VOICE BASED INTELLIGENT VIRTUAL ASSISTANT FOR WINDOWS USING PYTHON

*Rose Thomas, *Surya V S, *Tincy A Mathew, **Tinu Thomas
*Student, Dept. Of Computer Science & Engineering, Mangalam College of Engineering, India
**Assistant Professor, Dept. Of Computer Science & Engineering, Mangalam College of Engineering, India
rose.tj44@gmail.com suryavssurya001@gmail.com tincymathew2001@gmail.com tinu.thomas@mangalam.in

.

Abstract—**In this paper, Voice intelligent Assistance tool is used for searching purposes, summary extraction, setting reminders just by using voice commands. Voice recognition technology allows us to access any document or file we desire. If the user spells out the word it automatically types in the required field. It recognizes the speech and searches the appropriate content in the database and retrieves it. The audio from the microphone will be collected by this voice assistant, which will subsequently translate it into text. To ensure that the virtual assistant can comprehend them, the user should choose the proper language. If any wrong or invalid communication happens it invokes some messages in dialog box. It is a software application which performs tasks and events based on commands. Voice-Command and speech synthesis are enhancing the level of user-interaction in applications. This intelligent personal assistant (IPA) can interact with the user by opening a report, providing a brief summary via speech-to-text, and outlining the most crucial details in the appropriate context. Here, an attempt is made to build an intelligent voice personal assistant using Python, which offers the ability to control voice-activated devices and speech-activated smart devices for information extraction. NLP (Natural Language Processing) helps the virtual assistant to understand and respond to human speech and based on the voice commands the tasks are performed. BERT is designed for the computers to understand the meaning of ambiguous language.**

Keywords—Virtual Assistant, Speech Recognition, Extractive summarization, BERT

I. INTRODUCTION

The virtual assistants have become an essential component of our life nowadays as a result of all the functions and simplicity they provide. They can also automate some routine duties so that a user can concentrate on what is most important to them. A voice assistant is a digital assistant that uses speech synthesis, voice recognition and natural language processing (NLP). The software that can identify human voices and respond using an integrated speech system is known as a desktop-based voice assistant. A voice-based personal assistant is a helpful tool for searching, setting reminders, and taking notes just by speaking. Therefore, making use of these virtual assistant capabilities will enable one to save a lot of time and work. Everyone wants an assistant these days that will listen to their calls, anticipate their requirements, and take the appropriate action when necessary. The command open can be used by the user to launch any other application. The voice assistant records vocal input through a microphone and converts it into understandable computer language to give the user the information and answers they require. Keeping track of test dates, birthdays, or anniversaries is another challenging endeavor for most of the people. To solve this problem, the voice based virtual assistant helps to generate reminders. Therefore, making use of these virtual assistant capabilities will enable one to save a lot of time and work.

II. LITERATURE SURVEY

[1] A multi-functional Smart Home Automation System (SHAS) that can adapt to a user voice and recognize spoken instructions regardless of the speaker individual features, such as accent, was proposed by Yash Mittal. This system is affordable since processing and control are handled by an Arduino microcontroller board. Thus, the Smart Home Automation System (SHAS) prototype can be utilized to transform current homes into smart homes. [2] Home Automation Using Voice Commands in the Hindi Language. The voice recognition module and dedicated hardware, the Arduino Uno, were used in the planned home automation in Hindi language project to increase the system robustness and cost-effectiveness. The system can operate with several linked devices, such a lamp, fan, air conditioner, etc. With the use of voice assistants, this technology enables users to make decisions and control their home equipment. [3] In recent years, due to the progress of information technologies, the homes are built to smart homes. Smart home style can bring benefits to user, the technology becomes unavoidable in these years. Even enterprises still cannot integrate the functional divisions of smart home. Consumers struggle to find the products they require. In this paper, it builds a tailor-made function for users without their attempt; it makes use of Google Voice recognition in the house using machine learning to demonstrate the viability of a smart home pattern in order to meet user needs. This enabled user to interact with Google Home's voice recognition system while controlling devices by sending a Bluetooth signal to the Raspberry Pi. [4] Recent

