# Virtual desktop Voice Assistant

# Statement of Project Report

# Preparation

1. The title: Virtual desktop Voice Assistant

**2.** Degree for which the report is submitted: Bachelors in computer science and engineering

3 Project Supervisor was referred to for preparing the report.

1. Specifications regarding thesis format have been closely followed.
2. The contents of the thesis have been organized based on the guidelines.
3. The report has been prepared without resorting to plagiarism.
4. All sources used have been cited appropriately.

8 The report has not been submitted elsewhere for a degree.

Naman Raj Sharma

# ABSTRACT

A desktop voice assistant or intelligent personal assistant is a software agent that can perform tasks or services for an individual based on verbal commands i.e. by interpreting human speech and respond via synthesized voices. Users can ask their assistants’ questions, control home automation devices, and media playback via voice, and manage other basic tasks such as email, to-do lists, open or close any application, etc. with verbal commands. Desktop voice assistants work via Text, including: online chat (especially in an instant messaging app or other app), SMS Text, e-mail or other text-based communication channel, for example Conversica's Intelligent Virtual Assistants for business. Desktop voice assistants come in somewhat small packages and can perform a variety of actions after hearing your command. They can turn on lights, answer questions, play music, place online orders and do all kinds of AI-based stuff. Desktop voice assistants are not to be confused with virtual assistants, which are people who work remotely and can, therefore, handle all kinds of tasks. Rather, Desktop voice assistants are technology based. As Desktop voice assistants become more robust, their utility in both the personal and business realms will grow as well.

Having said that, how cool it would be to build a simple voice-based desktop/laptop assistant that has the capability to: - Open the YouTube, Google, and Wikipedia in the browser, Send an email to your contacts, Tells you the current weather and temperature of almost any city, Tells you the current time, Play you a song installed in your laptop/desktop, Tells you latest news feeds, Tells you about almost anything you ask.

*Keywords—Voice assistant, Speech Recognition, Low cost, Internet, Speech Synthesis, and Visually Challenged.*

6 |

## DECLARATION

Project Title: “Virtual desktop Voice Assistant”

Degree for which the project work is submitted**: Bachelors in Computer Science and Engineering**

I declare that the present project represents largely our own ideas and work in our own words.

Where others ideas or words have been included, I have adequately cited and listed in the reference materials. The report has been prepared without resorting to plagiarism. I have adhered to all principles of academic honesty and integrity. No falsified or fabricated data have been presented in the report. I understand that any violation of above will case for the above will cause for disciplinary action by the Institute, including revoking the conferred degree, if conferred and can also evoke penal action from the sources which have not been properly cited or from whom proper permission has not been taken.

Date:7-07-2021

7 |

**Muskan(18SCSE1010401),**

**Monish Verma(18SCSE1010353), Naman Raj Sharma(18SCSE1010066)**

# Table of Content

|  |  |  |
| --- | --- | --- |
| S.No. | Title | Page no. |
| 1. | Introduction |  |
| 2. | Literature Reviews/Comparative study |  |
| 3. | **Problem Formulation** |  |
| 4. | **Required tools** |  |
| 5. | **Methodology** |  |
| 6. | **Feasibility Analysis** |  |
| 7. | **Complete work plan layout** |  |
| 8. | **Implementation** |  |
| 9. | **Future Plan** |  |
| 10. | **Conclusion** |  |
| 11. | **References** |  |

8 |

1. **Introduction**

The Growth of Artificial Intelligence (AI) systems that cancoordinate natural interactions between humans and machines through voice,communication,gestures,facial expressions, and many more–– are gaining popularity today. It is no longer a human who learns to speak with a machine; instead, a machine is learning to interact with a human, analysing his actions, habits, and behaviour in the hopes of becoming his customised assistant.

Work has been going on for a long time to build and improve such customized assistants. These systems are continuously evolving and improving, go beyond personal computers ,and in various mobile devices and gadgets have already firmly developed themselves . One of the most popular voice assistant sare Siri, fromApple, Amazon Echo,which responds to the name of Alex from Amazon, Cortana from Microsoft, Google Assistant from Google, and the newly emerged "AIVA" intelligent assistant. A short introduction to the design and development of voice assistants is given in SectionI,II. Aproposed work schedule is given in Section III Section IV includes methodology of the work of a speech assistant AIVA. Section V provides future plan of voice assistant. SectionVI explains the voice assistant's test performance. SectionsVII and VIII define an assistant's conclusion and future scope using different artificial intelligent algorithms, and include a comparative assessment of algorithm learning ability. The main purpose of this work is to create a local voice assistant that performs the human task and the everyday task that a human being requires to perform in daily life. AIVA (2018) aimed at creating a personal voice-controlled assistant that does a lot of things,such as searching the Internet. It has some new features, such as making comments on social media platforms like Facebook, Twitter, etc. with only a few simple commands.

