STAGE II PROJECT REPORT

ON

Leah: A NLP Powered Voice Assistant Streamlining your Daily Tasks

SUBMITTED FOR THE EVALUATION OF STAGE II REPORT

BY

Mr. Aditya Yadav	**20190802060**
Mr. Raj Kansal	**20190802013**
Mr. Sidhant Subudhi	**20190802035**

Under The Guidance Of
Ms. Nikita Pagar & Dr. Bahubali Shiragapur

D Y PATIL
INTERNATIONAL
UNIVERSITY
AKURDI PUNE

School Of Computer Science Engineering & Applications

D Y Patil International University, Akurdi, Pune - 411044

ABSTRACT

This project aims to develop a voice assistant using Natural Language Processing (NLP) techniques. The voice assistant will be able to understand spoken language and respond with appropriate actions. It will be designed to recognize and interpret user commands, perform specific tasks, and provide relevant information. The project will involve building a custom dataset for training and evaluating the NLP models, implementing voice recognition and synthesis technologies, and integrating the assistant with various applications and services. As of now we have used NLP libraries like librosa and sounddevice to get started with and binary classification as our supervised machine learning model. The end goal is to create a user-friendly and efficient voice assistant that can enhance productivity and convenience in daily life.

Keywords: Voice assistant, Natural Language Processing (NLP), Voice Recognition, Supervised Machine Learning.

INDEX

Introduction

A voice assistant is an artificial intelligence (AI) software that utilizes natural language processing (NLP) to perform various tasks through voice commands. These assistants have become increasingly popular in recent years, with companies such as Amazon, Google, and Apple all releasing their versions of voice assistants.

NLP is a subfield of artificial intelligence that focuses on the interaction between computers and humans through natural language. Voice assistants that utilize NLP technology can understand and interpret spoken commands from users and provide relevant responses. This technology has made voice assistants more accessible and user-friendly, as users no longer need to rely on complicated commands or interfaces.

In a study by Wu et al. (2019), researchers examined the accuracy of various voice assistants, including Google Assistant, Amazon Alexa, and Apple Siri. The study found that these assistants had an average accuracy rate of 82%, with Google Assistant performing the best. However, the study also found that voice assistants were less accurate in noisy environments and when recognizing non-native speakers.

Another study by Zin et al. (2020) focused on the use of voice assistants in healthcare. The study found that voice assistants could be a valuable

tool in providing medical advice and answering questions related to health. However, the study also highlighted the need for voice assistants to be accurate and reliable in providing medical information.

Privacy and security are essential considerations when it comes to voice assistants. As voice assistants are always listening, there is a risk that they could capture and transmit sensitive information to third parties. In a study by Acar et al. (2018), researchers examined the privacy and security implications of voice assistants. The study found that many voice assistants were vulnerable to attacks such as eavesdropping and phishing.

In response to these concerns, companies have implemented various security features in their voice assistants. For example, Google Assistant now allows users to delete their voice data and has added a feature that requires users to authenticate themselves before making purchases. However, there is still a need for ongoing research and development to ensure that voice assistants are secure and protect user privacy.

User acceptance and satisfaction are crucial factors when it comes to the success of voice assistants. In a study by Jang et al. (2020), researchers examined the factors that influence user satisfaction with voice assistants. The study found that ease of use, accuracy, and perceived usefulness were significant predictors of user satisfaction. The study also found that users were more satisfied with voice assistants that had a more conversational tone.

In another study, Wang et al. (2020) examined the effect of personality on user acceptance of voice assistants. The study found that users were more likely to accept a voice assistant that had a personality similar to their

own. The study also found that users were more likely to continue using a voice assistant that had a more conversational tone and was more responsive to their needs.

Voice assistants that utilize NLP technology have become increasingly popular in recent years. They offer users a more accessible and user-friendly way to interact with technology. However, there are still challenges that need to be addressed, such as privacy and security concerns and user acceptance and satisfaction. Ongoing research and development are essential to ensure that voice assistants continue to improve and provide users with a valuable and engaging experience.

Literature Survey

Giovanni Campagna, Silei Xu, Mehrad Moradshahi, Richard Socher, Monica S. Lam. Genie: A Generator of Natural Language Semantic Parsers for Virtual Assistant Commands. A methodology and toolkit to reduce labor-intensive manual annotation of sentences for virtual assistants.Virtual Assistant Programming Language (VAPL) to formalize capability of virtual assistants. Neural semantic parser to translate natural language into VAPL code. Design principles to make VAPL languages amenable to natural language translation ThingTalk revised with design principles 62% accuracy on realistic user inputs 19% and 31% improvement over previous state of the art on music skill, aggregate functions and access control.

