A

Project Report

On

VOICE ASSISTANT FOR WEB

Submitted to

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE AND TECHNOLOGIES
R K VALLEY

In partial fulfillment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

In

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

Submitted by

SAKE BHARATH SAI(R180856)
GOLLA SAI RAM(R180918)

Under the Guidance of

Mrs. S Rajeswari, Guest Faculty

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
RAJIV GANDHI UNIVERSITY OF KNOWLEDGE AND TECHNOLOGIES
R K VALLEY
(catering the Educational Needs of Gifted Rural Youth of AP)
R.K Valley,Vempalli(M),Kadapa(Dist)—516330
2023 - 2024

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES
(A.P.Government Act 18 of 2008) RGUKT-RK Valley
Vempalli,Kadapa,Andhrapradesh – 516330.

CERTIFICATE OF PROJECT COMPLETION

This is to certify that I have examined the thesis entitled “**Voice assistant for web**” submitted by **Sake Bharath Sai (R180856), Golla Sai Ram (R180918)** under our guidance and supervision for the partial fulfillment for the degree of Bachelor of technology in computer science and Engineering during The academic session 2023-2024 at RGUKT-RK VALLEY.

Signature of Internal Guide
Mrs. S. Rajeswari,
Project Internal Guide,
Asst.Prof.in Dept of CSE,
RGUKT-RKValley.

Signature of HOD
Dr. P. Ravi Kumar,
Head of the Department ,
P.hD (University of AIZU, Japan),
M.E (IISc Bangalore),
RGUKT RK VALLEY.

Signature of External Examiner

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES
(A.P.Government Act 18 of 2008) RGUKT-RK Valley
Vempalli,Kadapa, Andhrapradesh- 516330.

DECLARATION

We hereby declare that the project report entitled “**Voice assistant for web**” done under guidance **of Mrs. S. Rajeswari** is submitted in partial fulfillment for the degree of Bachelor of Technology in Computer Science and Engineering during the academic session December 2023 – April 2024 at RGUKT-RKValley. I also declare that this project is a result of our own effort and has not been copied or imitated from any source. Citations from any websites are mentioned in the references. To the best of my knowledge, the results embodied in this dissertation work have not been submitted to any university or institute for the award of any degree or diploma.

With Sincere Regards,

Date: Sake Bharath SaiR180856)
Place:RK Valley Golla Sai Ram(R180918)

ACKNOWLEDGEMENT

I would like to express my deep sense of gratitude & respect to all those people behind the screen who guided, inspired and helped me crown all my efforts with success. I wish to express my gratitude to **Mrs. S. Rajeswari** for her valuable guidance at all stages of study, advice, constructive suggestions, supportive attitude and continuous encouragement, without which it would not be possible to complete this project. I would also like to extend our deepest gratitude & reverence to the **Director of RGUKT, RK Valley Prof.AVSS.Kumara Swami Gupta and HOD of Computer Science and Engineering Dr.P.Ravi Kumar** for their constant support and encouragement.

With Sincere Regards,

Sake Bharath Sai(R180856)
Golla Sai Ram(R180918)

INDEX

Page no

Abstract

Speech recognition technology is one from the fast growing engineering technologies. It has a number of applications in different areas and provide potebtial benefits. Nearly 20% people of the world are suffering from various diabilities many of them are blind or unable to use their hands effectively .The speech recognition system in those particular cases provide a significant help to them so that they can share with people operating computer through voice input.

This project is desinged and developed keeping that factor into mind, and a little effort is made to achive this aim. Our project is capable to recognize the speech and convert the input audio into text. We include this feature in website, which is used to navigate to the other web pages or to the particular section in the page by giving the voice command.

At the initial level effort is made to provide help for basic operations as discussed above, but the software can further be updated and enhanced in order to cover more operations.

1. Introduction

Overview of speech recognition:

Speech recognition is a technology that able to capture the words spoken by a human with the help of microphone.Those words are later on recognized by speech recognizer, and in the end system outputs the recognized words. The process of speech recognition consists of different steps that will be discussed in the following sections one by one.

An ideal situations in the process of speech recognition is that ,a speech recognition engine recognizes all words uttered by a human but,practically the performance of the speech recognition engine depends on number of factors. Vocabularies, multiple users and noisy environment are the major factors that are the major factors that are counted in as the depending factors for a speech recognition engine.

The speech recognition started at somewhere in 1940 ,practically the first speech recognition programwas appeared in 1952 at the bell labs that was about recognition of a digit in a noice free environment.

The key invention of this era were Deep Neural Network (DNN) and the stochastic language model.

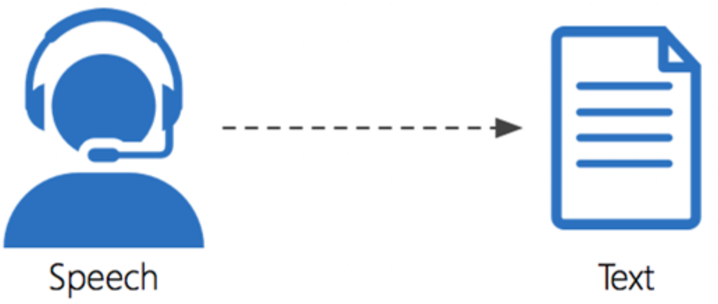
There are two modules in this model:

1) Speech recognition and converting into text

2) Navigating to the required output

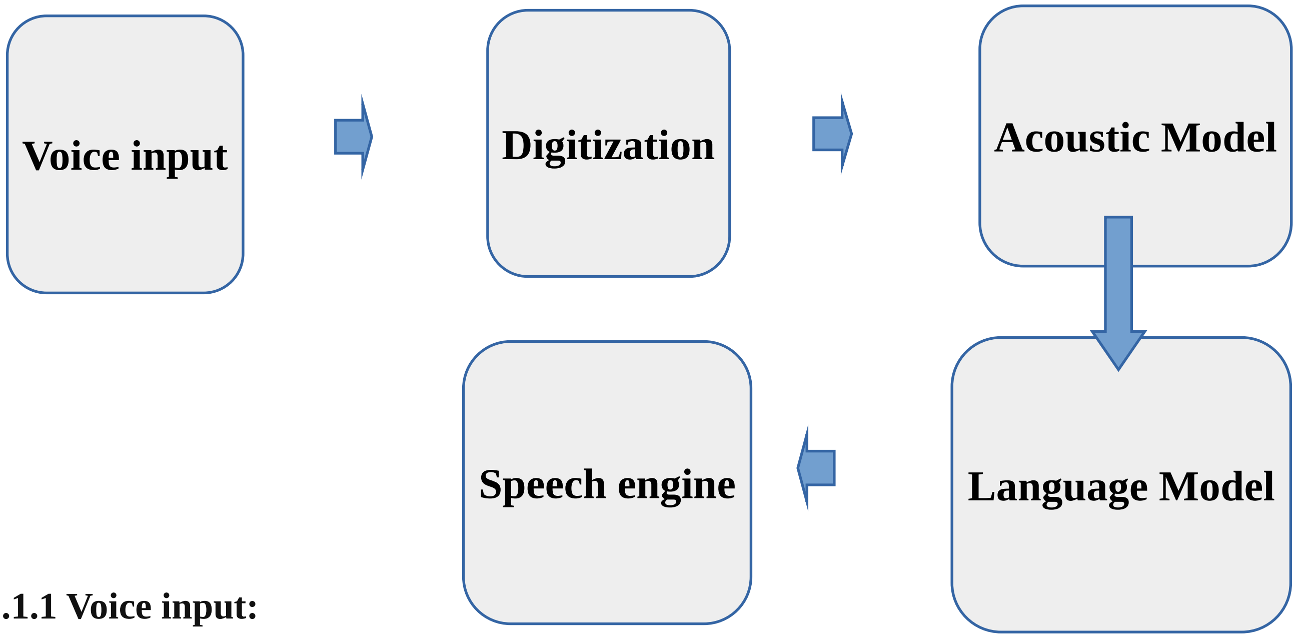
2. Speech Recognition model:

In the history of modern technology, the ability to convert spoken words into text is freely available to everyone who wants to experiment with it. **When it comes to creating speech-to-text applications, Python, one of the most widely used programming languages, has plenty of options.**



2.1 Components of speech recognition system:

- Voice input
- Digitization
- Acoustic model
- Language model
- Speech engine



2.1.1 Voice input:

With the help of microphone audio is the input to the system, the pc sound card produces equivalent digital representation of received audio.

<u>System Requirements:</u>

- High quality microphones

- Sound cards
- Computer/Processor

2.1.2 Microphones:

A quality microphone is key when utilizing the speech recognition system.Desktop microphones are not suitable to continue with speech recognition system, because they have tendency to pick up more ambient noise. The best chioce, and most common is the headset style. It allows the ambiest noise to be minimized, while allowing you to have the microphone at the tip of your toungue the time. Headsets are available without earphones and with earphones.

2.2 Digitization:

The process of converting analog to digital form is known as digitzation, it involves the both sampling and quantization procss. Sampling is convertiong a continuous signal into discrete signal, where the process of approximating a continuous range of values is known as quantization.

2.2.1 Analog to digital converter(ADC):

A converter that is used to change the analog signal to digital is known as an analog to digital converter or ADC converter. This converter is one kind of integrated circuit or IC that converts the signal directly from continuous form to discrete form. This converter can be expressed in A/D, ADC, A to D. The inverse function of DAC is nothing but ADC. The analog to digital converter symbol is shown below.

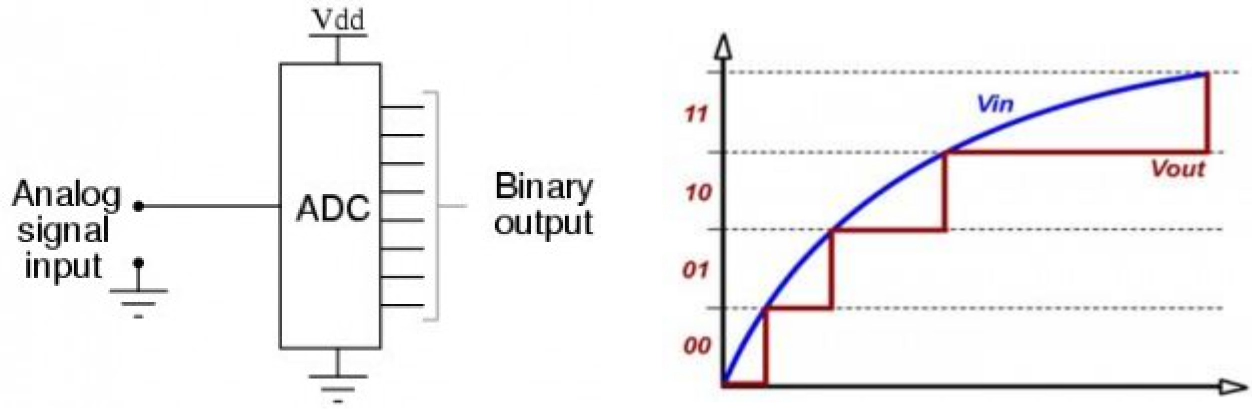


fig : ADC Converter

2.2.2 Sampling:

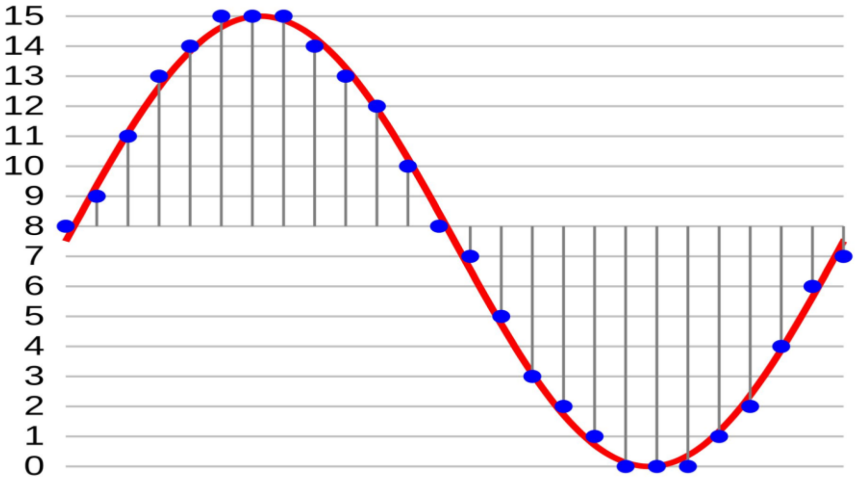
Sampling refers to the process of converting a continuous audio signal into a discrete digital representation. This is a crucial step in digitizing audio for storage manipulation and transmission.

Sampled signal should be represented the original signal faithfully. We should be able to reconstruct the original signal from its sampled version.

The sampling rate determines how may samples are taken per second. Common sampling rate includes 44.1 khz and 48 khz

<u>Nyquist theorem:</u>

The Nyquist theorem states that the sampling rate must be atleast twice the highest frequency presenting the signal.



2.2.3 Quantization:

Quantization is the process of mapping continuous values to a set of discrete values. In the context of digital signal processing, particularly in audio or image processing quantization invloves representing continuous amplitude values with a limited set of digital values.

✓ Determine the range:

Understand the range of values in the continuous signal

✓ Select the bit depth:

Choose a bit depth for your digital representation the bit depth determines the number of bits used to represent each sample .

Ex: n=16 bit

✓ Calculate quantization levels :

Determine the number of quantization levels based on the selected bit depth

The number of levels is given by 2^ n, where n is the bit depth.

✓ Determine the step size:

The step size is the range of values that each quatization level represents it is calcuated by dividing the total range of values by the number of quantization levels.

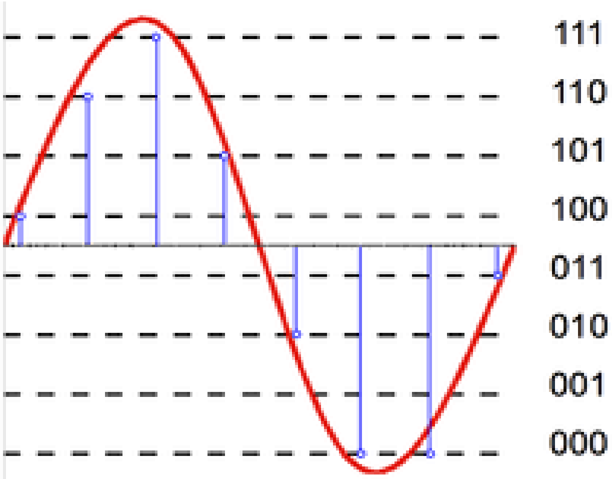
Ex: If you have a range 0 to 1 and 16 quantization levels the step size could be 1/16.