9 |

### Problem Statement

The aim of this project is to create a desktop application which work on commands given by the user. Whenever user give some input as a command then application recognize it and give response according to it The project should be simple to use, allowing even a beginner to use it.

10 |

**2.** **Literature Reviews/Comparative study**

 Artifical Intelligence refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal.

AI is a digital virtual assistant that makes collaboration easier. AI voice assistants mainly accept input by voiceWhen designing a virtual assistant, you’ll likely want to have a trigger word/wake word that wakes up your assistant and has it wait for instructions. You know probably don’t want to force your users to hit a button every time they want to interact with your assistant, which is why Amazon and Google have wake words like “Alexa” and “Okay, Google”. **Voice assistant**  is a software agent that can perform tasks or services for an individual based on commands or questions.

11 |

1. **Problem Formulation**

Have you ever wondered how cool it would be to have your own A.I. assistant? Imagine how easier it would be to send emails without typing a single word, doing Wikipedia searches without opening web browsers, and performing many other daily tasks like playing music with the help of a single voice command.

### The Drawbacks ofVoice Assistants

As the acceptance and usage of voice assistants continues to grow, it is only natural for some people to have reservations about using them. Below, we discuss some of the major issues regarding voice assistants.

**Privacy**: Privacy is a concern, especially involving smart speakers. While waiting for a wake word, smart speakers are always listening. On a smartphone, pressing a button or opening an app activates the assistant. Once you wake it up, it begins recording audio clips of what you say. These clips represent the files that go to a server to process the audio and formulate a response. The real brains are not in the little speakers in our homes: They’re on a massive server somewhere else. What the speaker sends is on an encrypted connection. Speakers do not record anything prior to the wake words.  
  
“People confuse ‘always listening’ with ‘always recording,’” says Mutchler of Voicebot.ai. “The genius of [smart speakers] is they can remove background noise and single out the wake word,” she continues. Only then do they begin recording.

**Accuracy**: Voice assistants don’t always understand what we are asking. Sometimes, it’s how we speak. Other times, it is simply because the artificial intelligence hasn’t yet learned how to do something.

12 |

Hackability and Security: Even though voice assistants communicate with their servers using encrypted connections, there is still a concern about hackability and security.  
  
In early 2018, some users of Amazon’s Echo reported it would suddenly emit an evil laugh for no reason. In the beginning, people thought someone had hacked into their smart speakers. Amazon investigated the problem and later announced that the Echo had been hearing words similar to “Alexa laugh,” so it began laughing. As a response, Amazon disabled the reaction and changed Alexa’s response to a user’s request that it laugh to “Sure, I can laugh,” followed by laughter.

**The Future of Voice Assistants**

The number of people using voice assistants is expected to grow. According to the Voicebot Smart Speaker Consumer Adoption Report 2018, almost ten percent of people who do not own a smart speaker plan to purchase one. If this holds true, the user base of smart speaker users will grow 50 percent, meaning a quarter of adults in the United States will own a smart speaker.

Smart speaker sales are expanding in other parts of the world, meaning they need to “learn” how to “understand” languages, accents, dialects, slang, and nuances in each country in which they are sold. Chinese companies are developing their own smart speakers. “The rest of the world is behind the U.S. and will catch up pretty quickly,” Mutchler says.

Voice assistants are always improving and “learning.” AI companies use data from existing systems to improve what assistants can do. Lucas believes that ultimately, the voice assistant might get so smart that it will automatically order a pizza if you say you’re hungry. It will use existing data from your previous purchases to come to the conclusion that saying you’re hungry equals ordering a pizza.

13 |

The experts predict that voice assistants will improve in many other ways. As described in a 2017 article for The Atlantic, “A subfield of AI called computational creativity forges algorithms that can write music, paint portraits, and tell jokes.” These capabilities will help smart speakers “show emotion” and “think” for themselves without being scripted. Systems that explain why they did what they did and what they’re going to do next are also on the horizon.

Voice assistants are not going anywhere. “I think people thought of it as a fad, but it’s not. It’s changing what people do in their homes. Voice assistants will grow and are here to stay,” Mutchler says. “I think they [voice assistants] will be in everything, and the smart speaker might fade away in a few years because many technologies, like televisions and refrigerators, will have their own voice assistants. The kids today won’t understand that there was a world where you couldn’t talk to things,” she concludes.

14 |

1. **Required tools**

The Voice Assistant for Desktop is made using the Python Language v.3.0+ As we know Python is a suitable language for script writers and developers. Let’s write a script for Voice Assistant using Python. The query for the assistant can be manipulated as per the user’s need.

Speech recognition is the process of converting audio into text. This is commonly used in voice assistants like Alexa, Siri, etc. Python provides an API called SpeechRecognition to allow us to convert audio into text for further processing. In this article, we will look at converting large or long audio files into text using the SpeechRecognition API in python.

**Modules needed**

**Subprocess**:- This module is used for getting system subprocess details which are used in various commands i.e Shutdown, Sleep, etc. This module comes buit-in with Python.

