ISSN No:-2456-2165

Library Management Using Voice Assistant

Samiksha Gaikwad
Department of Computer Engineering
International Institute of Information Technology
Pune, India

Shriya Purandare
Department of Computer Engineering
International Institute of Information Technology
Pune, India

Swathi Balaji
Department of Computer Engineering
International Institute of Information Technology
Pune, India

Kimi Ramteke
Department of Computer Engineering
International Institute of Information Technology
Pune, India

Abstract:- A Personal Assistant is a computer application that uses Artificial Intelligence (AI) to assist humans with their tasks. In today's world, the use of a voice personal assistant is becoming more common. Modern IPAs are capable of a broad range of functions, from simple ones like opening an app or setting an alarm to more complicated ones like taking notes.
Siri from Apple, Google Assistant and Alexa from Amazon are all examples of AI assistants.

Keywords:- *Intelligent Personal Assistant, NLP, ASR, Location of Books, Library Management.*

I. INTRODUCTION

A student's life revolves around the library. It is also beneficial to academic researchers. Libraries may now be found in almost any location. It's a huge endeavor to locate a book at the library. While looking for the needed book, some people lose interest. It takes a long time to look for books on the computer.

With open source software products like Moodle for Virtual Learning gaining traction in related fields, many librarians are looking for OSS alternatives to their present Library Management Systems.

An Intelligent Personal Assistant (IPA) is a computer application that uses Artificial Intelligence (AI) to help people do tasks. The IPA maintains a continuous dialogue with its users while responding to their questions or carrying out measures to fulfill their demands [2]. Modern IPAs are capable of a broad range of functions, from simple ones like opening an app or setting an alarm to more complicated ones like taking notes or making phone calls. Google Assistant from Google, Siri from Apple, and Alexa from Amazon are all instances of IPAs [3].

Although IPAs are not required in terms of communicating exclusively through voice, many current IPAs are pursuing Voice User Interface, which involves engaging with users only through voice, without the necessity of displays or physical interaction [3]. This necessitates the IPA's ability to (A) listen to human speech, (B) comprehend what is being indicated, and (C) conduct an action or respond with their own synthesized voicection. [4].

Natural Language Processing (NLP) is mostly about instructing machines to understand human languages and derive meaning from texts at their most basic level. Text mining, text categorization, text and emotion monitoring, and speech production and identification are just a few of its many applications. This is also why Natural language processing We constructed a working voice assistant for the purpose of this article, which can do simple tasks such as locating a book in the library. It can interpret "audio instructions" and obtain information from a database.

II. LITERATURE SURVEY

In comparison to previous assistants, he implemented a lot of stuff. It is quite beneficial in human life nowadays. It's a really straightforward application. It is also used in the corporate world, for example, in laboratories where people wear gloves and bodysuits for safety reasons, making it impossible to write. However, with a voice assistant, they can access whatever information they need, making their work easier. Only the most fundamental elements have been implemented in this study. For example, a Google term search, a YouTube song/video, a location search, and current news. There is a lot more that can be done. [1]

The focus of the research is on the technique used to create a multilingual and adjustable voice recognition and speech synthesis system. There can be no assumptions about the language identification of vocabulary items in voice calling, although the phone book can contain names in many languages. Each voice tag must be trained separately by the user. This keeps the amount of voice tags to a bare minimum. few, and the person has trouble recalling the exact statement that was uttered during the course of the training. [3]

With the desire for human-machine interactions, current voice recognition applications are becoming increasingly widespread. On traditional general-purpose computers, several speech-based interactive software programmes were run. It is

ISSN No:-2456-2165

speaker-dependent and only useful to a select group of people for whom the system has been trained. This type of technology provides good performance for a known user, but performance drops exponentially for new users. [11]

The design aspects of a speech recognition engine configured for mobile devices are outlined in this study. New features and crucial enhancements of the original concepts, which were ignored, are now discussed, despite the fact that main elements have previously been provided. We also demonstrate how these strategies may be successfully coupled to fulfill a variety of design goals with minimal influence on recognition performance. Speaker adaptation is a required process for optimizing recognition performance.

This set is utilized to compute the replacements for the original parameters once enough adaptation utterances have been observed.

However, doing so will significantly increase the amount of memory required for storing acoustic models. [12]

This article discusses recent advancements in voice recognition technology as well as the author's opinions on the subject. Dictation and human-computer conversation systems are the two primary domains where voice recognition technology is used. Automatic broadcast news transcription is now being researched in the dictation area, particularly under the DARPA project. The most pressing issue is determining how to make speech recognition systems resistant to acoustic and linguistic diversity in speech. In this setting, a paradigm change from speech recognition to understanding where the speaker's underlying messages, that is, the meaning/context that the speaker intended to express, are extracted rather than all the spoken words, will be critical. [13]

By giving a short survey on Fully automated Speech Recognition and discussing the major themes and progress made in the past 60 years of research, this paper provides a rapid technology perspective and a gratitude of the profound work that has been done in this important area of speech communication. Speech recognition has piqued the interest of scientists as a vital field. It has had a technical influence on society, and is projected to grow in this field of human-machine interaction in the future. [14]

This article provides an overview of the various types of voice assistants as well as the functions that voice assistants may perform. Before voice assistants can be used for anything that requires secrecy, privacy and security safeguards will need to be strengthened. [15]

A collection of 17 criteria for voice-based devices is described in this research. The heuristics cover a large portion of the design space and may also be used to generate early concepts. Error handling should receive more attention.