✓ Map continuous values to quatization levels:

For each quantiation levels in the signal find the closest qauntization level. This is offened done by rounding or concatinating the continuous values to the nearest quantization levels.

✓ Represent the Quantized values in binary :

Convert the quantized values into binary from based on the choosen depth this involves representing each quantized values used the specified number of bits.



2.3 Acoustic model:

An acoustic model is created by taking audio recordings of speech and there text transcriptions and using software to create stastical representation of the sounds that make up each word. It is used by a speech recognition engine to recognize speech. The software acoustic model breaks the eords into the phenonmes.

An acoustic model is used in automatic speech recognition to represent the relationship between an audio signal and the phonemes or other linguistic units that make up speech. The model is learned from a set of audio recordings and their corresponding transcripts. It is created by taking audio recordings of speech, and their text transcriptions, and using software to create statistical representations of the sounds that make up each word.

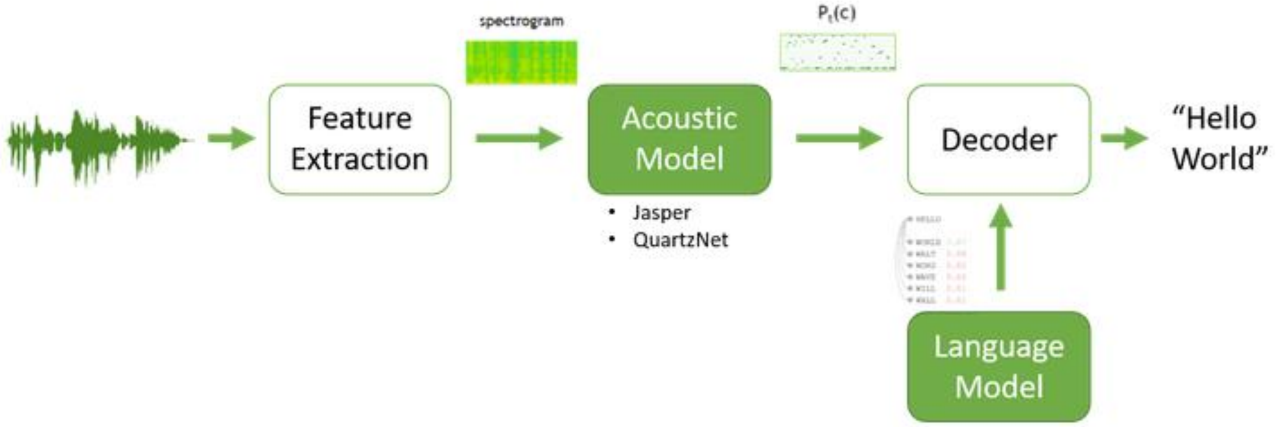


fig : Automatic Speech Recognition(ASR)

2.2.3 Language model:

Language modelling is used in many natural language processing applications such as speech recognitiontries to capture the properties of language and predict the next word in speech sequences.

A language model is a probabilistic model of a natural language. In 1980, the first significant statistical language model was proposed, and during the decade IBM performed 'Shannon-style' experiments, in which potential sources for language modeling improvement were identified by observing and analyzing the performance of human subjects in predicting or correcting text.

Language models are useful for a variety of tasks, including speech recognition, machine translation, natural language, optical character recognition, handwriting recognition, grammar induction, and information retrieval.

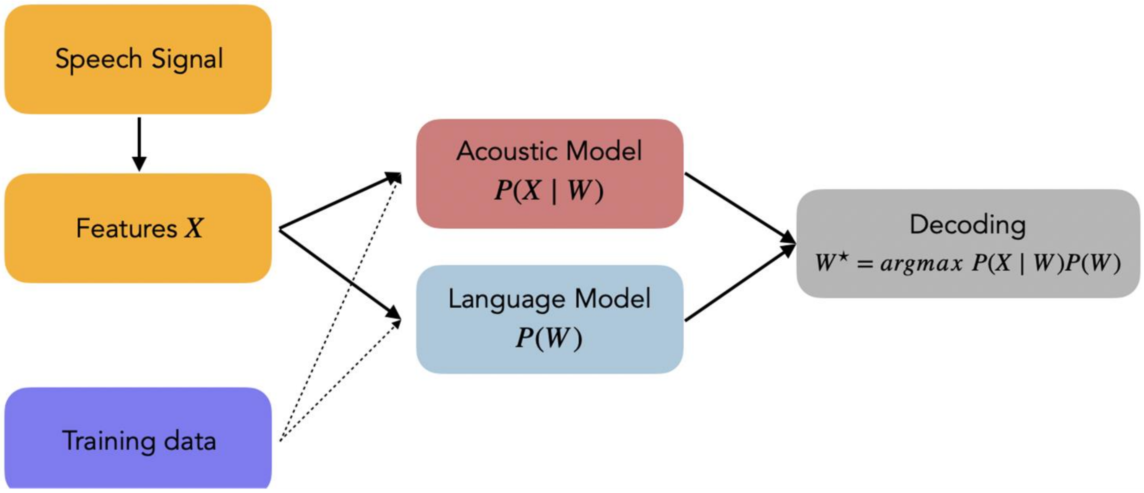


Fig : Process of Language model

2.2.4 Speech engine:

The job of speech engine is to convert audio input into text to accomplish this it uses all sort of data software algorithms and statistics.

4.Preliminaries

Automated Speech Recognition (ASR) has become increasingly sophisticated and accurate as a result of advances in deep learning, cloud computing, and the availability of large training sets (1, 2). The software converts speech into text using artificial intelligence models that have been trained on vast collections of speech containing millions of words. ASR software is widely available on most digital devices, including smartphones, tablets, or laptops. It is primarily used for voice commands (e.g., hey Siri!), at the workplace to create transcripts, or in class for taking notes. Recently, ASR has become available in online meetings (e.g., Microsoft teams) and video recordings (e.g., Google's Youtube) to provide automated captions. Also, several ASR-based speech-to-text apps have been developed for the hearing impaired and deaf, providing live captioning of conversations (2, 3), showing the potential of automation and artificial intelligence for hearing healthcare (4, 5). Early in 2020, we were confronted in our clinic with questions from patients related to the use of ASR apps for daily communication. These questions were especially common among patients with severe to profound hearing loss who visited our outpatient clinic to assess if they were eligible for a Cochlear Implant. Also, patients who had experienced sudden deafness, but had not yet been fitted with hearing aids, made use of an ASR app during their appointments. There was no or little experimental information at the time about the performance and usability of the ASR apps for hearing impaired persons beyond what was shared by developers. Nor did we have clear criteria for which groups of patients we might suggest the ASR apps to.

5. Web Speech API

The speech recognition part of the Web Speech API allows websites to enable speech input within their experiences. Some examples of this include Duolingo, Google Translate, Google.com (for voice search).

What does it do?

When a user visits a speech-enabled website, they will use that site's UI to start the process. It's up to individual sites to determine how voice is integrated in their experience, how it is triggered and how to display recognition results.

