

Voice Assistance using Python

Subhash Mani Kaushal¹ Dr. Megha Mishra² Dr. Vishnu Kumar Mishra³

¹Research Scholar ²Associate Professor ³Professor

^{1,2,3}Department of Computer Science and Engineering

^{1,2,3}Shrishankaracharya Group of Institute, Bhilai, C.G., India

Abstract— Artificial intelligence technologies are beginning to be actively used in human life, this is facilitated by the appearance and wide dissemination of cloud computing. Autonomous devices are becoming smarter in their way to interact with both a human and themselves. New capacities lead to creation of various systems for integration of smart things into Social Networks of the Internet of Things. One of the relevant trends in artificial intelligence is the technology of recognizing the natural language of a human. New insights in this topic can lead to new means of natural human-machine interaction, in which the machine would learn how to understand human's language, adjusting and interacting in it. One of such tools is voice assistant, which can be integrated into many other intelligent systems. In this paper, the principles of the functioning of voice assistants are described, its main shortcomings and limitations are given. The method of creating a local voice assistant without using python is described, which allows to significantly expand the applicability of such devices in the future.

Keywords: Voice Assistance, Python, Speech to Text Converter, Voice Input

I. INTRODUCTION

Virtual Assistant is a software agent that is used to perform task or service for individual. The software focuses on virtual assistant and structural elements of a virtual assistant system. In this software we tried to study virtual Environment and virtual Assistant. As mentioned above virtual assistant personal comes to the picture where the user can search files from system, weather forecast, time, website, find system IP address, jokes by using voice command. Today the development of artificial intelligence (AI) systems that are able to organize a natural human-machine interaction (through voice, communication, gestures, facial expressions, etc.) are gaining in popularity. One of the most studied and popular was the direction of interaction, based on the understanding of the machine by the machine of the natural human language. It is no longer a human learns to communicate with a machine, but a machine learns to communicate with a human, exploring his actions, habits, behavior and trying to become his personalized assistant. The work on creating and improving such personalized assistants has been going on for a long time. These systems are constantly improving and improving, go beyond personal computers and have already firmly established themselves in various mobile devices and gadgets. One of the most popular voice assistants are Siri, from Apple, Amazon Echo, which responds to the name of Alex from Amazon, Cortana from Microsoft, Google Assistant from Google, and the recently appeared intelligent assistant under the name "AIVA". Each company-developer of the intelligent assistant applies his own specific methods and approaches for development, which in turn affects the final product. One assistant can synthesize speech more

qualitatively, another can more accurately and without additional explanations and corrections perform tasks, others are able to perform a narrower range of tasks, but most accurately and as the user wants. Obviously, there is no universal assistant who would perform all tasks equally well. The set of characteristics that an assistant has depends entirely on which area the developer has paid more attention. Since all systems are based on machine learning methods and use for their creation huge amounts of data collected from various sources and then trained on them, an important role is played by the source of this data, be it search systems, various information sources or social networks. The amount of information from different sources determines the nature of the assistant, which can result as a result. Despite the different approaches to learning, different algorithms and techniques, the principle of building such systems remains approximately the same. The main technologies are voice activation, automatic speech recognition, Teach-To-Speech, voice biometrics, dialog manager, natural language understanding and named entity recognition.

II. PROPOSED WORK:

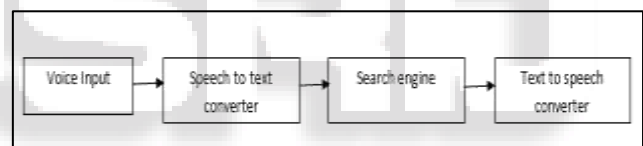


Fig. 2.1: Block diagram

A. Speech recognition:

The system uses Google's online speech recognition system for converting speech input to text. The speech input Users can obtain texts from the special corpora organized on the computer network server at the information center from the microphone is temporarily stored in the system which is then sent to Google cloud for speech recognition. The equivalent text is then received and fed to the central processor. Speech recognition is an interdisciplinary subfield of computer science and computational linguistics that develops methodologies and technologies that enable the recognition and translation of spoken language into text by computers. It is also known as automatic speech recognition (ASR), computer speech recognition or speech to text (STT). It incorporates knowledge and research in the computer science, linguistics and computer engineering fields.

