SMART TROLLEY FOR SUPERMARKET

R.Priyanka

(IT17033374)

B.Sc. (Hons) Degree in Information Technology
Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

September 2020

SMART TROLLEY FOR SUPERMARKET

R.Priyanka

(IT17033374)

B.Sc. (Hons) Degree in Information Technology
Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

September 2020

DECLARATION OF THE CANDIDATE & SUPERVISOR

We declare that this is our own work and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
R. Priyanka	**IT17033374**	

The supervisor/s should certify the proposal report with the following declaration.

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor: Date:

ABSTRACT

Speech-to-text technology enables spoken data to be translated into text by computer. The concept of a hand-free future in which a mouse and keyboard are outdated aspects of a computer has been implemented. There is a variety of artificial intelligence currently being built on the market, and the voice assistant is one of which has increasingly gained its existence or acquired knowledge. Their main advantages are increase sales, learning and updating and Management of multiple clients Interact with customers to let them know about deals and offers, promotional codes. Voice assistant is programmed for the customers to answer questions related to supermarket as a voice message. Voice assistant can understand what the customer is saying and have in build replies in text message to give to the user. It is done through APIs. As a solution to this problem we are developing a voice assistant for our voice based dictionary system. The system will be implemented using python. If the user use similar words which cannot be identified by voice assistant it will check the dictionary connected to the voice assistant about the meaning and give the answers according to that and give answers with the text message. The system is designed so that all services provided by mobile devices can be accessed by the end user through voice. It will take the user's voice as an input, analyse the speech, and respond with appropriate answers to user's questions by using technologies like Natural Language Processing (NLP) and Text Analysis. We aim to improve question answering (QA) by decomposing hard questions into easier subquestions that existing QA systems can answer. Since collecting labeled decompositions is cumbersome, we propose an unsupervised approach to produce sub-questions**.** Many current NLP structures depend on phrase embeddings, formerly educated in an unmonitored way on big corpora, as base functions. Efforts to reap embeddings for large chunks of textual content, which includes sentences, have but not been so successful. Several tries at learning unsupervised representations of sentences have now not reached great enough performance to be broadly followed. In this paper, we show how universal sentence representations trained using the supervised records of the Stanford Natural Language Inference datasets can continuously outperform unsupervised techniques like SkipThought vectors (Kiros et al., 2015) on a extensive variety of transfer duties. Much like how pc imaginative and prescient uses ImageNet to obtain functions, which could then be transferred to other duties, our work tends to signify the suitability of herbal language inference for transfer learning to other NLP duties

Keywords-voice assistant, python, Natural Language Processing

Acknowledgement

First and foremost, we would like to thank nature which has in influenced in the random coincidental creation of ourselves and making us work with passion on what drives us. We would like to express our sincere gratitude to our research supervisor, Mr.Jeasuthasan Alosius and co supervisor Janani Tharmaseelan who accepted to supervise our research project. We are in debt to our supervisor for guidance, support, assistance, flexibility and enthusiasm in assisting us. We show or gratitude for our parents for their love, care, support, sacrifice, blessings and encouragement provided to us. We would also like to thank our friends who assisted us in sharing their knowledge and showing their support. Finally, we would also like to convey our thanks to each and every member of the group who invested their effort and time in progressing with the project.

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	**Description**
NLP	Natural Language Processing
ASR	automatic speech recognition
TTS	Text to Speech
VFT	Value-focused Thinking

1. INTRODUCTION

Voice assistant are becoming increasingly popular tools in all businesses. In this document, we describe a extending the capabilities of a voice assistant by performing actions on user’s personal mobile device. Queries such as these handled through human to human interactions. Consequently, as the number of inquiries increases, the waiting time to answer these queries increases, which results in reduced customer satisfaction. With the recent development in Natural Language Processing, this process can be made much faster and simpler by creating voice assistant to handle these tasks. In some case queries are difficult to answer customer can use the dictionary system. The dictionary consists of many words and meanings, so it is being stored in database. The goal is to provide a quick and simple solution to this problem. System manipulates the input, obtained by converting the speech to text, then it attempts to look for the word from database and it searches the particular meaning with the same method, when a specific match is found it is shown on the console tab. Natural language processing and speech popularity have become greater and more sophisticated due to improved computing power, huge availability of linguistic records, stepped forward system learning techniques, and a higher information of human language. Question answering system (QAS) is coping with the fields related to statistics retrieval and herbal language processing (NLP) which affords an descriptive solutions to the questions posed by way of the people in herbal language. The QAS gives answers for factoid, non-factoid and Boolean type of questions. But typically, cutting-edge query answering system focus on factoid questions. The serps like Google, Yahoo presents a list of net hyperlinks as the consequences for the given question. As search engine gives answers in weblinks people flow in the direction of QAS because it affords descriptive answer in a minimal time