Jordan J. Bird, Anikó Ekárt, Diego R. Faria. Chatbot Interaction with Artificial Intelligence: Human Data Augmentation with T5 and Language Transformer Ensemble for Text Classification. Presents CI-AI framework as approach to training deep learning chatbots for task classification. Augments human-sourced data via artificial paraphrasing to generate a large set of training data. 483 responses recorded and split into training and validation sets. Seven transformer-based text classification algorithms benchmarked for both sets. All models

improved when training data augmented by T5 model (average increase of 4.01%)

RoBERTa model trained on T5 augmented data achieved 98.96% classification accuracy. Ensemble of five best-performing transformer models via Logistic Regression achieved 99.59% accuracy

Allows intelligent system to interpret human commands at social-interaction level via chatbot-like interface

Project Plan:

Schedule		Date	Project Activity
February	1st week	01/02/2023	Project Topic Selection
	2nd week	06/02/2023	Synopsis Submission
	3rd week	13/02/2023	Presentation on project ideas
	4th week	20/02/2023	Submission of Literature Survey & Feasibility
March	1st week	01/03/2023	Improvising the wake word functionality
	2nd week	06/03/2023	Study of various research papers
	3rd week	13/03/2023	Study and creation of the Intent Engine
	4th week	20/03/2023	Making different skill modules for our voice assistant
	5th week	27/03/2023	Integration of the modules with the program
April	1st week	03/04/2023	Documentation and creation of report along with diagrams
	2nd week	10/04/2023	Discussion on 2nd stage project report

	3rd week	**17/04/2023**	**Stage 2 Project report presentation, submission and evaluation**
	4th week	**24/04/2023**	
May	**1st week**	**01/05/2023**	
	2nd week	**08/05/2023**	
	3rd week	**15/05/2023**	

Chapter 1
Objectives, Problem statement, Feasibility, Methodology

1.1 Objectives

1. Develop a natural language processing (NLP) model that can accurately comprehend user requests and commands.

2. Implement a speech recognition system that can convert spoken language into text, which can be processed by the NLP model.

3. Integrate the voice assistant with various applications and services, such as calendars, emails, and weather services.

1.2 Problem Statement

The problem that the voice assistant project using NLP aims to address is the need for a more efficient and natural way for users to interact with technology. Current methods, such as typing on a keyboard or clicking on a mouse, can be time-consuming and cumbersome, especially for individuals with disabilities or those who prefer a more hands-free approach. By leveraging NLP techniques to develop a voice assistant, users can communicate with technology in a more intuitive and convenient way, using natural language commands and receiving spoken responses. Therefore, this project seeks to overcome these obstacles to create a robust and user-friendly voice assistant that can improve productivity and enhance the overall user experience.

1.3 Feasibility

Feasibility for a voice assistant project using NLP (Natural Language Processing) would depend on various factors, including:

1. Technical Feasibility: Many open-source libraries and frameworks exist to facilitate the development of such systems, and there is a wealth of research and expertise in the field.

2. Financial Feasibility: The costs associated with the project are primarily related to the hardware and infrastructure required for development and testing. However, most of the required software tools and technologies are free and open-source, which can help reduce the overall project costs.

3. Operational Feasibility: The voice assistant is user-friendly and easy to operate. Can it be integrated with the existing systems and workflows.

1.4 Methodology

1. Identify the Requirements: Define the scope of the project and gather the requirements from stakeholders. Determine the purpose, features, and functionalities of the voice assistant.

2. Data Collection: Collect and preprocess data to train the NLP model. The data can be collected from various sources, such as text documents, chat logs, and voice recordings.

3. Model Selection: Choose the appropriate NLP model based on the requirements and the available data. Common models used in voice assistants include Intent Recognition, Named Entity Recognition, and Sentiment Analysis.

4. Testing: Test the voice assistant to ensure that it meets the

requirements and functions as intended. Conduct unit testing, integration testing, and user acceptance testing.

5. Maintenance: Maintain the voice assistant by monitoring its performance, fixing bugs, and updating the NLP model and features as required.