**Wolframalpha**:- It is used to compute expert-level answers using Wolfram’s algorithms, knowledgebase and AI technology. To install this module type the below command in the terminal.

pip install wolframaplha

**Pyttsx3**:- This module is used for conversion of text to speech in a program it works offline. To install this module type the below command in the terminal.

pip install pyttsx3

15 |

**Wikipedia**:- As we all know Wikipedia is a great source of knowledge just like GeeksforGeeks we have used Wikipedia module to get information from Wikipedia or to perform Wikipedia search. To install this module type the below command in the terminal.

pip install wikipedia

**Speech Recognition**:- Since we’re building an Application of voice assistant, one of the most important things in this is that your assistant recognizes your voice (means what you want to say/ ask). To install this module type the below command in the terminal.

pip install SpeechRecognition

**Web browser**:- To perform Web Search. This module comes buit-in with Python.

**Pyjokes**:-Pyjokes is used for collection Python Jokes over the Internet. To install this module type the below command in the terminal.

pip install pyjokes

**Requests**: Requests is used for making GET and POST requests. To install this module type the below command in the terminal.

pip install requests

**BeautifulSoup**: Beautiful Soup is a library that makes it easy to scrape information from web pages. To install this module type the below command in the terminal.

pip install beautifulsoup4

16 |

1. **Methodology**

**Speech-Recognition**

The system which uses Google's online speech acknowledgment system for changing over speech contribution to text. Texts from the uncommon organization coordinated on the PC network worker at the data community can be gotten by clients from the mouth piece incidentally stored in the gadget ,and is then sent for speech acknowledgment to Google cloud.

**Python Backend**

The python back end gets the yield from the speech acknowledgment module and afterward decides whether an API Call, Context Extraction, and System Call are the order or the speech yield. The yield is then sent back to the backend of the python to send the client the essential yield.

**API Calls**

API represents Application Programming Interface. An API is software intermediary programming that empowers two applications to impart to one another. All in all, the courier that sends the solicitation to the help you demand it from is an API, and afterward returns the response to you.

17 |

**Context Extraction**

The way toward removing organized data automatically from unstructured as well as semi-organized machine-comprehensible reports is context extraction (CE). Much of the time, this movement includes preparing human language texts through the normal language handling (NLP). Most recent media archive preparing exercises ,like automated explanation and substance extraction from pictures/sound/video , might be deciphered as TESTRESULTS context extraction.

**System Calls**

A system bring in registering is the automatic way where a product program demands a help from a working system bit on which it is executed. This can incorporate equipment related administrations, (for example, getting to a hard plate drive), new cycles being produced and executed, and coordination with basic portion administrations, like planning of cycles. System calls give a fundament a linter face between the working system and an interaction.

**Text-To-Speech**

Text-to-Speech (TTS) alludes to the capacity to peruse text so anyone might hear on PCs. A TTS Engine makes an interpretation of composed text into a phonemic portrayal, and afterward changes the phonemic portrayal into sound yield waveforms. TTS motors are accessible to outsider distributers of different dialects, tongues and concentrated vocabulary.

18 |

1. **Feasibility Analysis**

As this program includes the functions and services of: transformation, mail exchange, music player service, checking weather, Google searching engine, and Bluetooth headset support. The list below indicates the information and the requirements of each individual function.

The program has two modes to well fetch the services and functions. The program will start with voice mode as its primary mode to provide the voice assistant, but the user can select switching to the text mode if he or she is not well working with the voice mode or the surrounds don’t support the voice recognition well.

* Mail exchange, customers are able to send the mail to the person with mail address in the contacts. By giving a correct command contains the mail request keyword together with the destination person; the mail should be received by the recipient after it has been sent.
* Music player service, the music player offers the services to the user to play a named or a randomly picked song in the pre-stored song list on the mobile phone. And it could be stopped when the user wants to terminate it.
* Checking weather, the user could check the weather in any place. In addition, the weather is returned with the temperature and humidity; the user could also check the weather for current day, tomorrow or in next four days.
* Google searching engine, the search engine enable the user to search anything on Google. The search engine will give result list back and displayed on the browser.
* Bluetooth headset support, since it is not possible to do the voice recognition while the music player is playing or the surroundings are noisy; the Bluetooth headset support makes it possible to speak to the headset rather than the desktop if the user enables it.

19 |

This analysis showed how the voice recognition system worked in an integrated voice based delivery system for the purpose of delivering instruction. An added importance of the study was that the voice system was an independent speech recognition system. At the time this study was conducted, there did not exist a reasonably priced speech recognition system that interfaced with both graphics and authoring software which allowed any student to speak to the system without training the system to recognize the individual student's voice. This feature increased the usefulness and flexibility of the system.