Speech recognition systems have evolved over the decades

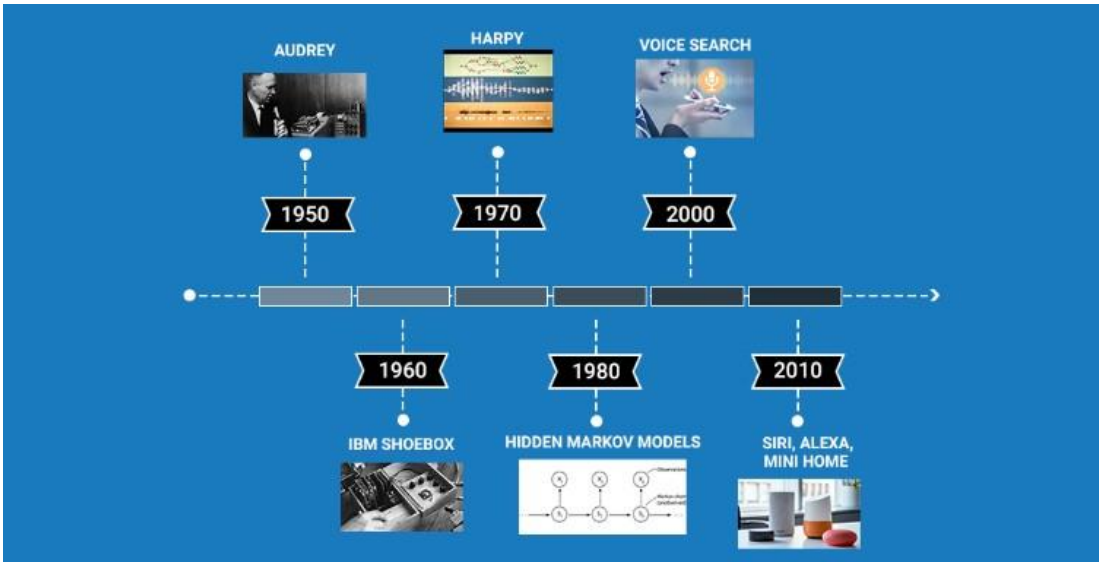
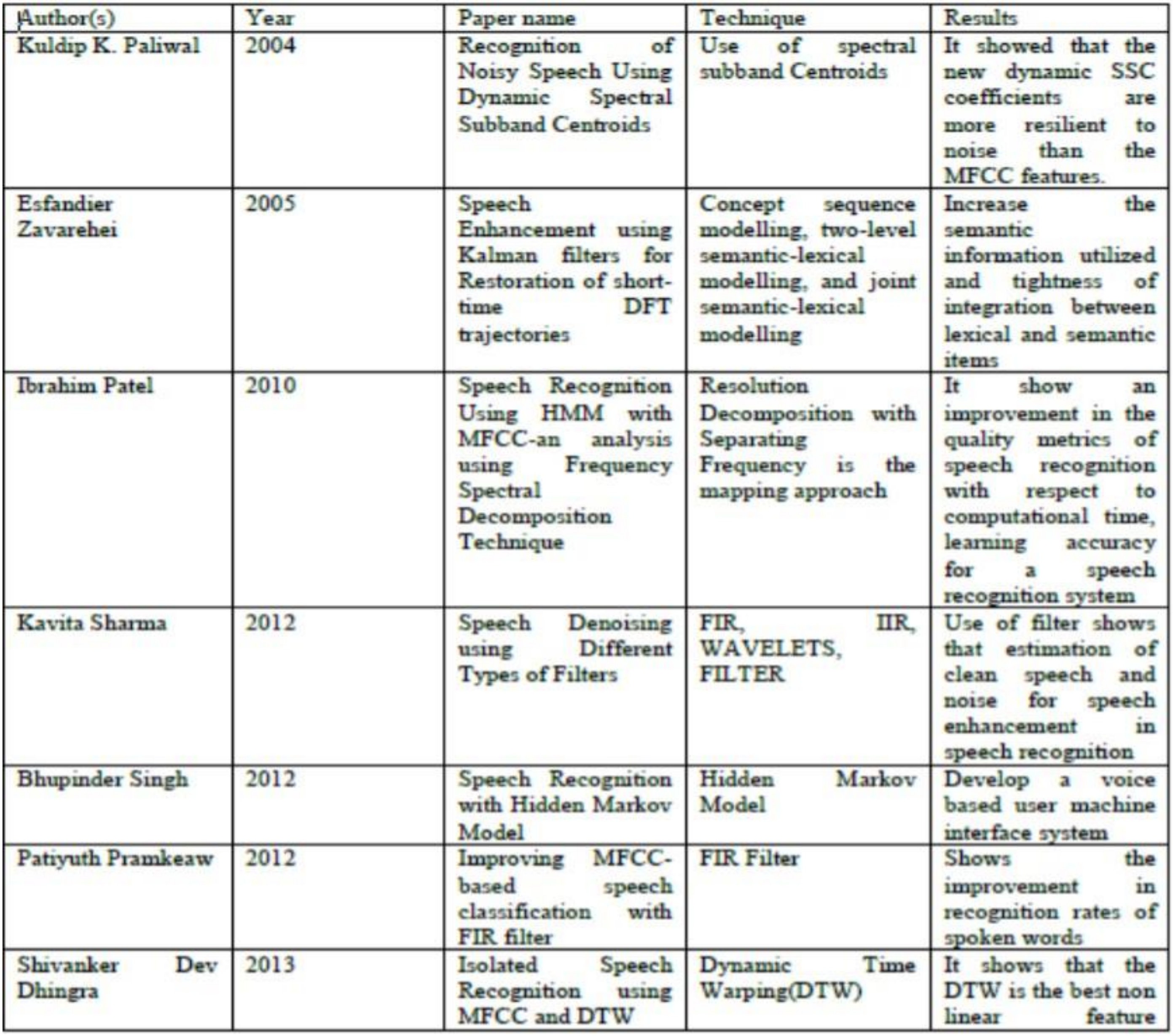