As an example, a user might see a microphone button in a text field. When they click it, they will be prompted to grant temporary permission for the browser to access the microphone. Then they can input what they want to say . Once they've finished their utterance, the browser passes the audio to a server, where it is run through a speech recognition engine. The speech recognizer decodes the audio and sends a transcript back down to the browser to display on a page as text.

Contextual Information Retrieval: After transcribing spoken language into text using speech recognition, developers can use web search APIs to retrieve relevant contextual information. For example, if a user asks a question like "What's the weather like in New York?", the speech recognition system can convert the spoken query into text and then use a web speech API to fetch current weather information from a weather website.

Knowledge Expansion: Speech recognition systems can utilize web speech APIs to expand their knowledge base and improve their understanding of natural language queries. For instance, if a user asks a question about a specific topic, the system can use a web speech API to fetch additional information from authoritative sources on the web, enriching its response with more detailed or up-to-date information.

Entity Recognition and Disambiguation: Web speech APIs can help speech recognition systems identify entities mentioned in spoken language and disambiguate them based on context. For example, if a user mentions a person's name or a place, the system can use a web speech API to retrieve information about that entity, such as its biography or location details, helping to provide more accurate and relevant responses.

Content Summarization: Speech recognition systems can leverage web speech APIs to summarize content retrieved from web pages or documents. For instance, if a user requests a summary of a long article or document, the system can use a web speech API to fetch the content and then apply text summarization techniques to generate a concise summary that can be read back to the user.

Fact Verification: Web speech APIs can assist speech recognition systems in fact-checking and verifying information mentioned in spoken language. By querying reputable sources on the web, the system can validate the accuracy of statements made by users and provide trustworthy responses based on reliable information.

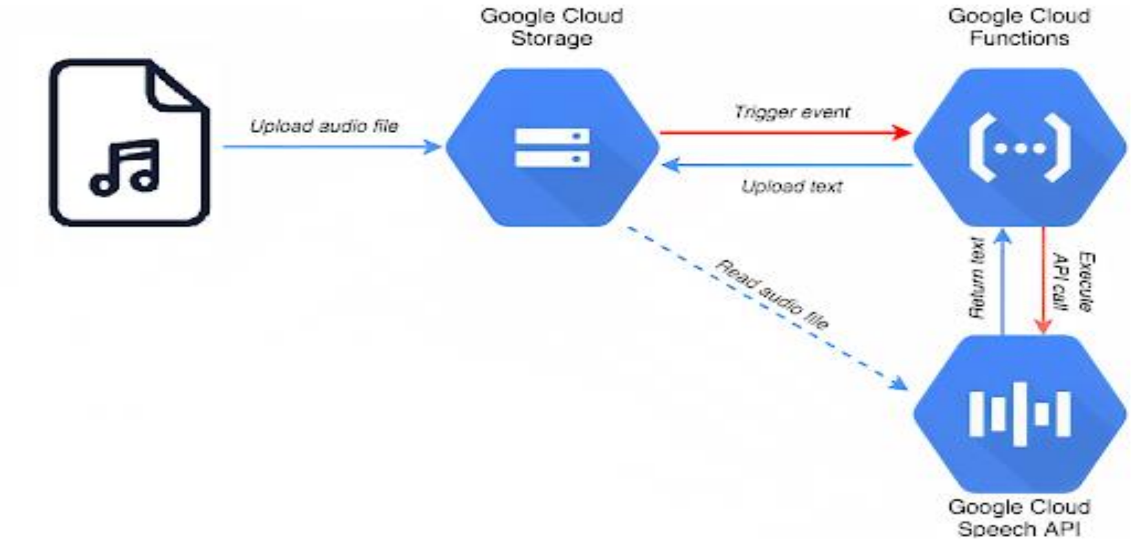


Fig : process of web speech

6. Google speech recognition technology

Google's speech recognition technology is one of the leading solutions in the field, offering high accuracy and robust performance across various languages and environments. Google employs advanced deep learning techniques, including deep neural networks (DNNs), to power its speech recognition systems. Here's an overview of Google's speech recognition technology and its key features:

Google Cloud Speech-to-Text API: Google provides the Cloud Speech-to-Text API, which enables developers to transcribe audio files or real-time speech into text. The API utilizes deep learning models trained on large datasets to achieve high accuracy in recognizing spoken language. It supports a wide range of languages and accents, as well as various audio formats and quality levels.

Deep Neural Networks (DNNs): Google's speech recognition technology leverages DNNs to model the acoustic properties of speech and perform feature learning from audio inputs. These neural networks are trained on massive datasets of annotated speech samples to learn complex patterns and representations, enabling accurate transcription of spoken language.

Language Modeling: In addition to acoustic modeling, Google's speech recognition systems incorporate language models to improve transcription accuracy. Language models capture the statistical properties of natural language, such as word sequences and probabilities, allowing the system to better predict the most likely words or phrases given the context of the speech.

Noise Reduction and Enhancement: Google's speech recognition technology includes algorithms for noise reduction and audio enhancement to improve accuracy in noisy environments or with poor audio quality. These algorithms help filter out background noise and enhance the clarity of the speech signal before transcription.

Real-Time and Batch Processing: Google's speech recognition systems support both real-time and batch processing of audio inputs. Real-time processing allows for immediate transcription of live speech streams, while batch processing enables transcription of pre-recorded audio files in bulk.

Integration with Google Products and Services: Google's speech recognition technology is integrated into various Google products and services, including

Google Assistant, Google Search, Google Translate, and more. This integration enables users to interact with these services using voice commands or input, facilitating natural and intuitive user experiences.

5.1 Deep neural network in machine learning:

Deep neural networks (DNNs) have significantly advanced the field of speech recognition, leading to notable improvements in accuracy and performance. Here's how DNNs are applied in speech recognition:

Feature Extraction: Speech signals are initially converted into a sequence of feature vectors that represent the acoustic properties of the speech signal. Commonly used features include Mel-Frequency Cepstral Coefficients (MFCCs) or filter bank energies. DNNs can be employed to learn hierarchical representations of these features, capturing complex patterns and variations in the speech signal.

Acoustic Modeling: DNNs are used to model the acoustic properties of speech. This involves training the network to map input feature vectors (e.g., MFCCs) to output units representing phonemes, sub-word units, or context-dependent units. DNN-based acoustic models have shown superior performance over traditional Gaussian Mixture Models (GMMs) due to their ability to capture non-linear relationships and dependencies in the data.

Language Modeling: In addition to acoustic modeling, DNNs are also employed for language modeling. Language models capture the statistical properties of natural language, such as word sequences and probabilities. DNN-based language models can effectively capture long-range dependencies and contextual information, leading to more accurate recognition of spoken utterances.