The methodology for this feasibility analysis was a development and evaluation model. This required a market analysis, development of the voice system and instructional course ware, testing the system using a sample population. The data collection approach was multifaceted. There were surveys to be completed by each subject: a student profile survey, a pretest, a posttest, and an opinion survey about how well the instruction met expectations. Data was also collected concerning how often the recognition system recognized, did not recognize, or misrecognized the voice of each subject. The information gathered was analyzed to determine how well the voice recognition system performs in a training delivery application.

20 |

1. **Complete work plan layout**

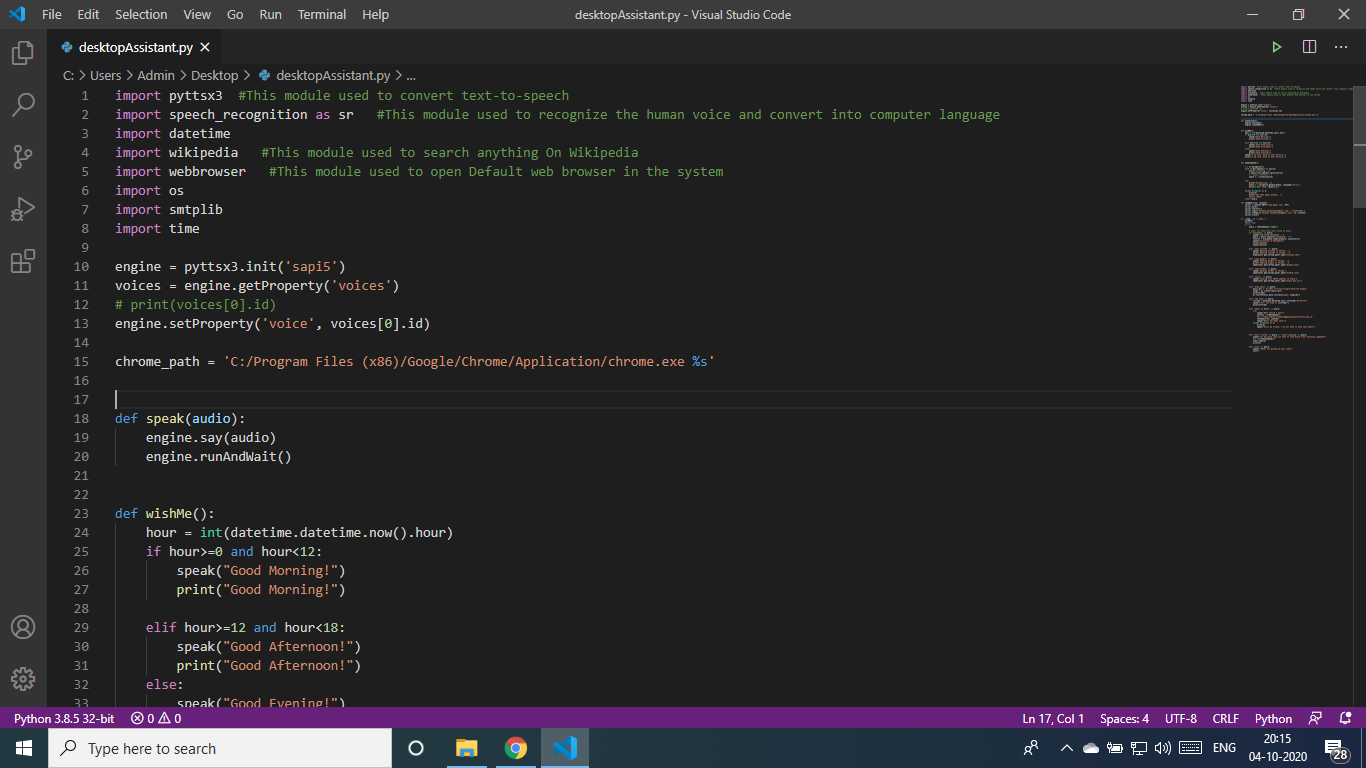
The project work is organized based on the actual task for the designing, implementation, test and optimization. As it has been primary planned, each of the developers worked 5 days a week; 3 days for implementation, and 2 days for testing and summarizing the work. Apart from the designing, implementation and testing, developers also defined the work plan every time before the implementation and improve the project after the accomplishment of each individual section.

Developers communicate though the Whatsapp , Facebook , instagram and Gmail for sharing the ideas and discuss the project. Data statistics and relative materials is collected and shared through gmail. Mostly the work was done by pairing programming, that is, every time developers made a meeting and set together for designing, figure out a valid solution and doing the implementation together.