6. User Feedback: Collect feedback from users to improve the voice assistant. Analyse user behaviour and interactions to identify areas for improvement.

By following these steps, the voice assistant project using NLP can be developed, tested, and deployed successfully.

Chapter 2
Diagrams

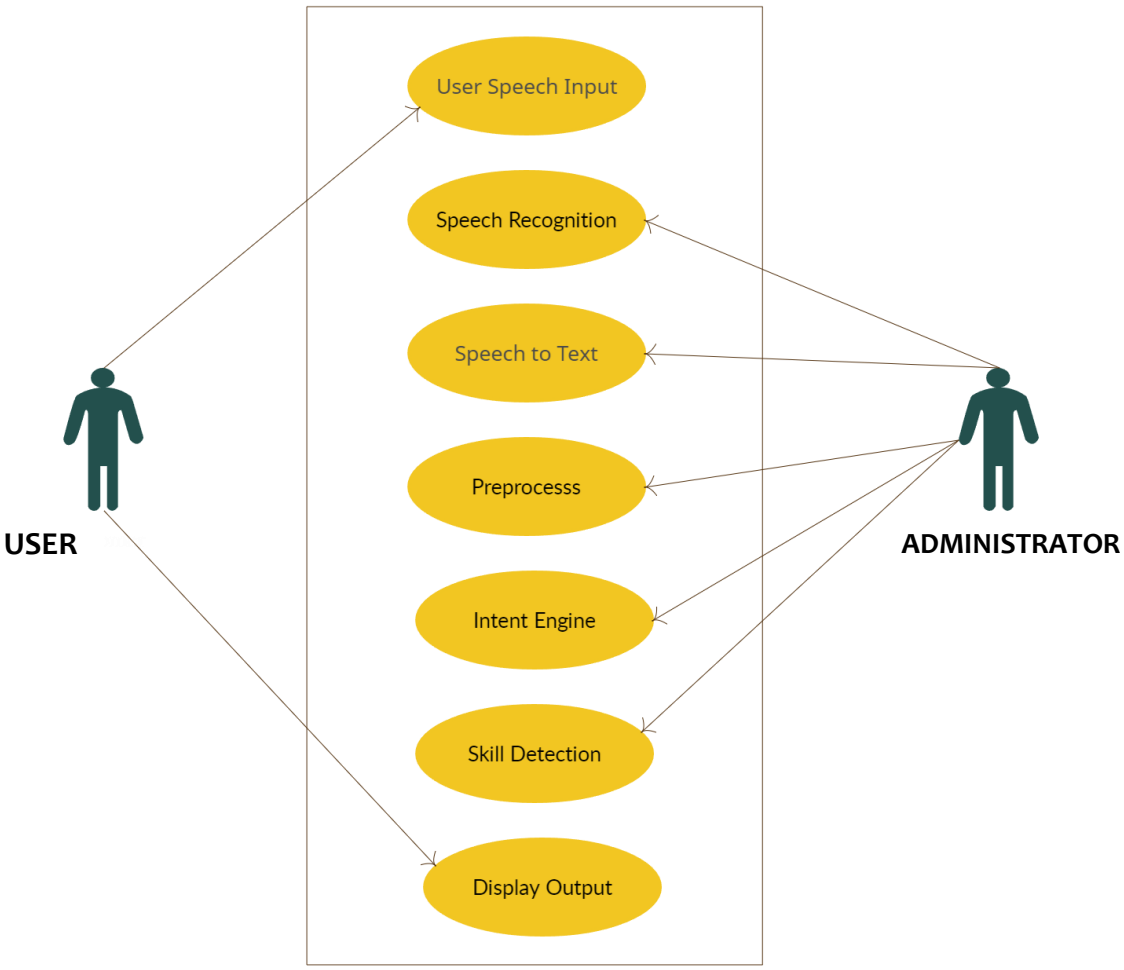
1.5 UML Diagram

A UML diagram is a diagram based on the UML (Unified Modeling Language) with the purpose of visually representing a system along with its main actors, roles, actions, artifacts or classes, in order to better understand, alter, maintain, or document information about the system.

1.5.1 Use-case Diagram

A use case diagram is a type of UML (Unified Modeling Language) diagram that represents the behavior of a system from the point of view of its users. It provides a visual representation of the interactions between the system and its users or actors.

