Designing of Virtual Desktop Assistant using Machine Learning

Vijaya Balpande[1], Vedanti Lute[2], Neha Pawar[3], Saniya Sadaf[4], Aayush Jain[5]

Head of the Department, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4,5]
Priyadarshini J. L College of Engineering, Nagpur, Maharashtra, India

Abstract: *The rise of automation, along with increased computational power and improved accessibility to data, has resulted in the birth of the digital assistant market, popularly represented by Apple's Siri, Microsoft's Cortana, Google's Google Assistant. This paper focuses on the design of a Virtual Desktop Assistant (VDA) for accessing the data sources available on the user-generated dated content and providing knowledge from the knowledge database using API. This desktop application acts as a search tool where we have to give voice input and get output through voice and display on the screen. Speech recognition uses methodologies and technologies that enable the recognition and translation of spoken language into text. To identify the spoken words speech recognizer uses Natural Language Processing Algorithm (NLP) which includes Natural Language Understanding (NLU) and Name Entity Recognition (NER) also it undergoes six steps such as Tokenization, Stemming, Lemmatization, POS tags, NER, and Chunking. According to the following voice command, the data get fetched from the respective path and gives output through voice. We already know some voice assistants which are used nowadays. It overcomes drawbacks in the existing system such as Siri, which is not capable of doing work on the desktop application, and Cortana which comes installed with the Windows operating system which is not capable to work in another operating system such as Unix or Ubantu.*

Keywords: Speech recognition, Lemmatization, Stemming, Name Entity Recognition, API.

I. INTRODUCTION

In the new age, automation is rapidly increasing human-machine interaction and virtual assistants are software programs that can perform your day-to-day task, it can show weather reports, set an alarm making list, etc. This all can happen by only voice command. This voice-based assistant needs a wakeup word or hot word to start the assistant features. Virtual assistants are specially designed for desktop users and physically challenged people and blind people. This virtual assistant is working under technologies like natural language processing and natural language understanding. So, the way humans can interact with each other. Similarly, machines are also able to understand human language with the help of this technology. It may be Text-to-Speech or Speech-to-text. Virtual assistants are becoming smarter than ever. It can allow your intelligent assistants to make email work for you. Also, deliver customized responses and explore important information. This project is designed with suitable data available on the internet, that can be utilized to build a virtual assistant that has access to making a smart decisions for routine activities.

II. LITERATURE REVIEW

The field of virtual assistants having speech recognition has seen some major advancements or innovations. This is mainly because of its demand in devices like smartwatches or fitness bands, speakers, bluetooth earphones, mobile phones, laptop or desktop, television, etc. Almost all the digital devices which are coming nowadays are coming with voice assistants which help to control the device with speech recognition only. A new set of techniques is being developed constantly to improve the performance of voice automated search

Dr. Kshama V. Kulhali et.al, has proposed in [2], the most application of iPhone 'SIRI' which helps the user to communicate with mobile using voice and respond to the voice command. They have concluded that the device designed by them can work with or without the internet. They named it Personal Assistant with Voice Recognition Intelligence. Their

work is based on the concept of speech recognition. The drawbacks of the system are it does not work on desktop applications.

Kishore Kumar et.al has proposed in [1], an economically effective and performance-wise effective VA using Raspberry Pi for home automation based on the concept of the internet of things and speech recognition. The author has concluded that the device responds through voice and fetches search results from the Internet along with controlling the home appliances. The drawback of this device is that, it is not capable to work on desktop applications.

Veton kepuska has proposed in [3], that one of the goals of Artificial intelligence (AI) is the realization of natural dialogue between humans and machines. However. in this paper, the author has specified that the multi-modal dialogue systems which Process two or more combined user input modes, such as speech, image, video, touch, and head and body movement to design the next generation of Virtual Personal Assistants model. In their work, they have tested the new VPAs model by the IBM Watson cloud server with Python, and Node-Red.

We are developing a Virtual Assistant which involves all the features in one system such as opening applications, playing music, setting reminders, etc. so that the client doesn't have to go for different options. If we take the example of Alexa which is the personal assistant of Amazon, Alexa can't perform on a desktop. it is not able to open applications and other functions like making list and clicking pictures.