Figure: 1.1 voice assistant

1.1 Background Literature

1.1.1Background

Kuldip K. Paliwal and et al in the year 2004 had discussed that without being affected by their popularity for front end parameter in speech recognition, the cepstral coefficients which had been obtained from linear prediction analysis is sensitive to noise. Here, the use of spectral subband centroids had been discussed by them for robust speech recognition [3]. In recent years, voice-based interactions with computers have been embodied in the persona of voice assistants or agents, which can be bespoke devices in the home, integrated into cars and wearables or on smartphones. At present, there is no standard software architecture for voice assistants [4]. The study of speech recognition and transcription began in the 1936 with AT & T's Bell Labs. Formerly, most research was funded and performed by Universities and the U.S. Government (primarily by the Military and DARPA- Defence Advanced Research Project Agency).In the early 1980's the Speech Recognition technology reached at commercial market [3].



Author(s)	Year	Paper name	Technique	Results
Kuldip K. Paliwal	2004	Recognition of Noisy Speech Using Dynamic Spectral Subband Centroids	Use of spectral subband Centroids	It showed that the new dynamic SSC coefficients are more resilient to noise than the MFCC features.
Esfandier Zavarehei	2005	Speech Enhancement using Kalman filters for Restoration of short-time DFT trajectories	Concept sequence modelling, two-level semantic-lexical modelling, and joint semantic-lexical modelling	Increase the semantic information utilized and tightness of integration between lexical and semantic items
Ibrahim Patel	2010	Speech Recognition Using HMM with MFCC-an analysis using Frequency Spectral Decomposition Technique	Resolution Decomposition with Separating Frequency is the mapping approach	It show an improvement in the quality metrics of speech recognition with respect to computational time, learning accuracy for a speech recognition system
Kavita Sharma	2012	Speech Denoising using Different Types of Filters	FIR, IIR, WAVELETS, FILTER	Use of filter shows that estimation of clean speech and noise for speech enhancement in speech recognition
Bhupinder Singh	2012	Speech Recognition with Hidden Markov Model	Hidden Markov Model	Develop a voice based user machine interface system
Patiyuth Pramkeaw	2012	Improving MFCC-based speech classification with FIR filter	FIR Filter	Shows the improvement in recognition rates of spoken words
Shivanker Dev Dhingra	2013	Isolated Speech Recognition using MFCC and DTW	Dynamic Time Warping(DTW)	It shows that the DTW is the best non linear feature

Figure: 1.1.1.1 Speech Recognition

In this paper [15] we focus on the cross-lingual setting by learning representations on unlabeled data that generalize across languages. We build on a concurrently introduced pretraining approach [6] which jointly learns contextualized representations of speech as well as a discrete vocabulary of latent speech representations. The latter serves to effectively train the model with a contrastive loss (§ 2). These discrete latent speech representations are shared across languages.

This paper presents a compositional, attentional model for answering questions about a variety of world representations, including images and structured knowledge bases. The model translates from questions to dynamically assembled neural networks, then applies these networks to world representations (images or knowledge bases) to produce answers. We take advantage of two largely independent lines of work: on one hand, an extensive literature on answering questions by mapping from strings to logical representations of meaning; on the other, a series of recent successes in deep neural models for image recognition and captioning. By constructing neural networks instead of logical forms, our model leverages the best aspects of both linguistic compositionality and continuous representations.[16]
Kuldip K. Paliwal and et al in the year 2004 had discussed that without being affected by their popularity for front end parameter in speech recognition, the cepstral coefficients which had been obtained from linear prediction analysis is sensitive to noise. Here, the use of spectral subband centroids had been discussed by them for robust speech recognition [3]. In recent years, voice-based interactions with computers have been embodied in the persona of voice assistants or agents, which can be bespoke devices in the home, integrated into cars and wearables or on smartphones. At present, there is no standard software architecture for voice assistants [4]. The study of speech recognition and transcription began in the 1936 with AT & T's Bell Labs. Formerly, most research was funded and performed by Universities and the U.S. Government (primarily by the Military and DARPA- Defence Advanced Research Project Agency).In the early 1980's the Speech Recognition technology reached at commercial market [3].

Current scientific knowledge of the impact of localization of voice assistants on engagement and loyalty is understudied. One reason for this may be that voice assistants are still in the early stages of the product lifecycle for many companies and consumers. While there are white papers (Invoca, 2018) discussing the positive impact on engagement and customer loyalty through AI, there is limited knowledge of consumer perceptions toward using localized voice assistants for tasks in their daily lives[1]. Speech Technology in language learning field has been extensively researched and developed in recent years. It can be deployed to improve dictionary applications. In this paper, a speech-based dictionary system refers to an application that is created for computer and provides users with word or phrase definition by using speech interface [2].