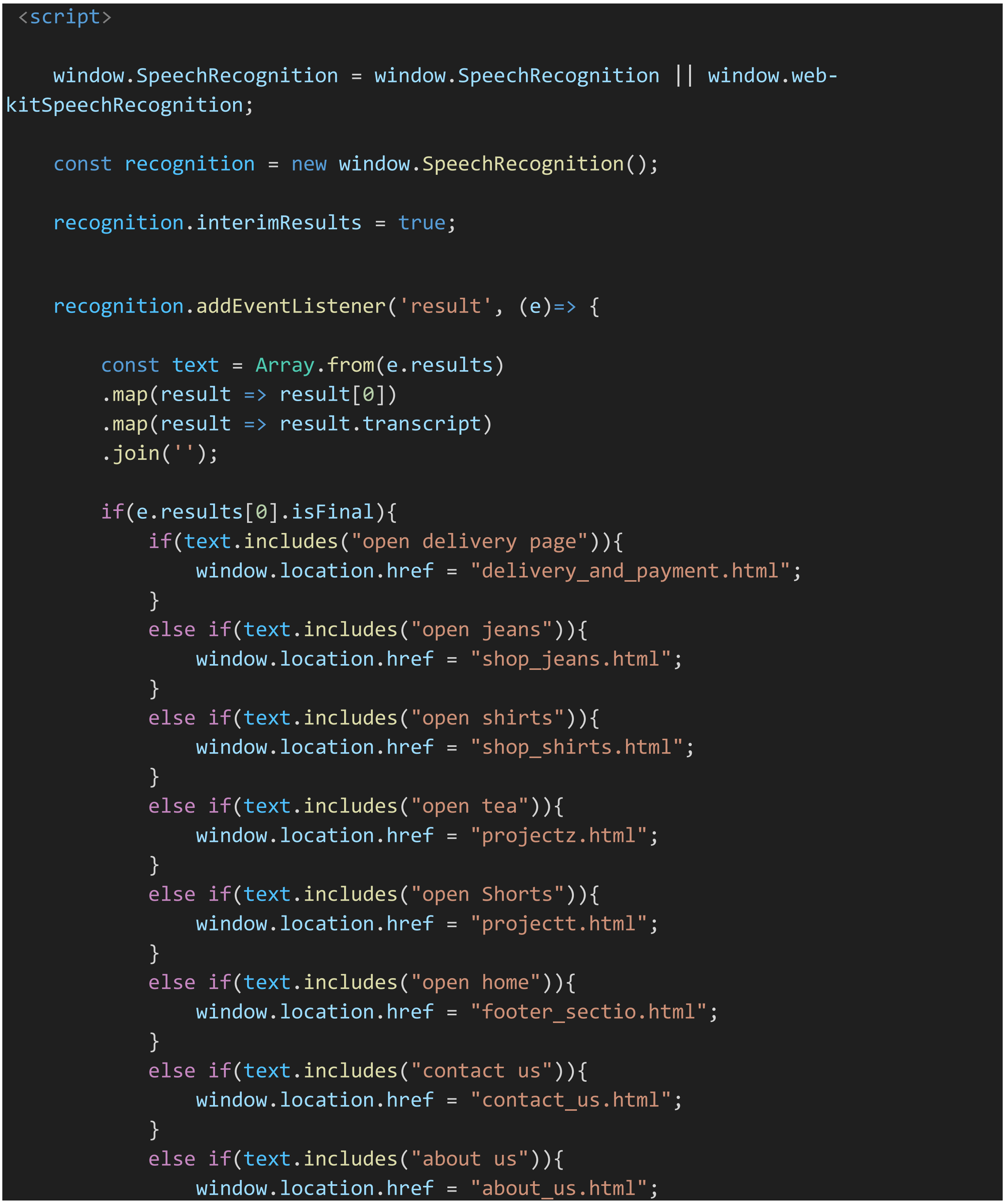
End-to-End Speech Recognition: DNNs have enabled the development of end-to-end speech recognition systems, where a single neural network is trained to directly transcribe speech input into text output without relying on intermediate components such as phoneme or word alignments. End-to-end models simplify the training process and can potentially achieve better performance by jointly optimizing acoustic and language modeling.

Speaker Adaptation: DNNs can be adapted to specific speakers or environments through techniques such as speaker normalization or speaker adaptation. These techniques enable speech recognition systems to better accommodate variations in speech patterns, accents, or background noise.

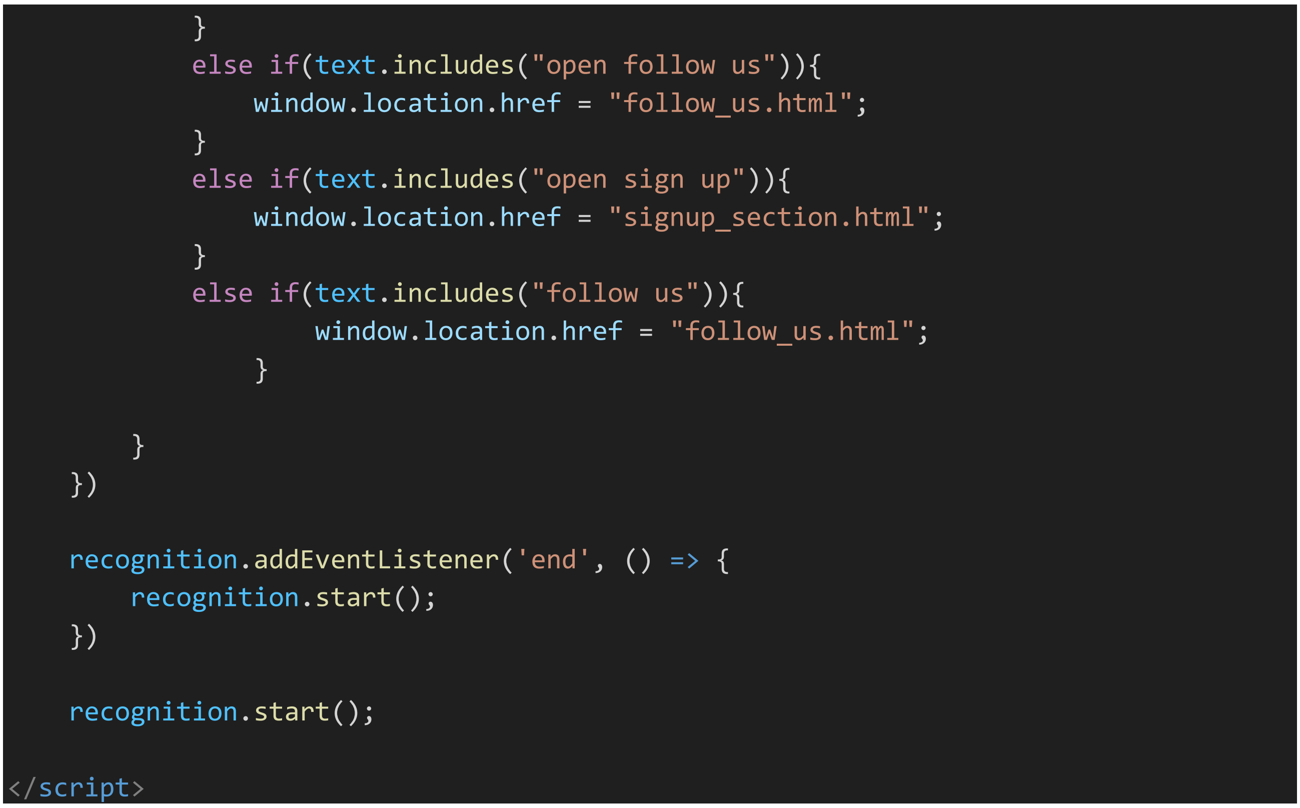
Large-Scale Training: DNN-based speech recognition systems benefit from large-scale training data, which allows the models to learn robust representations of speech features and linguistic patterns. Advances in deep learning frameworks, computational resources, and data collection techniques have facilitated the training of large-scale DNN models for speech recognition.