Some speech recognition systems require "training" (also called "enrollment") where an individual speaker reads text or isolated vocabulary into the system. The system analyzes the person's specific voice and uses it to fine-tune the recognition of that person's speech, resulting in increased accuracy. Systems that do not use training are

called "speaker independent"^[1] systems. Systems that use training are called "speaker dependent".

Speech recognition applications include voice user interfaces such as voice dialing (e.g. "call home"), call routing (e.g. "I would like to make a collect call"), domestic appliance control, search key words (e.g. find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g. a radiology report), determining speaker characteristics, speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed direct voice input).

The term voice recognition or speaker identification refers to identifying the speaker, rather than what they are saying. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on a specific person's voice or it can be used to authenticate or verify the identity of a speaker as part of a security process.

From the technology perspective, speech recognition has a long history with several waves of major innovations. Most recently, the field has benefited from advances in deep learning and big data. The advances are evidenced not only by the surge of academic papers published in the field, but more importantly by the worldwide industry adoption of a variety of deep learning methods in designing and deploying speech recognition systems.

B. Search engine using cloud computing:

Google Cloud Platform (GCP), offered by Google, is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail, file storage, and YouTube. Alongside a set of management tools, it provides a series of modular cloud services including computing, data storage, data analytics and machine learning.^[2] Registration requires a credit card or bank account details.

Google Cloud Platform provides infrastructure as a service, platform as a service, and serverless computing environments.

In April 2008, Google announced App Engine, a platform for developing and hosting web applications in Google-managed data centers, which was the first cloud computing service from the company. The service became generally available in November 2011. Since the announcement of the App Engine, Google added multiple cloud services to the platform.

Google Cloud Platform is a part of Google Cloud, which includes the Google Cloud Platform public cloud infrastructure, as well as Google Workspace (formerly G Suite), enterprise versions of Android and Chrome OS, and application programming interfaces (APIs) for machine learning and enterprise mapping services.

C. Text-To-Speech Text-to-Speech (TTS):

Refers to the ability of computers to read text aloud. A TTS Engine converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be output as sound. TTS engines with different

languages, dialects and specialized vocabularies are available through third-party publishers. Text to speech (TTS) is the use of software to create an audio output in the form of a spoken voice. The program that is used by programs to change text on the page to an audio output of the spoken voice is normally a text to speech engine. TTS engines are needed for an audio output of machine translation results.

D. Python:

Python is an interpreted, high-level and general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

III. CONCLUSION

In this paper, we discussed the design and implementation of a Digital Assistance. The project is built using open source software modules with PyCharm community backing which can accommodate any updates in the near future. 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 give responses to the user on the basis of query being asked or the words spoken by the user such as opening tasks and operations. It is greeting the user the way user feels more comfortable and feels free to interact 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.

IV. REFERENCES:

- [1] AbhayDekate, ChaitanyaKulkarni, RohanKilledar, "Study of Voice Controlled Personal Assistant Device", International Journal of Computer Trends and Technology (IJCTT) – Volume 42 Number 1 – December 2016.
- [2] Deny Nancy, Sumithra Praveen, AnushriaSai, M.Ganga, R.S. Abisree, "Voice Assistant Application for a college Website", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6S5, April 2019.
- [3] Deepak Shende, RiaUmahiya, Monika Raghorte, AishwaryaBhisikar, AnupBhange, "AI Based Voice Assistant Using Python", Journal of Emerging Technologies and Innovative Research (JETIR), February 2019, Volume 6, Issue 2,
- [4] Dr.Kshama V. Kulhalli, Dr.KotrappaSirbi, 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).
- [5] Isha S. Dubey, Jyotsna S. Verma, Ms. Arundhati Mehendale, "An Assistive System for Visually Impaired using Raspberry Pi", International Journal of Engineering Research & Technology (IJERT), Volume 8 Issue 05, May-2019.
- [6] 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 (IJECE), Volume 5, Issue 7, July 2018.
- [7] M. A. Jawale, A. B. Pawar, D. N. Kyatanavar, "Smart Python Coding through Voice Recognition", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-10, August 2019.
- [8] Rutuja V. Kukade, Ruchita G. Fengse, Kiran D. Rodge, Siddhi P. Ransing, Vina M. Lomte, "Virtual Personal Assistant for the Blind", International Journal of Computer Science and Technology (JCST), Volume 9, Issue 4, October - December 2018.
- [9] Tushar Gharge, Chintan Chitroda, Nishit Bhagat, Kathapriya Giri, "AI-Smart Assistant", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 01, January 2019.
- [10] Veton Këpuska, "Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google