In a use case diagram, use cases are represented as ovals, while the actors are represented as stick figures. The use cases are connected to the actors using lines or arrows that represent the flow of information or actions between them.



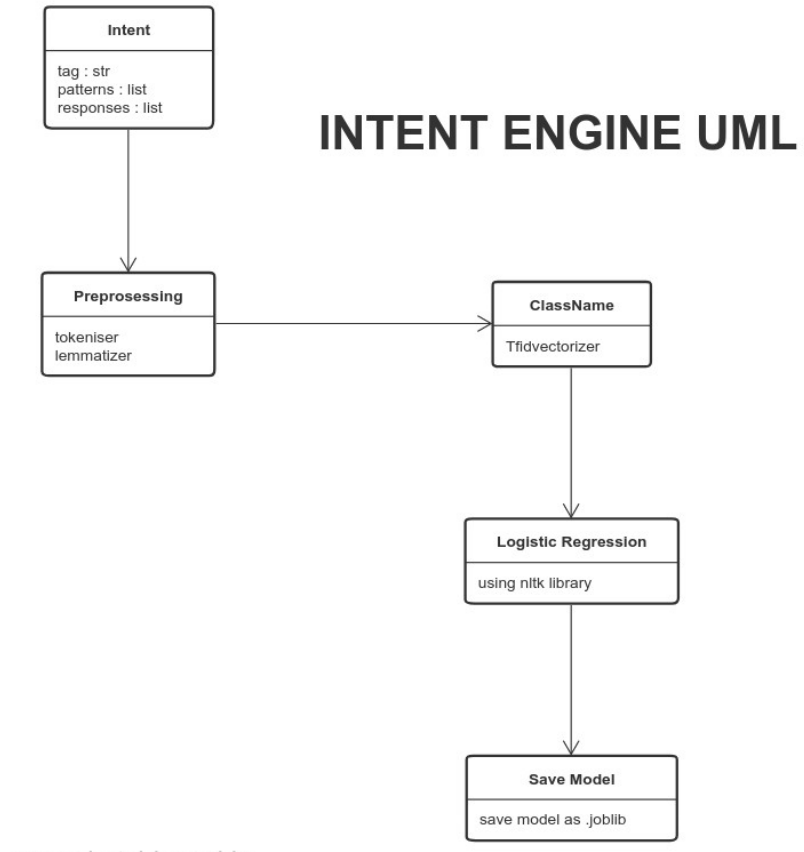
1.5.2 Class Diagram

A class diagram is a type of UML (Unified Modeling Language) diagram that represents the static structure of a system by showing the classes, their attributes, operations, and relationships among them.

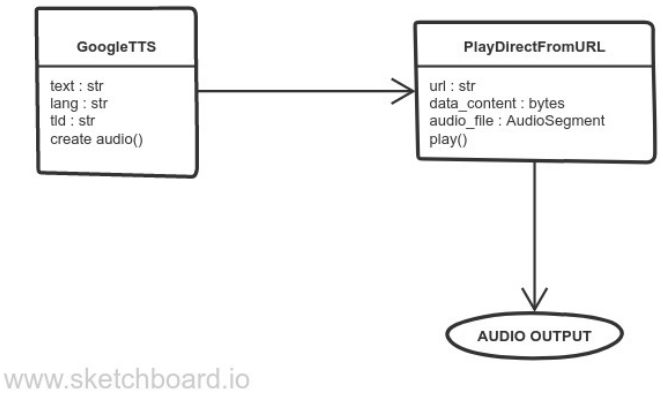
In a class diagram, classes are represented as boxes with the class name written inside. The class attributes (i.e., variables) are shown as smaller boxes within the class box, while the class operations (i.e., methods) are shown as a list under the class name.

Relationships between classes are shown using lines or arrows between the classes. For example, an inheritance relationship is shown as a line with a hollow arrowhead pointing from the subclass to the superclass. An association relationship is shown as a line with an arrowhead pointing from one class to another.

Class diagrams are used in software development to help developers understand the structure of the system being developed and to communicate the system design to stakeholders. They are particularly useful for planning and designing object-oriented systems.