8. Implementation & Result

Code:



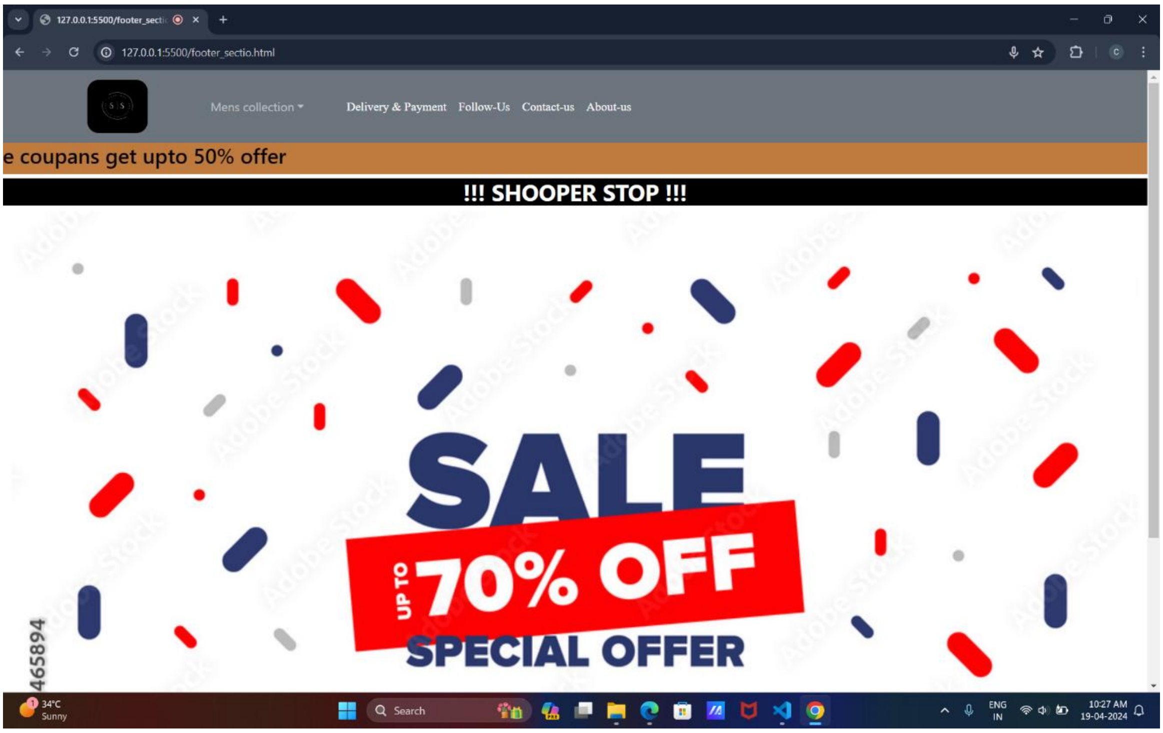
```
 <script>

    window.SpeechRecognition = window.SpeechRecognition || window.webkitSpeechRecognition;

    const recognition = new window.SpeechRecognition();

    recognition.interimResults = true;


    recognition.addEventListener('result', (e)=> {

        const text = Array.from(e.results)
        .map(result => result[0])
        .map(result => result.transcript)
        .join('');

        if(e.results[0].isFinal){
            if(text.includes("open delivery page")){
                window.location.href = "delivery_and_payment.html";
            }
            else if(text.includes("open jeans")){
                window.location.href = "shop_jeans.html";
            }
            else if(text.includes("open shirts")){
                window.location.href = "shop_shirts.html";
            }
            else if(text.includes("open tea")){
                window.location.href = "projectz.html";
            }
            else if(text.includes("open Shorts")){
                window.location.href = "projectt.html";
            }
            else if(text.includes("open home")){
                window.location.href = "footer_sectio.html";
            }
            else if(text.includes("contact us")){
                window.location.href = "contact_us.html";
            }
            else if(text.includes("about us")){
                window.location.href = "about_us.html";
```



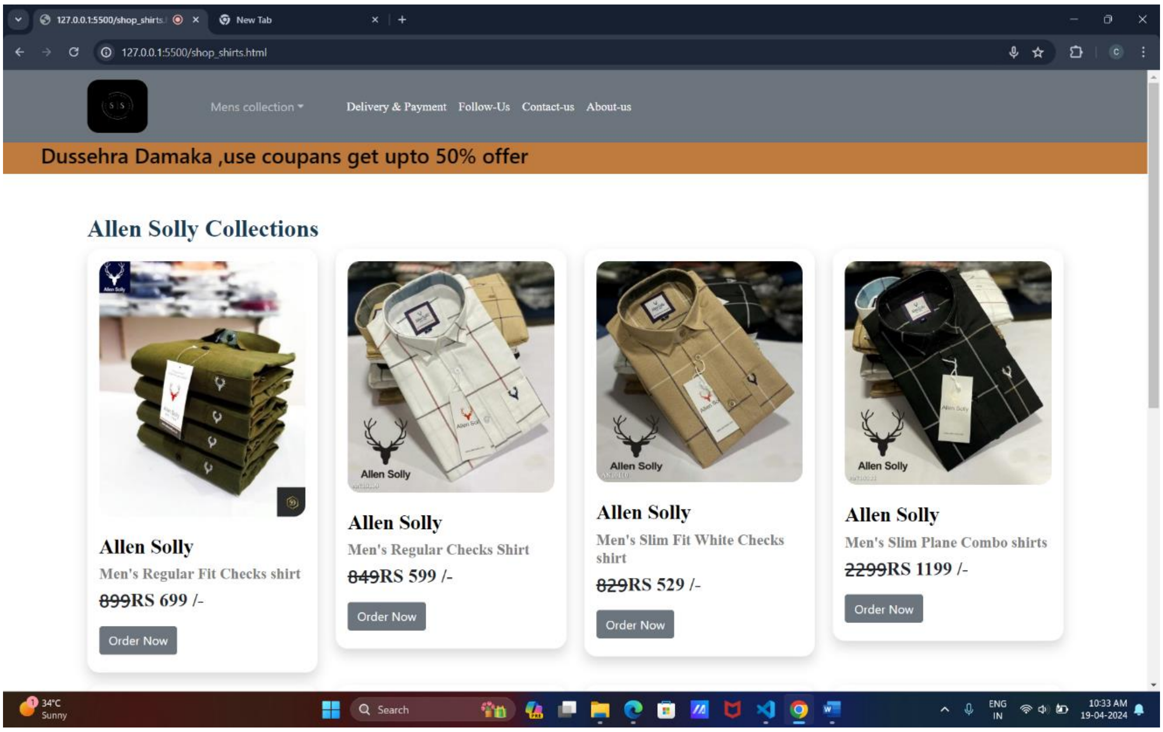
```
            }
            else if(text.includes("open follow us")){
                window.location.href = "follow_us.html";
            }
            else if(text.includes("open sign up")){
                window.location.href = "signup_section.html";
            }
            else if(text.includes("follow us")){
                    window.location.href = "follow_us.html";
                }

        }
    })

    recognition.addEventListener('end', () => {
        recognition.start();
    })

    recognition.start();

</script>
```

Result :

1. When we open the website, it asks for the permission to allow the access to microphone. The symbol of the microphone will be displayed at the right side of the url, which indicates that the microphone is enabled and ready to give commands.