III. METHODOLOGY

For implementing a desktop virtual desktop assistant, where the speech recognition library has an inbuild function that will help the virtual assistant to understand commands given by the user and it will respond to the user invoices. The NLP algorithm will convert the user voice into text and according to the keywords present in the text respective action will be performed by the desktop virtual assistant. Speech recognition uses Natural Language Processing for the recognition of speech.

3.1 Working of the System

Speech recognition, or speech-to-text, is the ability of a machine or program to identify words spoken and convert them into readable text. Speech recognition systems use computer algorithms to and interpret spoken words and convert them into text. A software program turns a microphone into written language that computers and humans can understand.
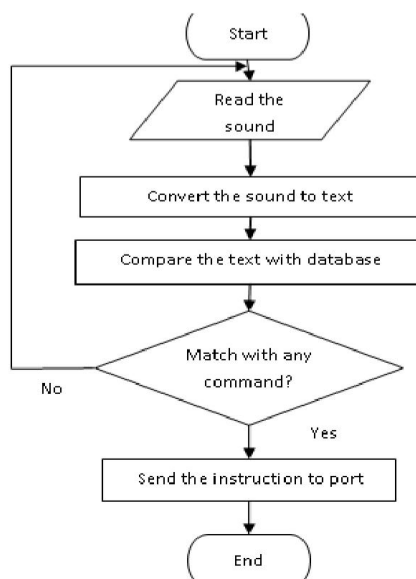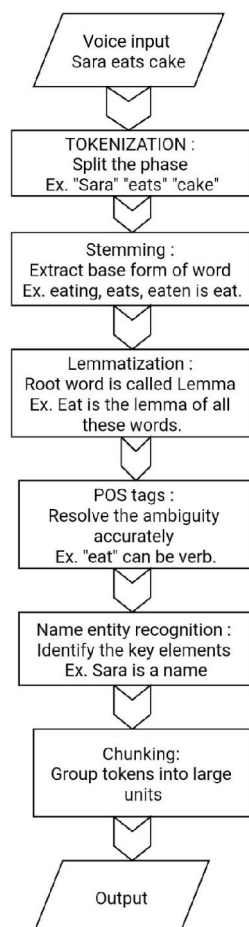

Figure 1: Flowchart for speech recognition [4]

3.2 Algorithm

- Natural Language Understanding (NLU)
- Named Entity Recognition (NER)

NLU is a type of natural language processing (NLP) makes it possible to carry out a dialog with a computer with the help of natural language.

Figure 2: Steps for conversion of speech to text

- **Tokenization:** Tokenization is essentially splitting a phrase, paragraph, or an entire text document into tokens. Each of these smaller units is known as a token. For example, Sara eats cake. Now the sentence can be broken down as ''Sara'' ''eats'' ''cake''.
- **Stemming:** Stemming is a technique that is used to extract the base form of the words by removing affixes. For example, the root of the words eating, eats, eaten is eat.
- **Lemmatization:** In Lemmatization root word is known as 'Lemma'. For example, eats, eating, eat are all forms of the word eat, therefore eat is the lemma of all these words.
- **POS tags:** PoS tagger is to resolve the ambiguity accurately based on the context of use. For example, the word "eat" can be a verb.
- **Name entity recognition:** Named entity recognition (NER) helps you easily identify the key elements in a text like names of people, places, brands, etc. For example, Sara is a name.
- **Chunking:** Chunking is a process to take small pieces of information and group them into large units.

After the conversion of speech to text, According to the following voice command, the data get fetched from the respective path such as website, desktop application, webpage, etc., and gives output through voice or display on the screen.

3.3 Experimental Implementation and Analysis

For implementing features like random, OS, Turtle, Myalarm Etc. each corresponding to these techniques we will use the libraries random for gaming, OS to implement operating System function like battery percentage, sleeping system, volume up, volume down. By using pyautogui will be able to take screenshots psutil library is useful for checking battery status. Some other modules are:

1. Speech recognition - Speech recognition is an important feature used in various forms like house automation and artificial intelligence devices. The important function of this library is to try to understand whatever the humans speak and converts the speech into text form.
2. CV2 - With this module we can capture images using the camera
3. DateTime - It works on time and date.
4. OS - It provides the function to interact with the operating system.
5. ime - It helps to showtime.
6. Web browser - It extracts data from the internet and this is an in-built package in python.
7. Subprocess - It is a standard library used to process various system commands like to sleep or to restart and shut down your PC.
8. Json- The JSON module is used for storing and exchanging data from API (Application Programming Interface).
9. request- The request module is used to send all types of HTTP requests. Its accepts URL as limitations and gives access to the given URL'S.