Speech recognition systems can be classified on basis of the following parameters [8]:

- **Speaker**: It have a different kind of voice. The models hence are either designed for a specific speaker.

- **Vocal Sound**: The way the speaker speaks also plays a role in speech recognition. Some models can recognize either single utterances or separate utterance with a pause in between

- **Vocabulary**: The size of the vocabulary plays an important role in determining the complexity, performance, and precision of the system.

In this paper[17], they look at the task of studying prevalent representations of sentences, i.E., a sentence encoder version this is trained on a large corpus and ultimately transferred to other obligations. Two questions need to be solved with a purpose to build such an encoder, namely: what's the most efficient neural community structure; and how and on what assignment ought to any such community study. Following present paintings on gaining knowledge of phrase embeddings, most cutting-edge approaches bear in mind studying sentence encoders in an unsupervised manner like SkipThought (Kiros et al. , 2015) or FastSent (Hill et al. , 2016)

In this paper[18], they introduce How2, a massive-scale dataset of tutorial videos masking a wide variety of topics across 80,000 clips (approximately 2,000 hours), with word-level time alignments to the ground-reality English subtitles. In addition to being multimodal, How2 is multilingual: we crowdsourced Portuguese translations of the subtitles. We gift outcomes for monomodal and multimodal baselines on several language processing obligations with interesting insights on the software program of diverse modalities. We desire that via making the How2 dataset and baselines to be had they'll encourage collaboration throughout language, speech and imaginative and prescient communities. They introduce How2, a dataset of tutorial motion pictures paired with spoken utterances, English subtitles and their crowdsourced Portuguese translations, in addition to English video summaries. The pervasive multimodality of How2 makes it an ideal resource for growing new fashions for multimodal knowledge.

In this paper[19], they introduce recurrent neural community grammars (RNNGs), a new generative probabilistic model of sentences that explicitly models nested, hierarchical relationships amongst phrases and terms. RNNGs perform through a recursive syntactic technique harking back to probabilistic context-unfastened grammar technology, but selections are parameterized the usage of RNNs that condition on the whole syntactic derivation records, significantly enjoyable context-free independence assumptions. The basis of this paintings is a top-down model of transition-primarily based definitely parsing . We offer two variations of the algorithm, one for parsing (given an found sentence, rework it right into a tree), and one for technology.

In this paper[20], they have provided a method for including contextual differences to phrase embeddings with a second section of embedding. This contextual statistics gains energy in distinguishing amongst distinctive aspects of phrases. Experimental effects with embedding of the English variation of Wikipedia (over 2 billion words) indicates giant upgrades in each semantic- and syntactic- primarily based phrase embedding overall performance. The result also provides a extensive range of interesting standards of phrases in expressivity evaluation. These results strongly guide the concept of the use of context embeddings to take advantage of context data for troubles in NLP. As we highlighted in advance, context embeddings are underutilized, despite the fact that word embeddings had been notably exploited in a couple of applications.

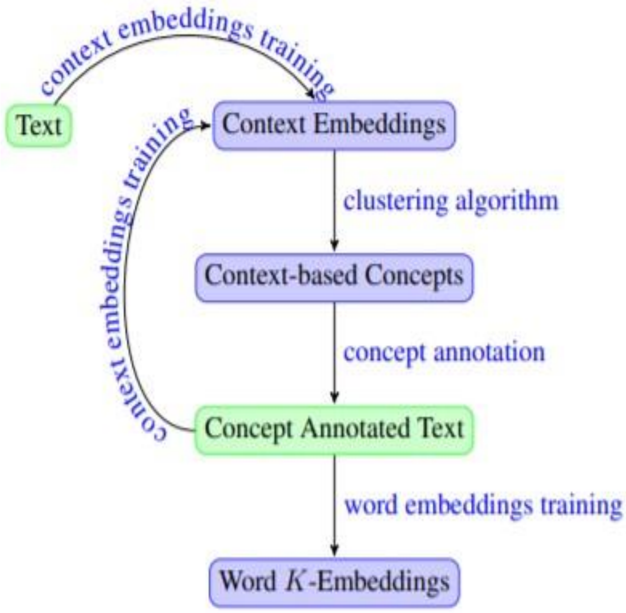


Figure 1.1.1.2 Training Word K-Embeddings

In this paper [21], the Factored NMT approach has been further explored. Factors primarily based on linguistic a priori knowledge were used to decompose the target phrases. This technique outperforms a strong baseline gadget the usage of subword units computed with byte pair encoding. Our FNMT machine is able to model an nearly 6 instances larger word vocabulary with only a moderate growth of the computational cost. By these method, the FNMT machine is capable of halve the generation of unknown tokens as compared to wordlevel NMT. Using a simple unknown phrase alternative process regarding a bilingual dictionary, we're capable of attain even higher outcomes. Also, using external linguistic resources allows us to generate new word bureaucracy that might not be blanketed in the popular NMT gadget shortlist. The advantage of this technique is that the new generated words are managed with the aid of the linguistic understanding, that avoid producing wrong phrases, in preference to actual structures using BPE. We tested the performance of this type of machine on an inflected language (French). The consequences are very promising to be used with fantastically inflected languages like Arabic or Czech.