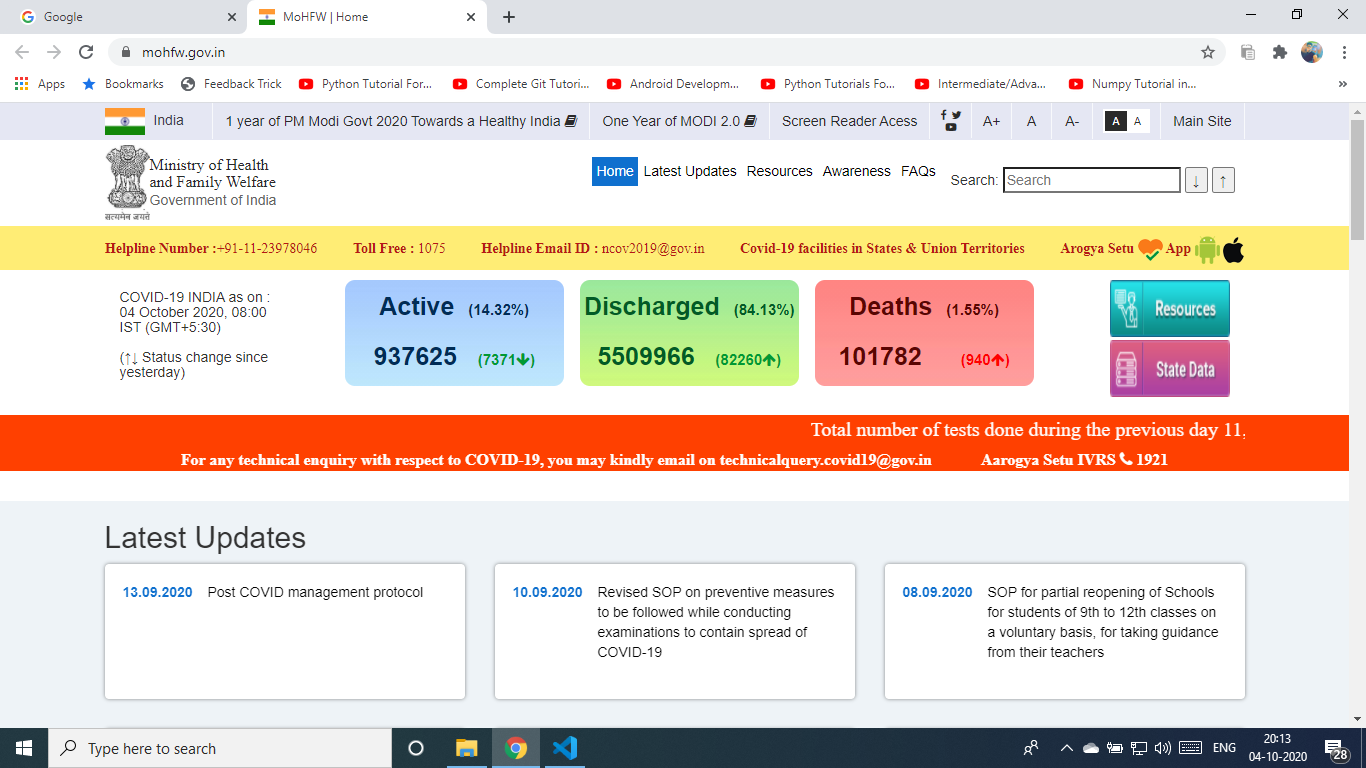
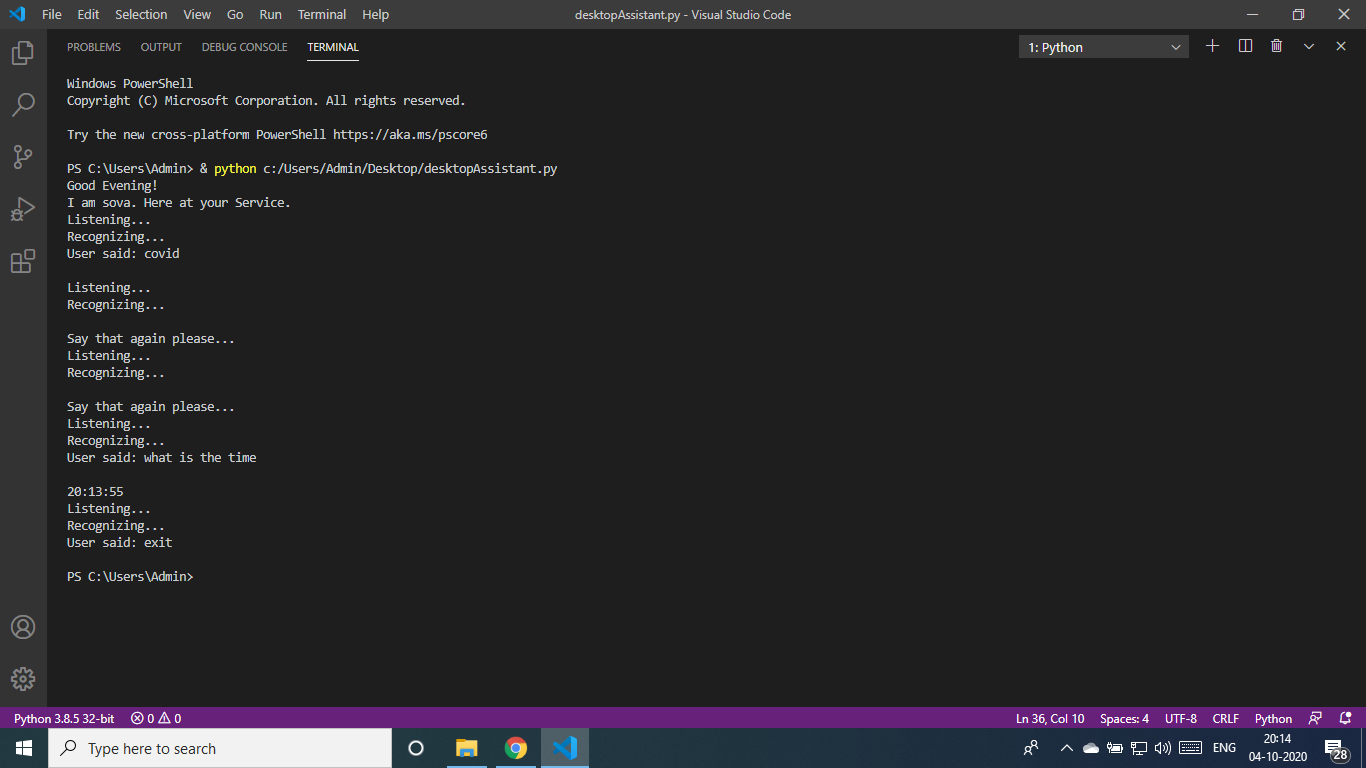
The designing and the framework was done together, and the individual implementation of functions was assigned to different developers, but the developer was not only caring his own part, but also considering the whole program.