2. Now, when we say command "open shirts", it will navigate to the particular section "open home " of the same page.



Dussehra Damaka ,use coupans get upto 50% offer

Allen Solly Collections

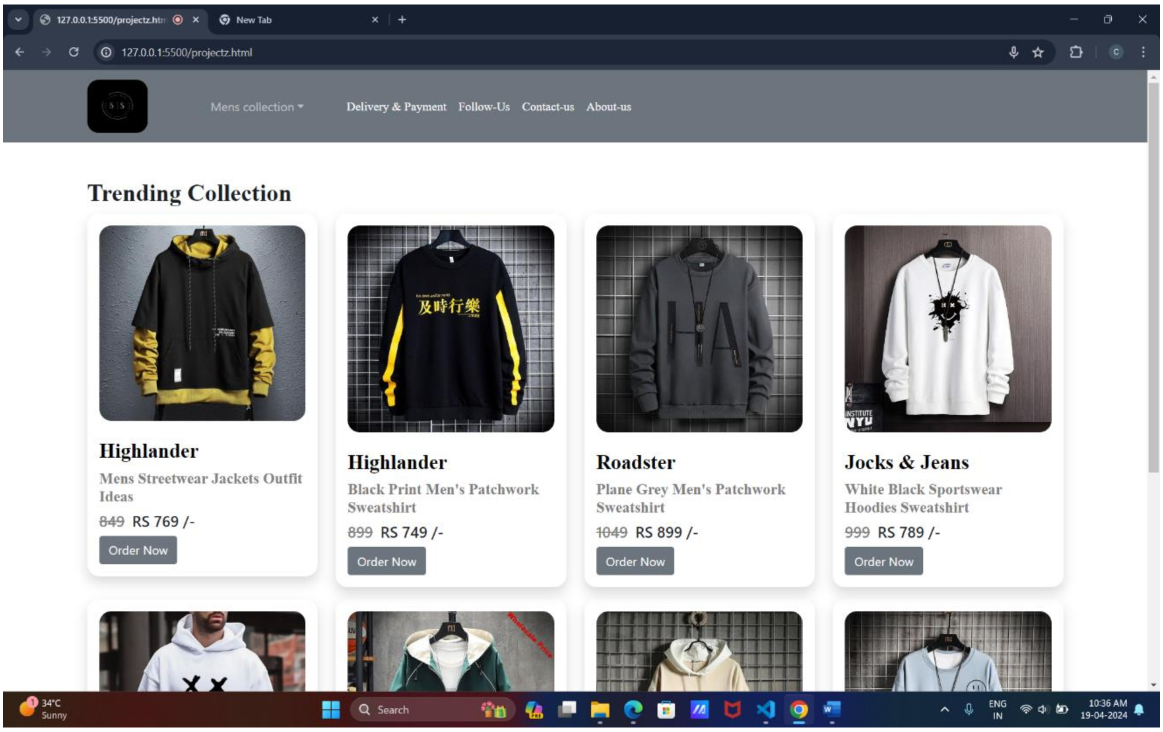
Allen Solly
Men's Regular Fit Checks shirt
~~899~~RS 699 /-
Order Now

Allen Solly
Men's Regular Checks Shirt
~~849~~RS 599 /-
Order Now

Allen Solly
Men's Slim Fit White Checks shirt
~~829~~RS 529 /-
Order Now

Allen Solly
Men's Slim Plane Combo shirts
~~2299~~RS 1199 /-
Order Now

3. When we say command "open tee", it will navigate the page tshirts page.



Trending Collection

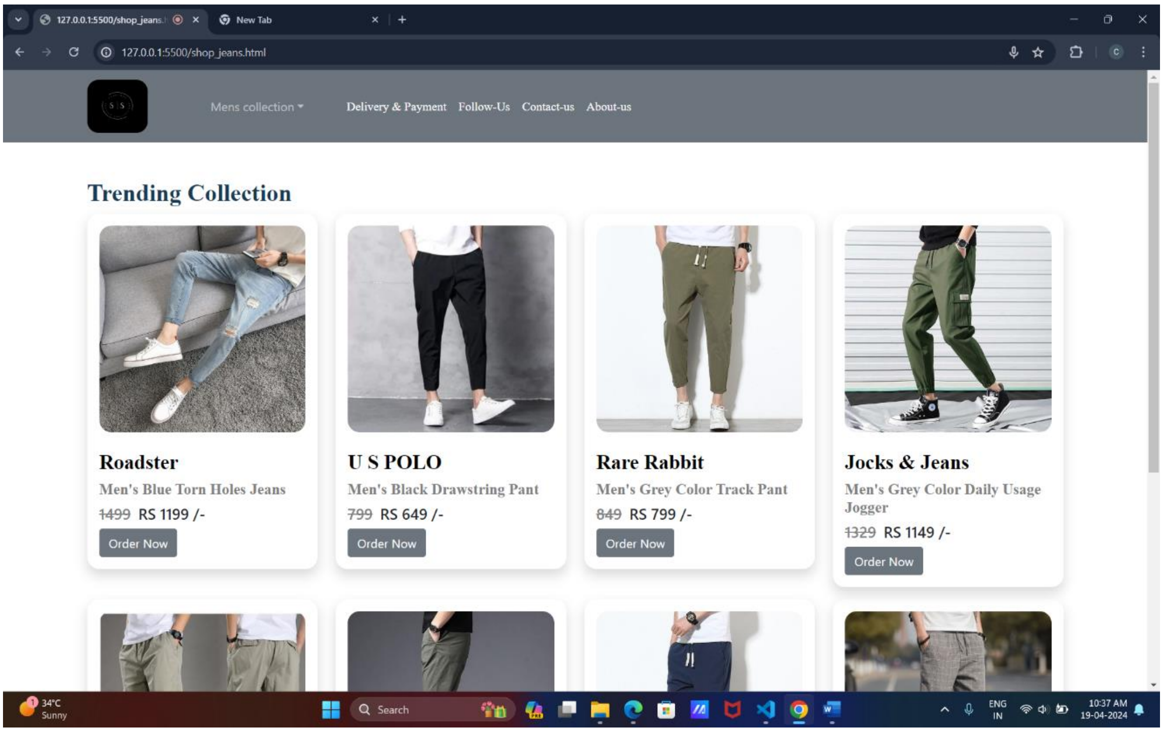
Highlander
Mens Streetwear Jackets Outfit Ideas
~~849~~ RS 769 /-
Order Now

Highlander
Black Print Men's Patchwork Sweatshirt
~~899~~ RS 749 /-
Order Now

Roadster
Plane Grey Men's Patchwork Sweatshirt
~~1049~~ RS 899 /-
Order Now

Jocks & Jeans
White Black Sportswear Hoodies Sweatshirt
~~999~~ RS 789 /-
Order Now