In this paper[22], we focus at the cross-lingual putting by way of studying representations on unlabeled records that generalize across languages. We build on a simultaneously introduced pretraining method [6] which jointly learns contextualized representations of speech as well as a discrete vocabulary of latent speech representations. The latter serves to efficiently educate the version with a contrastive loss . These discrete latent speech representations are shared throughout languages.

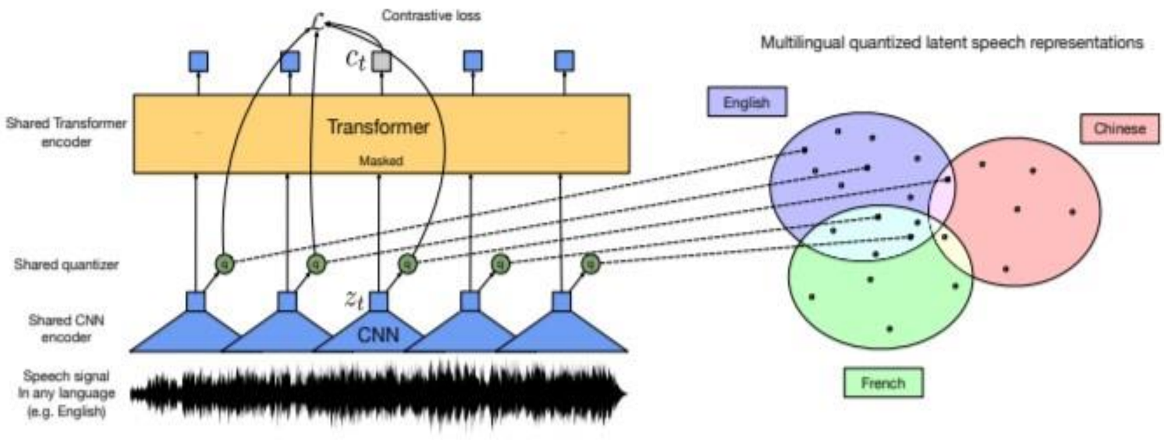

Contrastive loss
$\mathcal{L}$
c_t
Shared Transformer encoder
Transformer
Masked
Shared quantizer
q
z_t
Shared CNN encoder
CNN
Speech signal In any language (e.g. English)
Multilingual quantized latent speech representations
English
Chinese
French

Figure: 1.1.1.3 The XLSR approach

A shared quantization module over characteristic encoder representations produces multilingual quantized latent speech units whose embeddings are then used as objectives for a unmarried Transformer trained with contrastive studying. The version learns to percentage discrete tokens across languages, developing bridges across languages. Their method is inspired through [15, 36] and builds on pinnacle of wav2vec 2.Zero. It calls for most effective uncooked unsupervised speech audio from multiple languages.

In this paper [23] They added a hard and fast of responsibilities probing the linguistic knowledge of sentence embedding methods. Their reason isn't always to inspire the development of ad-hoc models that acquire pinnacle performance on them, however, to assist exploring what records is captured via unique pre-skilled encoders. Captured by using one-of-a-kind pre-trained encoders. We accomplished an extensive linguistic evaluation of modern-day sentence encoders. Our results advocate that the encoders are taking pictures a wide range of properties, well above the ones captured through a set of sturdy baselines. We further uncovered interesting styles of correlation among the probing duties and more complex “downstream” obligations and supplied a hard and fast of interesting findings approximately the linguistic houses of various embedding strategies.

In this paper [10] goal is to provide a short and simple approach to this trouble. The Virtual Assistant continues a natural conversation with the user. It can solution queries associated with the infrastructure, publications, money owed, school, and events of our institute. The Voice Assistant receives a question from the user, is familiar with the problem, and affords appropriate answers. It does this by means of changing an English sentence right into a machine- pleasant query, extracting the relevant keywords, then going thru the necessary facts and in the end returning the answer in a herbal language sentence. In other words, it solutions the questions like a human does.

In this paper [24] PARI is Designed to help Native and especially for Blind men and women which fits on their Voice Commands. PARI additionally has the capability of recognizing the voice commands without internet connection. PARI has diverse functionalities of mobile devices like network connection and managing diverse programs on just the voice commands. Contains key features like Voice Pattern Detection, Keyword Learning, and so on. Which beneficial for give up user to use numerous functionalities and offerings of the cellular devices. Hence, PARI is language barrier impartial which actively responds to person’s voice commands quicker than the Online Voice Search packages.

In this paper [1] This observe grants empirical help for the direct, indirect and mediating outcomes of engagement on technology adoption constructs and the moderating effect of localization at the respective relationships, thereby imparting critical implications for advertising concept and practice. If Google and Apple offer localized service, they may be rewarded with higher patron loyalty. Localization has been mentioned extensively within the international advertising and marketing literature. However, there may be little known about the effect of localizing a voice assistant for engagement, goal to apply and client loyalty. Our findings support the perception that era localization is grounded inside the more potent customer–firm relationships as characterized not simplest by multiplied engagement however also higher customer loyalty.