years, voice assistants (VA) such as Alexa, and Google Assistant are becoming popular. Devices are dependent on the wake-up keyword for every utterance, which hinders the natural way of interacting with devices. This paper proposes an on-device solution which listens to the user continuously for a predefined period. Most frequently used domains in mobile phones, such as navigation, call, messaging, etc. are selected to create the dataset, through this it can distinguish background voices. The framework of this paper is done based on two stages i.e., Stage 1 and Stage 2, where both deals with the root command and follow up command. [5] Due to the growing popularity of virtual assistants, speaker verification (SV) has recently considered in research interest. At the same time, requirement for an SV system is increasing: it should be robust to short speech, especially in noisy environments. In this paper, it considers one more important requirement that is the system should be robust to an audio stream containing long non-speech, where a voice activity detection (VAD) is not applied. These requirements meet by using Feature Pyramid Module (FPM)-based multi-scale aggregation (MSA) and self-adaptive soft VAD (SAS-VAD). By the development of deep learning, a deep neural network (DNN)-based acoustic model has been integrated into the i-vector/PLDA system and used to generate senone posteriors for i-vector computation instead of the conventional Gaussian Mixture Model-Universal Background Model (GMM-UBM). Deep speaker embedding learning is another approach, which is most extensively studied approach. [6] With the arrival of the 5G and Artificial Intelligence of Things (AIoT) eras, related technologies such as the Internet of Things, big data analysis, cloud applications, and artificial intelligence have opened up new possibilities in a variety of application fields, including smart homes, autonomous vehicles, smart cities, healthcare, and smart campuses. The majority of university campus apps are offered as static web pages or app menus. The primary goal of this research was to create an emotionally aware campus virtual assistant based on Deep Neural Networks (DNN). It included Chinese Word Embedding to the robot dialogue system, which significantly improved dialogue tolerance and semantic interpretation.[7] This paper demonstrates the creation of software that assists the visually impaired in accessing the internet. The software adds a new dimension to accessing and providing commands to any website by using voice commands instead of the standard keyboard and mouse. The developed software can automate any website by reading out the content of the website and then using speech to text and text to speech modules, as well as selenium. The system may also provide a summary of the website's content and answer user queries with reference to the summary using a BERT model trained on the Stanford Question Answer Dataset.

III. PROPOSED SYSTEM

This system is based on a desktop application. The proposed system will provide a good understanding of the intelligent assistant that can recognize user commands. The voice assistant can readily recognize the user's verbal orders and makes the appropriate responses. The user's command is heard by the voice assistant through the microphone. For implementing a voice assistant for windows users, Speech Recognition that has many in-built functions that captures the voice by the user and convert the voice to text and vice versa is used. Users will have an option for opening the text files. The user's voice will be considered as input. The voice will then be processed and converted to text by the system. Following text processing, the system will look up text (file name) in the database. The system will open the file if the file name is found in the database. If it is a text file, the system will first summarize and read it to the user. We will use extractive text summarization using BERT model. BERT is a powerful language model that perform NLP task efficiently. The opened text files will first undergo preprocessing and produce a summary of the text using BERT. The user will next hear the written summary read by the system. Every function has a set of pre-defined instructions. An example of a predetermined instruction is "open Google". The voice assistant searches for related commands as soon as it gets a command, and each command further performs a task. The voice assistant does the action specified in the command when a match is made with the input. It will also generate reminders for specified date and time. Also, it also performs restarting, sleep or turning off our PC with a single voice command.

IV. TECHNOLOGIES USED

Python was chosen as the programming language for this project because of its adaptability and accessibility to a large number of libraries. Python programming language supporting Microsoft Visual Studio Code (IDE) is used to create the Virtual Assistant. Python has a speech recognition package that includes certain built-in functions. We will first define a function that will turn the text into speech. We employ the pyttsx3 library for that. The say() method is used to provide text as an argument and the result will be a voice reply. Another function is used to recognize the voice command. In order to transform the relevant analog voice command into a digital text format for our project, we select Google's Speech Recognition Engine. The Assistant will look for the keyword after receiving that text as input. The relevant function will be invoked and perform the actions like opening Google, Wikipedia, generating reminders, accessing files etc., if the input command contains a word that matches the relevant term.

A. BERT EXTRACTIVE SUMMARIZATION

BERT (Bidirectional Encoder Representations from Transformers) is a powerful natural language processing technique that can be used for various tasks, including extractive summarization. Extractive summarization is a technique used to generate a summary of a longer text by selecting the most important sentences or phrases from the original text. BERT can be used for this task by fine-tuning a pre-trained BERT model on a specific dataset for extractive summarization.