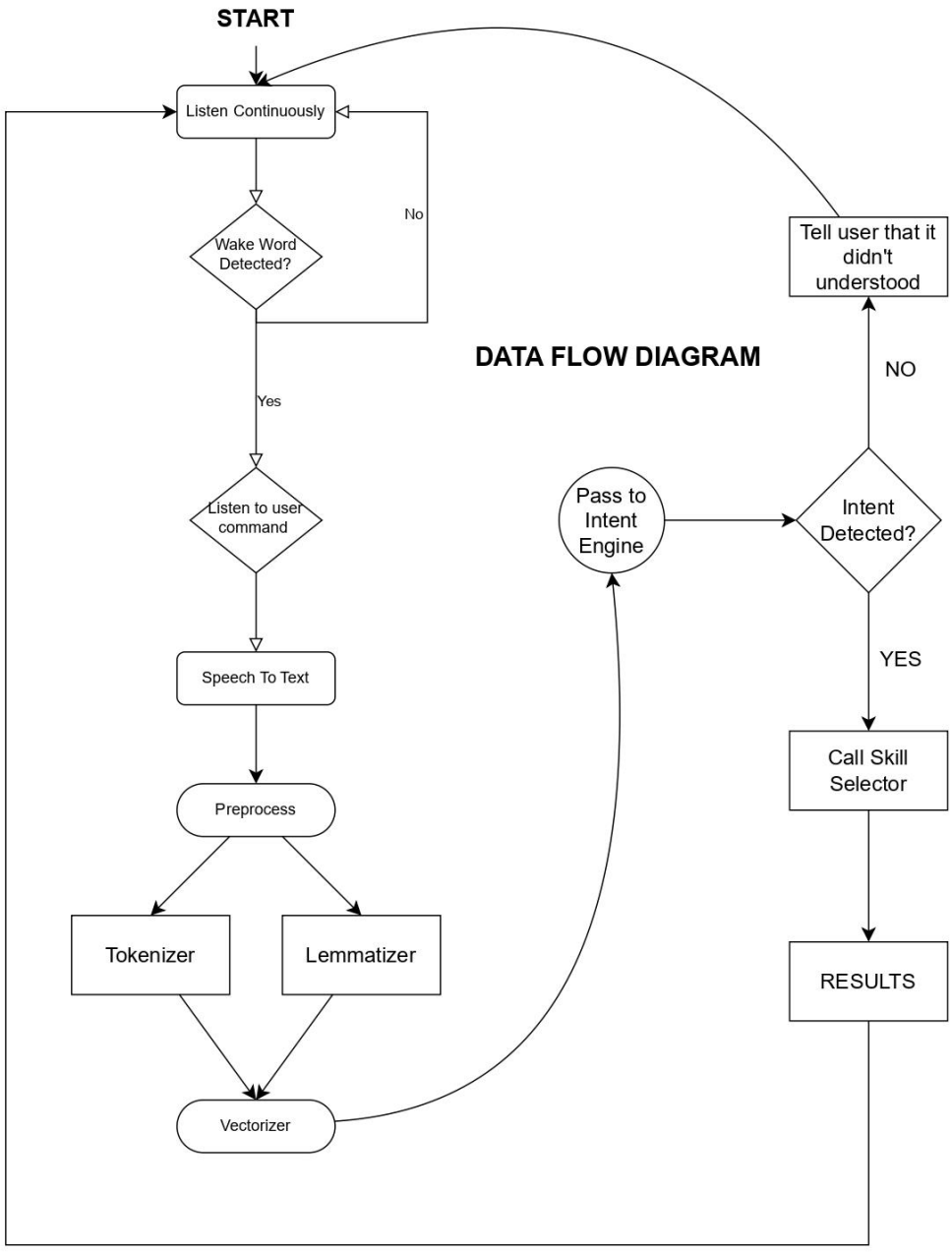
TTS Class Diagram



1.6 Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of how data flows through a system, modeling its inputs, outputs, and processes. It is a type of structured analysis and design tool that is used to describe and analyze the flow of data within an organization or system.

A DFD consists of a set of interconnected data flow diagrams that show the flow of data through different processes within the system. Each process in the diagram represents a specific activity or operation that transforms the input data into output data. The input and output data are represented as data flows, and they are connected to the processes using arrows.


START
Listen Continuously
Wake Word Detected?
No
Yes
Listen to user command
Speech To Text
Preprocess
Tokenizer
Lemmatizer
Vectorizer
DATA FLOW DIAGRAM
Pass to Intent Engine
Intent Detected?
NO
YES
Tell user that it didn't understood
Call Skill Selector
RESULTS