In this paper [4], we brought UIVoice, a machine and framework that allows voice interactions for current mobile packages with out requiring any code modifications inside the original software. As voice assistants have significant reliance at the cloud for his or her ASR and NLU capabilities, we trust our version which entails significant operations on quit-person gadgets, poses exciting challenges for the edge computing network.

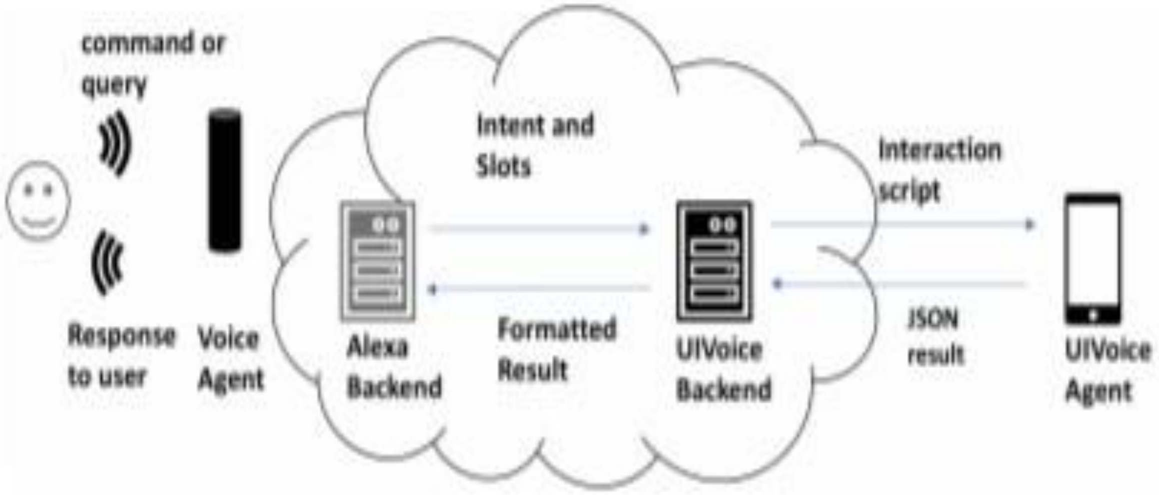


Figure 1.1.1.4 System design

In this paper[25] As voice assistants grow to be higher at learning consumer alternatives and behavior, they will more and more have an impact on purchaser behaviors (Simms, 2019). In doing so, VAs can also anticipate a central relational position in the patron marketplace and step by step mediate market interactions. These rapid-changing market dynamics in the context of voice purchasing may additionally have a extreme impact on customer manufacturers and shops. Loss of brand visibility, the increased relevance of retailers' personal labels, and the growth in advertising charges are just a number of the effects predicted with the aid of advertising and marketing and technology experts. In mild of those potential dynamics, researchers are referred to as to study the interplay between purchasers, manufacturers, and retailers' behaviors in response to "machine behaviors" (Rahwan et al., 2019). Providing structure and steerage to researchers and marketers to be able to further discover this emerging circulate of research is fundamental.

In This paper [26] proposes an incremental studying strategy for neural word embedding strategies, such as SkipGrams and Global Vectors. Since our method iteratively generates embedding vectors one dimension at a time, acquired vectors equip a unique assets. Namely, any right-truncated vector matches the answer of the corresponding decrease-dimensional embedding. Therefore, a unmarried embedding vector can control a extensive range of dimensional requirements imposed by way of many one of a kind makes use of and programs. The main reason of this paper is to further beautify the 'usability' of obtained embedding vectors in real use. In addition, ITACO can also be a very good alternative of SGNS and GloVe in phrases of the execution pace of a single run. Now, we're free from retraining unique dimensions of embedding vectors by using using ITACO. Our technique appreciably reduces the full calculation fee and storage, which improves the 'usability' of embedding vectors.

In this paper [27] goal is to investigate algorithms for combining supervised studying with self-play — which we call supervised self-play (S2P) algorithms — the usage of classic emergent communication tasks: a Lewis signaling recreation with symbolic inputs, and a greater complex photo-based referential

recreation with herbal language descriptions. Our first locating is that supervised studying observed through self-play outperforms emergent verbal exchange with supervised great-tuning in those environments, and we offer three reasons for why that is the case. We then empirically check out several supervised-first S2P strategies in our environments. Existing approaches in this location have used numerous advert-hoc schedules for alternating among the two kinds of updates (Lazaridou et al., 2017), but to our understanding there was no systematic have a look at that has compared those strategies. Lastly, we advocate the usage of population-based totally techniques for S2P, and discover that it leads to stepped forward performance inside the greater hard picture-primarily based referential game. Our findings highlight the want for further work in combining supervised getting to know and self-play to expand more pattern-green language learners.