Special Issue - 2023

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICCIDT - 2023 Conference Proceedings

Here is a general overview of how BERT-based extractive summarization works:

- Preprocessing: The input text is split into sentences or phrases, and each sentence or phrase is tokenized into sub words using the BERT tokenizer.
- Fine-tuning: A pre-trained BERT model is fine-tuned on a dataset of text and corresponding summaries, where the model learns to predict which sentences or phrases I the text are most important for generating an accurate summary.
- Selection: The fine-tuned model is then used to score each sentence or phrase in the input text based on its importance for generating the summary. The top-scoring sentences or phrases are then selected and concatenated to form the final summary.
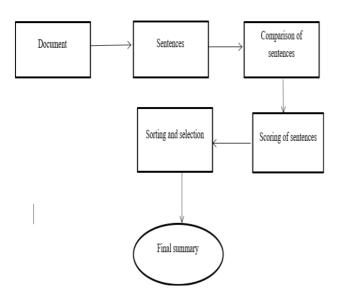

Fig.1: Bert Extractive summarization
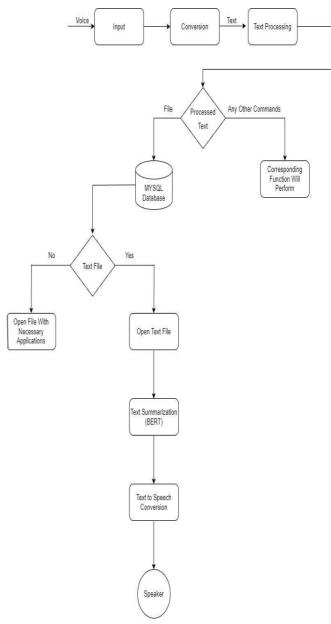
V. SYSTEM ARCHITECTURE

Fig. 2: System Architecture

VI. RESULT

The Python programming language's essential packages have been installed, and the code was implemented using Microsoft Visual Studio code (IDE). The following are just some of the outputs produced by our voice assistant.

A. Opening Google

As illustrated in the below fig 3, if the user give command to voice assistant to open google then google will be opened.

Special Issue - 2023

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICCIDT - 2023 Conference Proceedings

Fig. 3: Opening Google

B. Extractive summarization using BERT

As illustrated in the figure 4, If the user wants to generate a summary of the text file, that file will be checked in the database, if it presents then summary will be generated.
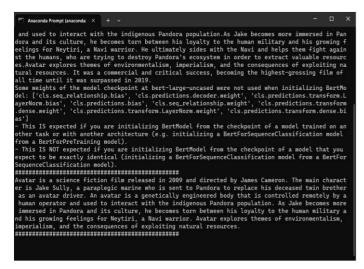

Fig.4: Extractive summarization using BERT

VII. CONCLUSION

The concept and implementation of a Python-based voice-enabled personal computer assistant is thoroughly described in this paper. This voice-activated personal assistant will be more effective at saving time in today's lifestyle than it was in earlier eras. The key characteristic of this Personal Assistant is its simplicity of usage. The Assistant effectively completes some duties that users assign it. Additionally, this assistant can perform a wide range of tasks, including restarting or turning off our PC with a single voice command, generating reminders and summaries of text documents etc.

REFERENCES

[1] Yash Mittal, Pradhi Toshniwal "A voice-controlled multi-functional Smart Home Automation System".

[2] Prerna Wadikar, Nidhi Sargar, Rahool Rangnekar, Prof.Pankaj Kunekar "Home Automation using Voice Commands in the Hindi Language".

[3] Chen-Yen Peng and Rung-Chin Chen "Voice Recognition by Google Home and Raspberry Pi for smart Socket Control".

[4] Abhishek Singh, Rituraj Kabra, "On-Device System for Device Directed Speech Detection for Improving Human Computer Interaction" 22 September 2021.

[5] Youngmoon Jung, Yeunju Choi, Hyungjun Lim, Hoirin Kim, "A Unified Deep Learning Framework for Short-Duration Speaker Verification in Adverse Environments" 22 September 2020.

[6] Po-Sheng Chiu, Jia-Wei Chang, Ming-Che Lee, Ching-Hui Chen, Da-Sheng Lee "Enabling Intelligent Environment by the Design of Emotionally Aware Virtual Assistant: A Case of Smart Campus" 30 March 2020.

[7] Vinayak Iyer, Kshitij Shah, Sahil Sheth, Kailas Devadkar "Virtual Assistant For The Visually Impaired"26 July 2020.