To make the system status clearer, these gadgets should deliver more effective multimodal feedback. [16]

The interactions of well known voice assistants with 26 children are described in this study. These kids were able to converse and play with the assistance.

Later on, the kids provide comments about the assistant. Overall, all of the youngsters were enthusiastic helpers. We found that the vast majority of participants had difficulty getting the agents to grasp their inquiries. Several students attempted to raise their voice volume or add additional pauses to their remarks in the hopes of making their questions more understandable. However, it does not always result in greater agent recognition. [17]

III. PROPOSED WORK

We employ speech recognition to discover books in our project, which makes this mammoth process a lot simpler. Even those with no prior understanding of the system can use it to discover books more quickly. It cuts down on time spent looking for books and makes the library more user-friendly.

The system will connect with its users in real time as it responds to their questions. Inquiries include book locations based on a unique identifier or the title of the book.

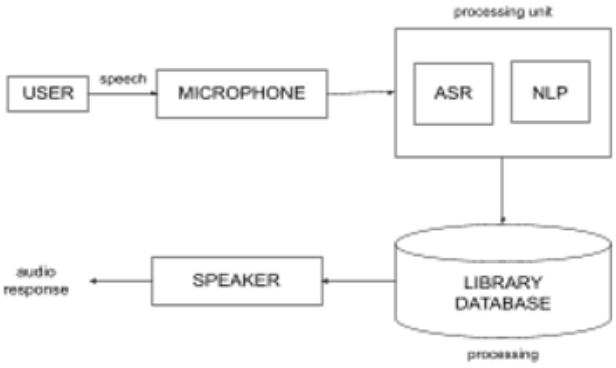

Fig. 1. Architecture Diagram.

The steps taken will be as follows :

Step 1 : User Speaks

User speaks into the microphone and then these signals are then sent to the processor.

Step 2 : Audio Processing (ASR and NLP)

Here, Automatic Speech Recognition (ASR) is used which takes an audio file and converts it into a sequence of words. The NLP module picks up keywords and identifies the book and its location in the database.

Step 3 : Relevant information gathered from the database

Information is retrieved from the database (book and its location).

Step 4 : Response/Error is received

The response is received through the speaker.

IV. MODULES

A. Automatic Speech Recognition (ASR) module

Speech recognition is defined as the automatic recognition of human speech and is recognized as one of the most important tasks when it comes to making applications based on Voice User Interfacing (VUI).

Python comes with several libraries which support speech recognition. We will be using the speech recognition library because it is the simplest and easiest to learn.

This module involves 2 steps -

- *Loading Audio*

To load audio the AudioFile function is used. The function opens the file, reads its contents and stores all the information in an AudioFile instance called source.

We will traverse through the source and do the following things:

- Every audio has some noise involved which can be removed using the adjust_for_ambient_noise function.
- Making use of the record method which reads the audio file and stores certain information into a variable to be read later on.

- *Reading data from audio*

We can now invoke recognize_google() method and recognize any speech in the audio. After processing the method returns the best possible speech that the program was able to recognize from the first 100 seconds. The output comes out to be a bunch of sentences from the audio which turn out to be pretty good. The accuracy can be increased by the use of more functions but for now it does the basic functionalities.

B. Natural Language Processing (NLP) module

SpaCy is a Natural Language Processing package for Python that is open-source. It's primarily intended for use in the workplace, where it may create real-world initiatives and manage massive amounts of text data. Moreover, sixty languages are supported by spaCy, which has prepared pipelines for various languages and activities. Because this toolset is designed in Python and Cython, it is quicker and more accurate when dealing with massive amounts of text data.

C. Subprocess

Utilized to obtain information on system subprocesses, which are then used in different commands like shutdown, sleep, and so on. Python includes this module by default.

D. WolframAlpha

Wolfram's algorithms, knowledgebase, and AI technologies are used to compute expert-level responses.

E. Pyttsx3

In a software, this is used to convert text to speech.

F. Datetime

Used to know date and time. It comes built in with Python.

G. Speech Recognition

Used to make the assistant understand your voice.

V. USES AND APPLICATION

This voice based assistant will be very useful for a student to search a book in the library.Voice based assistant will communicate with the user and will give the exact location of the book at the library. This assistant can be built in various colleges to reduce the work performed by the librarian.

VI. SCOPE

In the near future, we can implement a system to keep the records of book transactions. For example, the date of a book borrowed and returned can be recorded. The voice assistance can be made multilingual so that the user can give input in any of the languages. We can also implement a student and faculty in- out system which will keep the record of entry and exit time in the library.

VII. CONCLUSION

The method described is appropriate for developing an Intelligent Voice Assistant that can interpret domain-specific direct voice commands from users. We will be using this approach to build an automated library management system based on an intelligent voice assistant.