When the user uses the wake-up word the system gets activated and according to the command fetches the data from the respective path. The system is capable to perform operating system tasks like volume up, volume down, show battery percentage, tell date and time, etc.
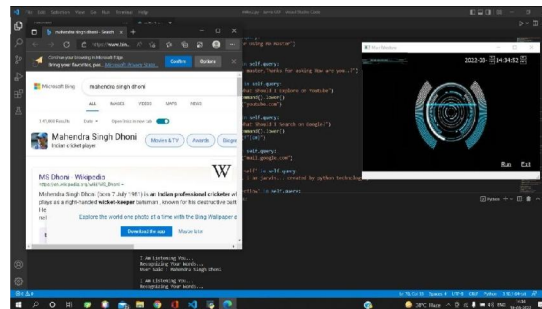
Figure 4: Wikipedia search

The application is capable of doing a web-based search like Wikipedia, Google Chrome, Netflix, Pinterest, etc.
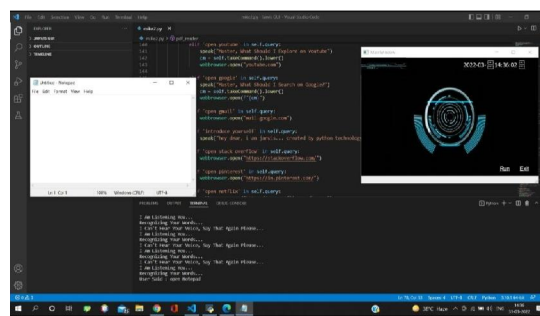
Figure 5: Open desktop application

The system is capable of opening desktop applications like notepad and able to write a note on it through voice.

IV. FUTURE SCOPE

The Virtual Assistants which are currently available are fast and responsive but still, there is a need to make them more advance. The future of this VA is incorporated with AI which includes Machine Learning, Neural networks, etc. Jarvis, that

is the VA development by us has set new standards of what we can achieve using voice activation. More features can be added such as functioning without a wake-up word, sending mail by typing the text through voice, recording video just by giving start and stop voice commands, etc.

V. CONCLUSION

In this paper, we have discussed a VA developed using python. This assistant currently works online and performs basic tasks like weather updates, streaming music, searching Wikipedia, playing music, opening desktop applications, etc. The system requires an internet connection to work. This Personal Assistant has been designed with ease of use as the main feature. The Assistant works properly to perform tasks given by the user.

It will overcome the drawbacks of the existing solutions. A similar application is Cortana which comes with Windows operating system and does not work in other operating systems but our system will work on any of the operating systems. There are many benefits of using this assistant such as it reduces labor costs, improves work quality, increases productivity, increases flexibility, etc.

REFERENCES

[1]. Kishore Kumar R, Ms. J. Jayalakshmi, Karthik Prasanna, "A Python-based Virtual Assistant Using Raspberry Pi for Home Automation", International Journal of Electronics and Communication Engineering (IJE CE), Volume 5, Issue 7, July 2018.

[2]. Dr. Kshama V. Kulhalli, Dr. Kotrappa Sirbi, Mr. Abhijit J. Patankar, "Personal Assistant with Voice Recognition Intelligence", International Journal of Engineering Research and Technology. ISSN 0974- 3154 Volume 10, Number 1 2017.

[3]. Veton Këpuska, "Next-Generation of Virtual Personal Assistant (Microsoft Cortana, Apple Siri, Amazon Alexa and Google 11. Home)", Pycon, Cleveland, 2018.

[4]. www.semanticscholar.org/paper/Speech-Recognition-for-Robotic-Control-Gosavi-Khot/54e5bf6735476d0217367f929d374f8739901e4f