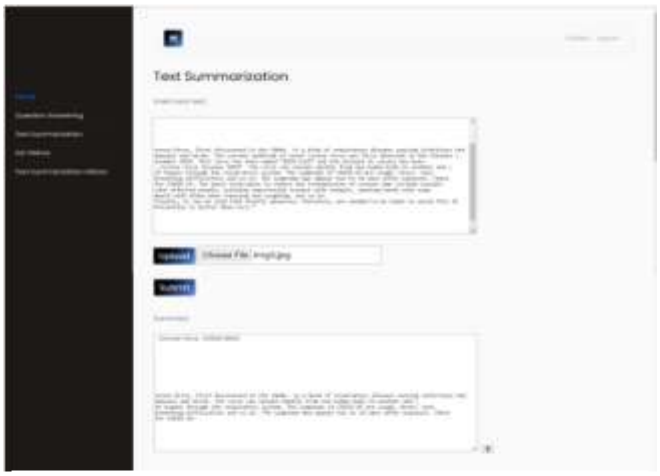


Fig 10. Text Summarization section for the uploaded image context with summary

Fig 11. Text Summarization History section

V. CONCLUSION

Answering questions in natural language is the goal of Question Answering (QA). Automated methods of receiving responses to natural language inquiries are provided by question answering systems (QAS).

The idea of question-answering systems represents a significant advancement in information retrieval technology, especially in terms of its ability to access knowledge resources naturally by posing questions and finding pertinent material in short sentences.

The increasing growth of content available on the Internet in a variety of formats necessitates in-depth research for automatically summarizing information. The key content is conveyed in a summary version created from one or more papers, which is much shorter than the original text. Simultaneously, summarizing the enormous text collection presents a number of issues. Aside from the temporal complexity, one of the primary challenges in text summarizing is the semantic similarity degree. The summarized text aids the user in quickly and easily comprehending the enormous corpus. In this paper, a novel technique is presented for extracting text from an image and converting it into text files that can be used by machine learning models to answer the users' questions and provide a concise overview of the same context.

VI. REFERENCE

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, 2016, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, Google AI Language, IEEE Access, Vol 4.

[2] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov, 26 July 2019, “RoBERTa: A Robustly Optimized BERT Pretraining Approach”, University of Washington, Seatle, USA.

[3] Victor Sanh, Lysandre Debut, Julien Chaumond, Thomas Wolf, 1 March 2020, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter”, 5th Edition co-located with NeurlIPS.

[4] Marco Antonio Calijorne Soares, Fernando Silva Parreiras, (2020), “A literature review on question answering techniques, paradigms and systems”, Computer and Information Sciences 32, 635–646.

[5] Sanjay K Dwivedia, Vaishali Singhb (2013), “Research and reviews in question answering system”, Department of Computer Science, B. B. A. University (A Central University), Procedia Technology 10, 417 – 424.

[6] Sarah Saad Alanazi1 Nazar Elfadil, Mutsam Jarajreh, Saad Algarni, (2021), “Question Answering Systems: A Systematic Literature Review”, International Journal of

Advanced Computer Science and Applications, Vol. 12, No. 3.

[7] Eva Sharma1* Luyang Huang2* Zhe Hu1*, "An Entity-Driven Framework for Abstractive Summarization", Lu Wang1 1Khoury College of Computer Sciences, Northeastern University.

[8] Luca Cagliero, "Summarization of emergency news articles driven by relevance feedback", Dipartimento di Automatica e Informatica Politecnico di Torino Torino.

[9] YAN DU AND HUA HUO, 11th August 2020, "News Text Summarization Based on Multi- Feature and Fuzzy Logic", School of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China, Vol 8.

[10] SHANSHAN YU, JINDIAN SU, DA LUO, 2016, "Improving BERT-based Text Classification with Auxiliary Sentence and Domain Knowledge", College of Computer Science and Engineering, South China University of Technology, Guangzhou 510640, China, VOLUME 4.

[11] Nikhil Mishra, 2017, "Image Text to Speech Conversion using Raspberry Pi & OCR Techniques", Department of Electronics & Telecommunication Engineering AGPCET Nagpur, Vol. 5, Issue 08.

[12] Jafar A Alzubi, Rachana Jain, Anubhav Singh, Pritee Parwekar, Meenu Gupta, 2021, "COBERT: COVID-19 Question Answering System Using BERT", DOI: 10.1007/s13369-021-05810-5.