REFERENCES

[1]. Subhash S, Prajwal N Srivatsa, Siddesh S, Ullas A, Santhosh B, "Artificial Intelligence- based Voice Assistant" 2020, IEEE.
[2]. J. Kurpansky, "What Is an Intelligent Digital Assistant?" Medium, 2017.
[3]. J. Iso-Sipila, M. Moberg and O. Viikki, "Multi-lingual speaker independent voice user interface for mobile devices," in ICASSP 2006. IEEE, 2006, vol. I, pp. 1081-1084.
[4]. "Intelligent Personal Assistants. In: Speech Technology" S. Springenberg, Seminar, Institute of Computer Science Hamburg, 2016, pp. 5-7
[5]. Saadman Shahid Chowdhury,Atiar Talukdar, Ashik Mahmud, Tanzilur Rahman, "Domain Specific Intelligent Personal Assistant with Bilingual Voice Command Processing" TENCON 2018 - 2018 IEEE Region 10 Conference

[6]. Ishita Garg Hritik Solanki Sushma Verma, "Automation and Presentation of Word Document Using Speech Recognition" 2020 International Conference for Emerging Technology (INCET) [7]Dinesh K, Prakash R and Madhan Mohan P, "Real-time Multi Source Speech Enhancement for Voice Personal Assistant by using Linear Array Microphone based on Spatial Signal Processing" 2019 International Conference on Communication and Signal Processing (ICCSP)
[7]. Alisha Pradhan and Amanda Lazar, University of Maryland Leah Findlater, University of Washington, "Use of Intelligent Voice Assistants by Older Adults With Low Technology Used" ACM Transactions on Computer-Human Interaction
[8]. George Terzopoulos Maya Satratzemi, "Voice Assistant and Artificial Intelligence in Education"
[9]. The 9th Balkan Conference Juha Iso-Sipilä, Marko Moberg, Olli Viikki, "Multilingual Speaker-Independent Voice User Interface for Mobile Devices" Nokia Technology Platforms, Tampere, Finland
[10]. D Satya Ganesh, Dr. Prasant Kumar Sahu School of Electrical Sciences Indian Institute of Technology Bhubaneswar, Odisha, India, "A Study on Automatic Speech Recognition Toolkits" 2015 IEEE at ICMOCE
[11]. Marcel Vasilache, Juha Iso-Sipilä, Olli Viikki, "On a Practical Design of a Low Complexity Speech Recognition Engine" 2014 IEEE at ICASSP
[12]. Sadaoki Furui Department of Computer Science Tokyo institute of Technology, "Automatic Speech Recognition and its Application to Information Extraction" 2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8552 Japan
[13]. M.A.Anusuya, S.K.Katti Department of Computer Science and Engineering Sri Jayachamarajendra College of Engineering, Mysore, India, "Speech Recognition by Machine : A Review" IJCSIS Vol. 6, No. 3, 2009
[14]. Matthew B. Hoy, "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants" Medical Reference Services Quarterly, 37:1, 81-88, DOI: 10.1080/02763869.2018.1404391
[15]. Zhuxiaona Wei, James A. Landay, "Evaluating Speech-Based Smart Devices Using New Usability Heuristics" Published by the IEEE Computer Society April–June 2018
[16]. Stefania Druga, Cynthia Breazeal, Randi Williams, Mitchel Resnick, "'Hey Google is it OK if I eat you?' Initial Explorations in Child-Agent Interaction" IDC 2017, June 27–30, 2017, Stanford, CA, USA
[17]. Polyakov E.V., Mazhanov M.S., Rolich A.Y., Voskov L.S., Kachalova M.V., Polyakov S.V., "Investigation and Development of the Intelligent Voice Assistant for the Internet of Things Using Machine Learning" 2018 IEEE
[18]. Thimmaraja Yadava, Pravenn Kumar, H S Jayanna, "An End to End Spoken Dialogue System to Access the Agricultural Commodity Price Information in Kannada Language/Dialects" [20]Ekenta Elizabeth Odokuma, Orluchukwu Great Ndidi, "Development of a Voice-Controlled Personal Assistant for the Elderly and Disabled" IRJCS, 2016 [21]Vassil Vassilev, Anthony Phipps, Matthew Lane, Khalid Mohamed, Artur Naciscionis "Two-factor authentication for voice assistance in digital banking using public cloud services" April, 2020
[19]. Shinya Iizuka, Kosuke Tsujino, Shin Oguri, Hirotaka Furukawa, "Speech Recognition Technology and Applications for Improving Terminal Functionality and Service Usability" 2012 NTT DOCOMO Technical Journal Vol. 13 No. 4
[20]. Alisha Pradhan1, Kanika Mehta1, Leah Findlater2, "Accessibility came by accident:Use of voice controlled intelligent personal assistants by people with disabilities" April 21–26, 2018, Montréal, QC, Canada
[21]. Hasrat Jahan, "Reshaping learning through Artificial Intelligence" Department of Education, Bhopal Degree College, 393, Ashok Vihar, Bhopal - 462 023 (M.P.) India.