21 |

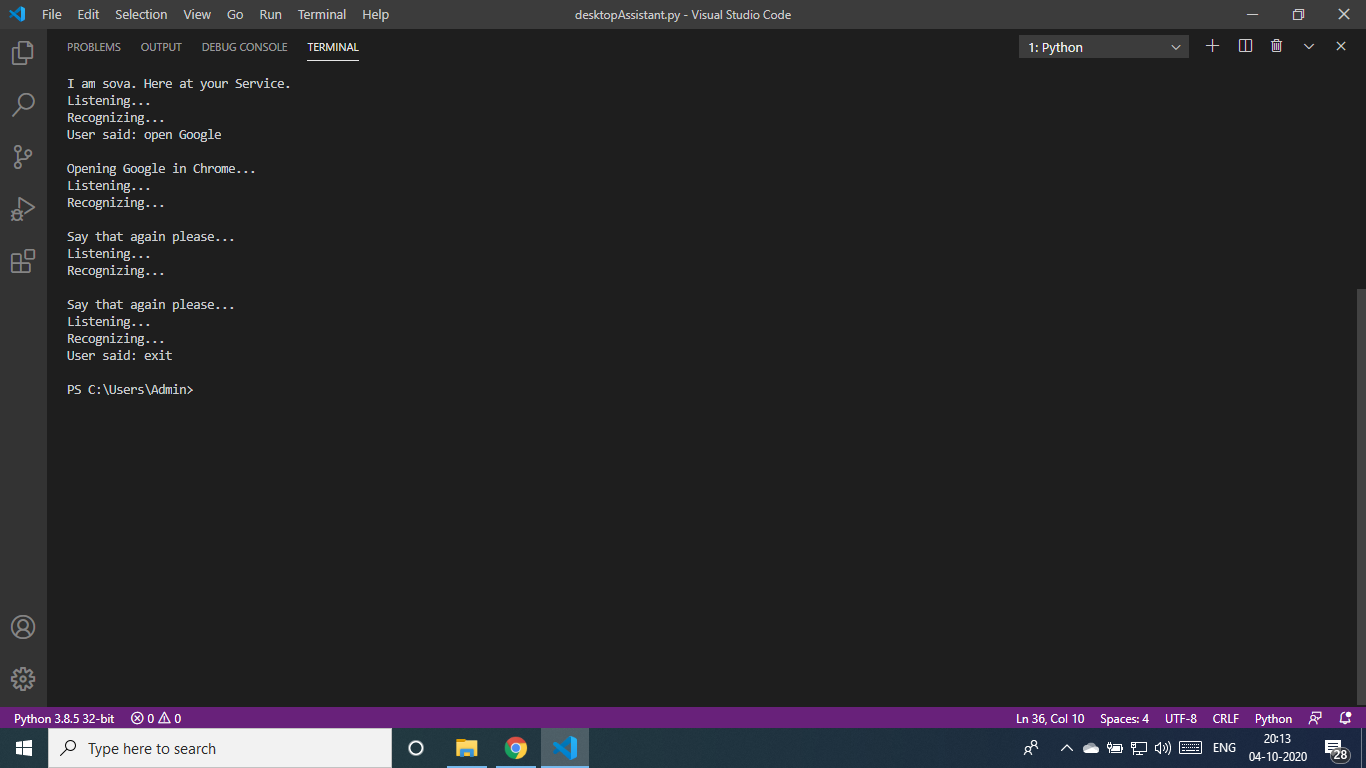
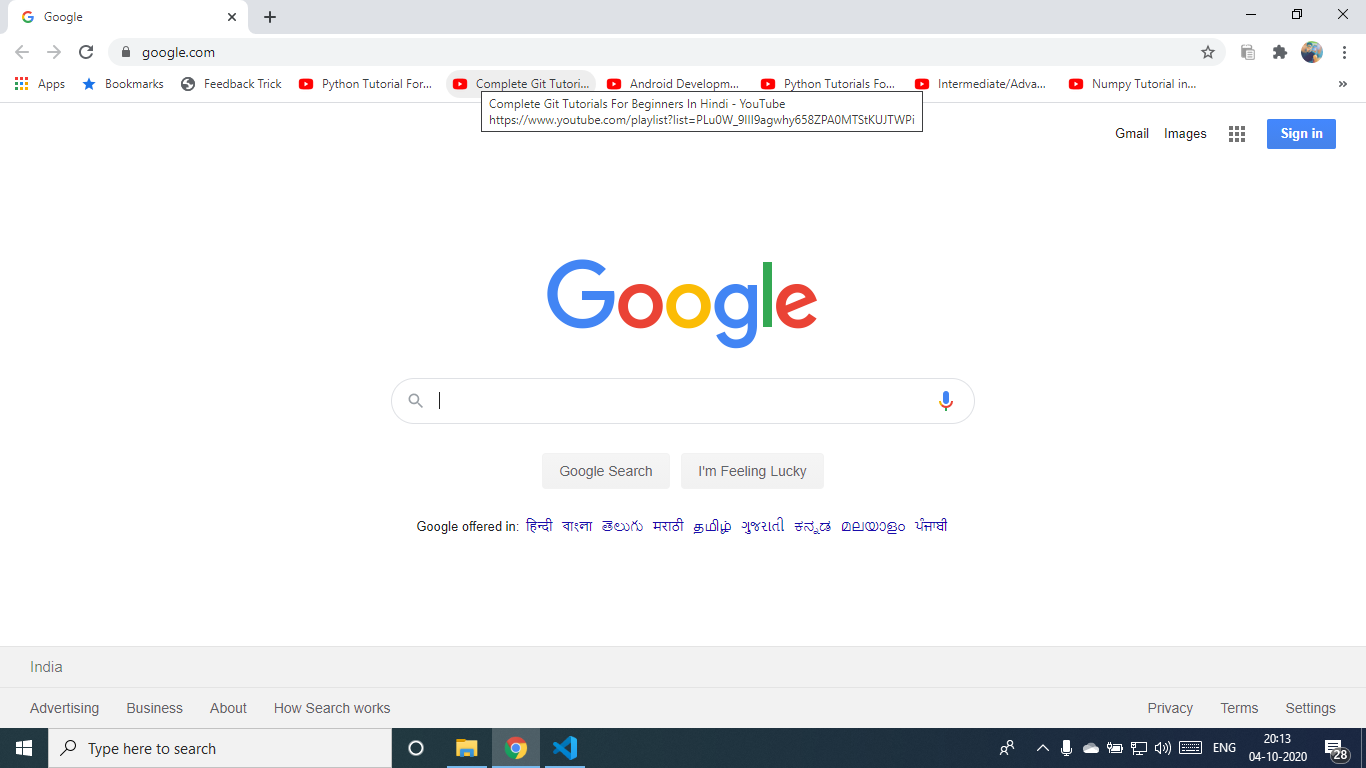
1. **Implementation**

****

22 |



23 |

**9. Future Plan**

24 |

We propose that an android application be developed based on the survey, which fulfils the desires of multiple users. The key reason the user needs to use the voice assistant is to make their life simpler, so it is possible to facilitate the user by implementing the features mentioned below.

* + From sports to technology, read the news.
  + Voice control of your songs.
  + Use Google to find what you need.
  + Available in Spanish, German, French, Italian and more than70 other languages.
  + Link with Facebook ,Twitter and WhatsApp social media.
  + For anyplace , get weather forecasts.
  + Manage your schedule to figure out what your day feels like easily.
  + In the area, find the best restaurant or bar.
  + Get directions to friends , restaurants and shops and navigate.
  + Find places on Google Maps easily.
  + Set alerts and notifications and, if necessary, warn you at a particular time.
  + Read and send emails and text messages.
  + Making phone calls ,find specifics of contacts.

25 |

**10.Conclusion**

In this research paper we discussed the design and implementation of a Digital Assistance. The project is built using open source software modules that can handle any updates in the near future with PyCharm community backing.

The modular nature of this project makes it more flexible and easy to add additional features without disturbing current system functionalities.

It not only works on human commands, but also provides the user with responses based on the question being requested or the user's words, such as opening tasks and operations. It greets the user in a way that makes the user feel more comfortable and free to communicate with the voice assistant.