In this paper [28], we mentioned the topics relevant to the improvement of SpeechTo Text structures. The speech to text conversion may additionally seem effective and efficient to its users if it produces natural speech and by making several modifications to it. This device is beneficial for deaf and dumb humans to Interact with the other peoples from society. Speech to Text synthesis is a important research and alertness location within the discipline of multimedia interfaces. In this paper gathers important references to literature related to the endogenous versions of the speech sign and their importance in automated speech recognition. A database has been comprised of the numerous domain phrases and syllables. The desired speech is produced with the aid of the Concatenative speech synthesis method. Speech synthesis is nice for individuals who are visually handicapped. This paper made a clear and easy evaluation of operating of speech to text device (STT) in little by little procedure. The system offers the input facts from mice in the shape of voice, then preprocessed that statistics & converted into text format displayed on PC.

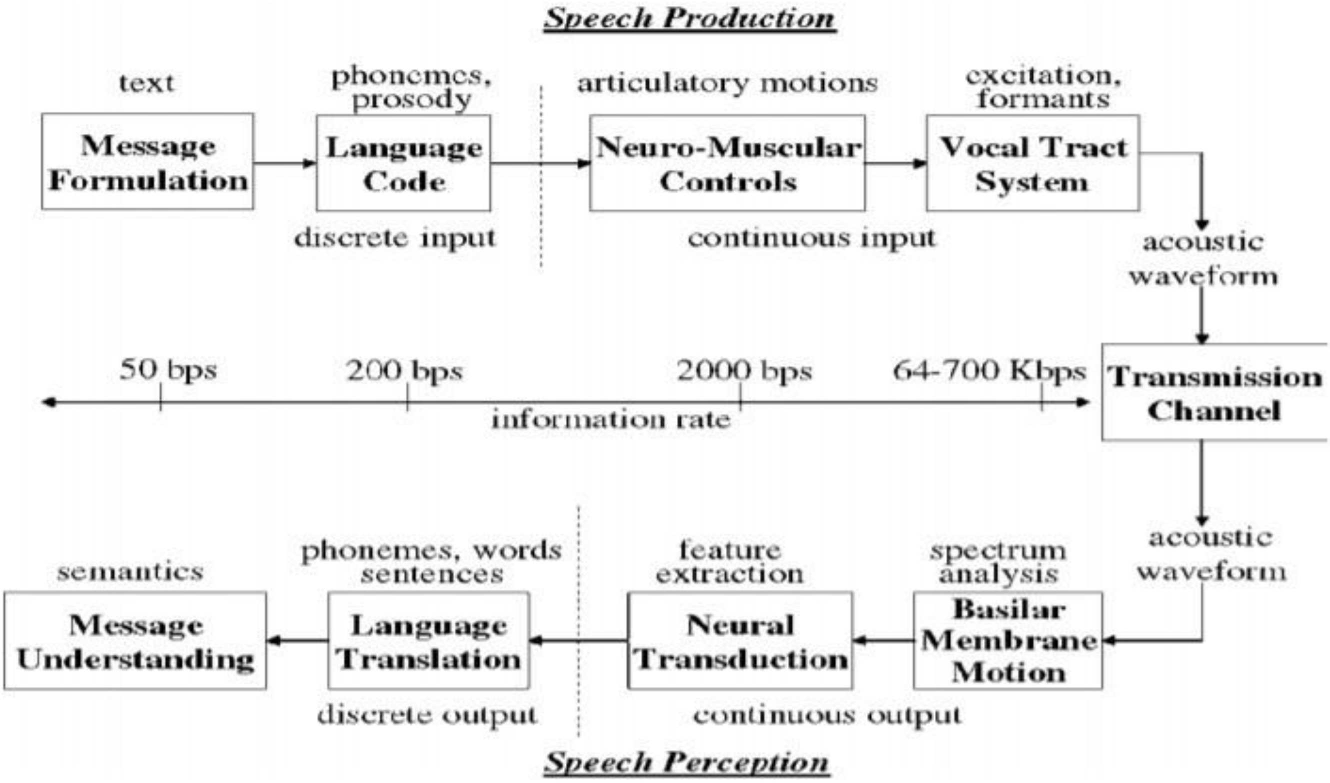


Figure 1.1.1.5 Speech chain

In this paper [30] The aim of our method is to put into effect the speech reputation using Google API. From the literature survey, we've come to a end that Hidden Markov Models and Deep Neural Networks are the great guess to enforce speech popularity and the toolkit we've chosen had been designed with the equal hidden Marko fashions and deep neural networks concept. The acoustic fashions is trained to educate the device to apprehend the speech patterns. The whole corpus information have is constructed absolutely if you want to make the speech-to-text conversion gadget to be best. Feature extraction within speaker identification ought to be much less stimulated with the aid of noise or the man or woman's health [3]. This is an essential aspect that is being taken into consideration at the same time as constructing the device.

In This paper [31] is based as follows: First, the theoretical background covers extant studies on speech interplay, voice assistants and means-quit chain theory. Second, VFT (Value Focused Thinking) as research technique and the information series and analysis method are described. Third, we gift and discuss findings from the interviews, compare them to extant research, and provide implications for each researchers and practitioners.