4. When We say command “open jeans”, it will navigate from tshirts page to jeans page.



Trending Collection

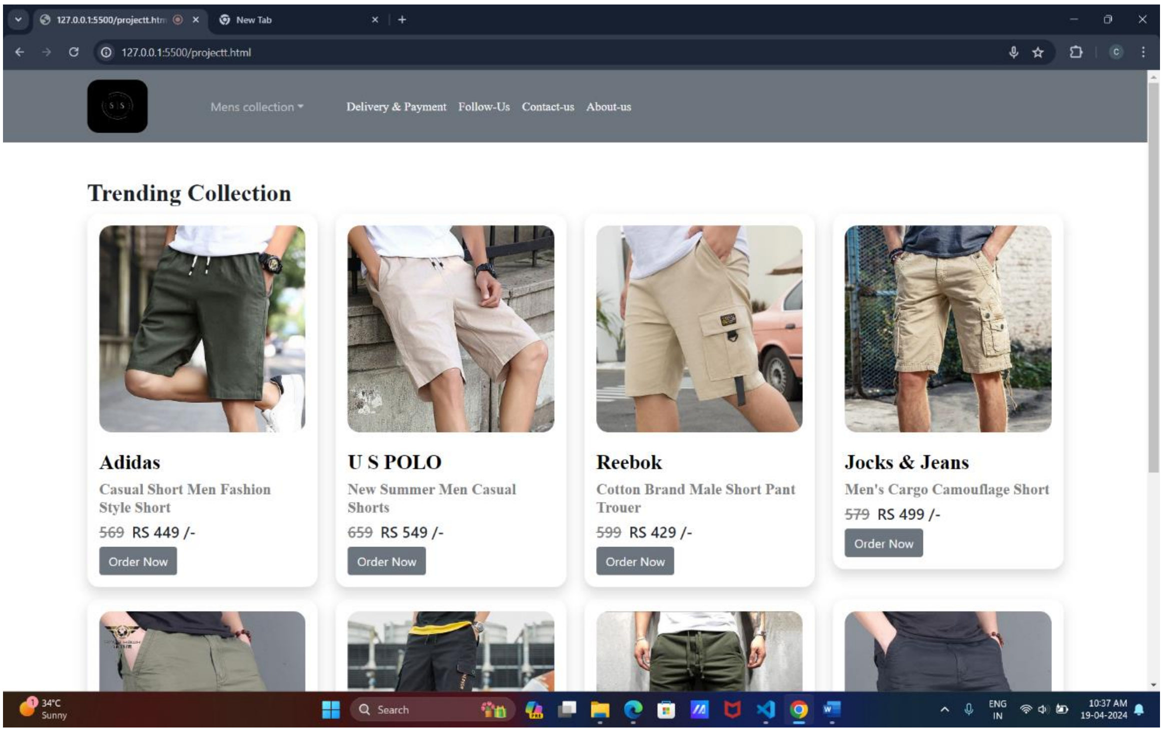
Roadster
Men's Blue Torn Holes Jeans
~~1499~~ RS 1199 /-

U S POLO
Men's Black Drawstring Pant
~~799~~ RS 649 /-

Rare Rabbit
Men's Grey Color Track Pant
~~849~~ RS 799 /-

Jocks & Jeans
Men's Grey Color Daily Usage Jogger
~~1329~~ RS 1149 /-

5. When we say command “open shorts”, it will navigate from shorts page to shorts page.



Trending Collection

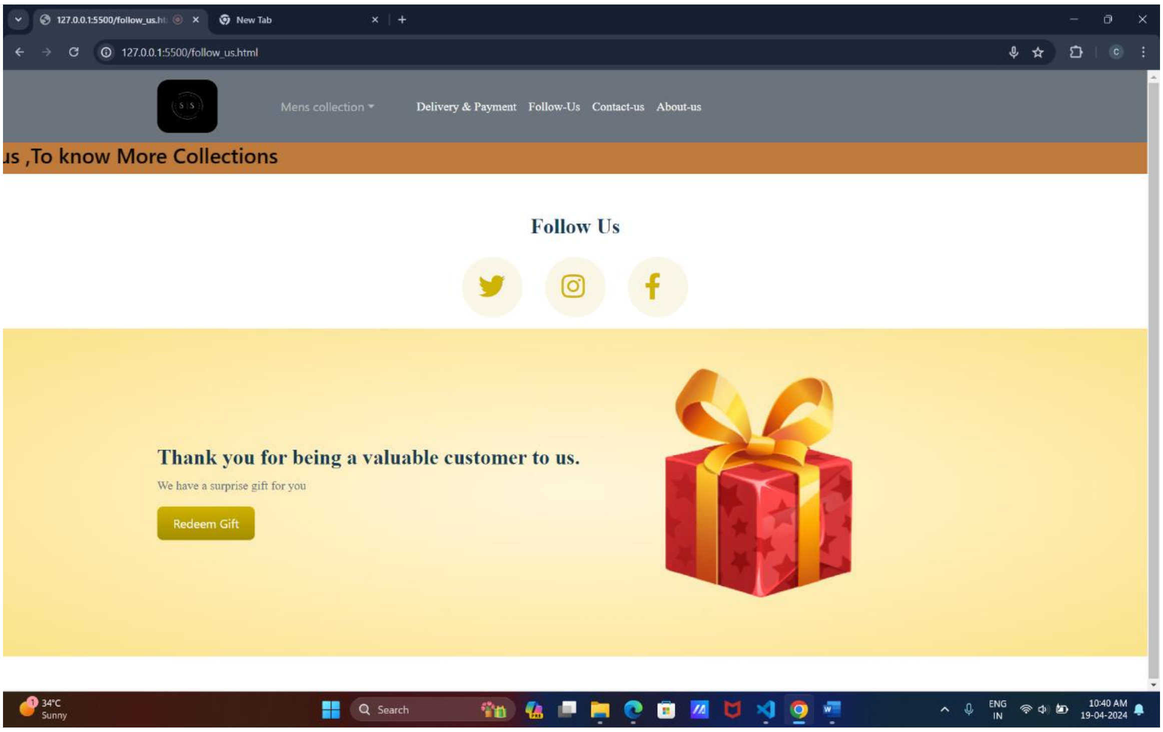
Adidas
Casual Short Men Fashion Style Short
~~569~~ RS 449 /-

U S POLO
New Summer Men Casual Shorts
~~659~~ RS 549 /-

Reebok
Cotton Brand Male Short Pant Trouer
~~599~~ RS 429 /-

Jocks & Jeans
Men's Cargo Camouflage Short
~~579~~ RS 499 /-

6. When we say command “follow us”, it will navigate to the follow us section in the home page from the shorts.



us ,To know More Collections

Follow Us

Thank you for being a valuable customer to us.

We have a surprise gift for you

7. CONCLUSION

The personal voice assistant will be easy to use and will reduce the manual human efforts for performing various tasks. The functionality of the current voice assistant system is limited to working on Desktop based system and working online only. Virtual assistance is the future of work, providing clients with cost savings, access to a global talent pool, scalability, increased efficiency, productivity, and flexibility.

A virtual assistant is a self-employed worker who specializes in offering administrative services to clients from a remote location, usually a home office. Typical tasks a virtual assistant might perform include scheduling appointments, making phone calls, making travel arrangements, and managing email accounts.

8. REFERENCES

✓ www.pythonprogramming.net

✓ www.codecademy.com

✓ www.tutorialspoint.com

✓ www.google.co.in

✓ Other Wesbites related to the topics

Books referred

✓ Python programming

✓ python code for artificial intelligence