The application should also eliminate any kind of unnecessary manual work required in the user life of performing each and every task. The entire system works on the verbal input rather than the text one.

It chips away at human orders, yet in addition furnishes the client with reactions dependent on the inquiry being mentioned or the client's words, like opening errands and tasks. It welcomes the client such that causes the client to feel more good and allowed to speak with the voice assistant. The application ought to likewise take out any sort of super fluous manual work needed in the client life of playing out every single errand. Rather than the content one, the entire system works on verbal info.

26 |

# 11. REFERENCES

1. Mammone, R., Xiaoyu Zhang, & Ramachandran, R. (1996). strong speaker recognition: a function- primarily based approach. IEEE signal Processing mag,thirteen(five),fifty-eight. https://doi.org/10.1109/seventy nine.536825.
2. Nuance offers free voice reputation on Dragon cell Assistant. (2014), 2014(5), 2.

https://doi.org/10.1016/s0969-4765(14)70078-five

1. ColllinaSzy, J., Bundzel, M., &Zolotova, I. (2017). Implementation of wise software program using IBM Watson and Bluemix. Acta Electrotechnica Et Informatica, 17(1), fifty-eight-63. https://doi.org/10.15546/aeei-2017-0008
2. Nedić, B. (2019). Gartner's pinnacle Strategy technology traits. proceedings on Engineering Sciences, 1(2), 433-442.

https://doi.org/10.24874/pes01.02.041

1. Ji, X., & Rau, P. (2018). development and alertness of a classification device for Voice wise sellers. international journal Of Human–laptop interaction, 35(9), 787-795.

https://doi.org/10.1080/10447318.2018.1496969

1. Chang, J., Lee, M., & Wang, T. (2016). Integrating a semantic-based retrieval agent into case-primarily based reasoning structures: A case observe of a web bookstore. computer systems in enterprise, seventy- eight, 29-42.

https://doi.org/10.1016/j.compind.2015.10.007.

1. Garcı́a-Serrano, A., Martı́nez, P., & Hernández, J. (2004). the usage of AI techniques to assist superior interaction skills in a digital assistant for e- commerce. professional systems with

27 |

applications,26(3),413-426. https://doi.org/10.1016/j.eswa.2003.09.012 .

1. Von der Pütten, A., Krämer, N., Gratch, J., & Kang,

S. (2010). “It doesn’t count number what you are!” Explaining social consequences of retailers and avatars. computer systems In Human conduct,26(6),1641-1650. https://doi.org/10.1016/j.chb.2010.06.012

1. Li, J. (2015). A survey of experimental works evaluating copresent robots, telepresent robots and digital dealers. worldwide journal Of Human-computer studies,seventy-seven,23-37. https://doi.org/10.1016/j.ijhcs.2015.01.001.
2. Purwaningsih, R. and Yenifi, I., 2015. Usability assessment of worldwide workplace website of Diponegoro college with situation-based Usability evaluation approach and Wammi technique. ComTech: laptop, arithmetic and Engineering programs, 6(three), p.329.
3. M, H. (2019). developing humans's information competencies: fostering statistics literacy in instructional, administrative center and network contexts. Webology, 16(1), 210-212. doi: 10.14704/net/v16i1/bookreview30
4. Berdasco, López, Diaz, Quesada, & Guerrero. (2019). user revel in contrast of wise private Assistants: Alexa, Google Assistant, Siri and Cortana. proceedings, 31(1), 51. doi: 10.3390/proceedings2019031051
5. Smart item detector for visually impaired. (2017). worldwide magazine of latest trends In Engineeringresearch,three(three),192-195. doi: 10.23883/ijrter.conf.20170331.038.5c8ti
6. Digital Assistant App for Disabled people. (2019). global magazine of innovative era And Exploring Engineering, 9(2S), 326-329. doi: 10.35940/ijitee. b1086.1292s19

28 |

1. Baranchuk, V. (2019). digital assistance statistics model for public offerings. Yugra nation college Bulletin,15(2),7-15.doi: 10.17816/byusu20190207-15
2. Zahour, O. (2020). toward a Chatbot for educational and vocational steering in Morocco: Chatbot E-Orientation. worldwide magazine Of advanced developments In laptop science And Engineering,9(2),2479-2487doi:10.30534/ijatcse/2020/237922020
3. Gong, L.: purpose of entry, CA (US) u. s. US 2003.01671.67A1 (12) Patent
   1. Application faucet houselication c (10) tavern. No.: North American nation 2003/0167167 A1 Gong (43) faucet house. Date: 4 Sept 2003 for Intelligent Virtual Assistant
4. Gong, L.: purpose of entry, CA (US) u. s. US 2003.01671.67A1 (12) Patent
   1. computing (AI), generally stated as machine
   2. B. Marr, The marvellous ways that within which GoogleUses Deep Learning AI.
5. Cortana Intelligence, Google Assistant, Apple Siri.

29 |