In this paper[32] The proposed system of our task offers answers to open domain questions in the mode of audio layout. Web pages extraction, answer identification and text to speech conversion are the processes concerned in this proposed system. Then user gives the question inside the user interface, the question is searched in the search engine and related web pages are retrieved from it. Web pages are stored in nearby documents, from the internet pages snippets and provide to the user in the shape of audio as well as text.

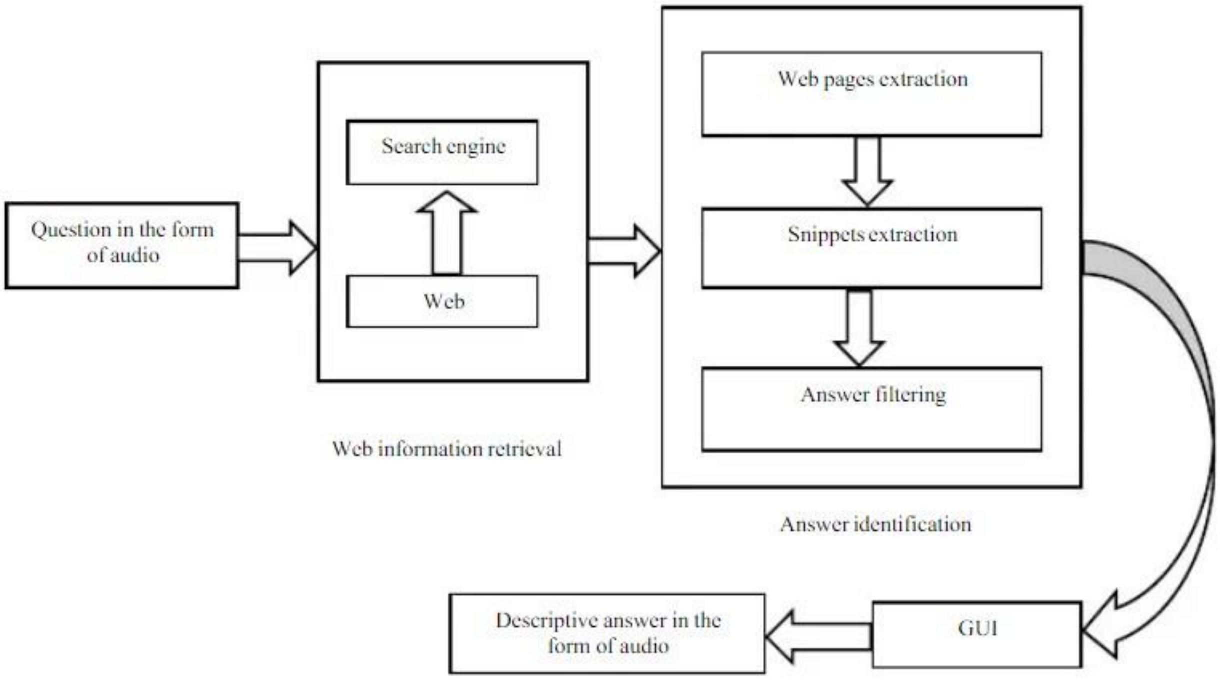


Figure 1.1.1.6 Proposed system architecture

1.2 Research Gap

	Speaker	**Speech to Text**	**Dictionary**	**Updating Offers &Promotions**
How Voice Can Change Customer Satisfaction: A Comparative Analysis between E- Commerce and Voice Commerce	✓	✗	✗	✗
Speech-based Dictionary Application	✗	✓	✓	✗
Voice Recognition System: Speech to Text	✗	✓	✗	✗
In My System	✓	✓	✓	✓

Table: 1.2.1 research gap table

When we analyse research papers regarding voice assistant that says speaker, speech to text and dictionary mechanism have in some and all three mechanisms mentioned above are not in one particular voice assistant. so, we planned to implement all these three mechanisms in one voice assistant and as an addition to this we are implementing sending notifications about promotions and discounts also which in not in all three mentioned before.

1.3Research Problem

1. Customer has to ask manually to sales representatives –

If the customer has some queries like damaged items returnable or price related queries every time they have to ask the sales representatives manually about these queries.

2. Customer doesn’t know about promotions and offers.-

In some circumstances customer doesn’t know about discounts for credit debit cards or foodcity cards or offers until they are going to cash counters. If it is informed earlier through the notifications they there is a high possibility to motivate them to buy the items which has discounts.

1.4 Objectives

1.4.1Main Objective

The main Objective of this study to implementing voice assistant to give a quick reply for customer queries and notify them about discounts and promotions .By giving the notification about discounts and promotions earlier we motivate our customers to buy things. So the sales will be increased.

1.4.2Specific Objectives

- Let the user to ask queries using microphone
- Voice assistant will check the queries with the answers already implemented in the database.
- If the user use similar words which cannot be identified by voice assistant it will check the dictionary connected to the voice assistant about the meaning and give the answers according to that and give answers with the textmessage
  Ex: price =cost

Fundamental Objectives of voice assistant	
Maximize efficiency - Ensure faster task completion	Maximize convenience - Minimize physical effort
Maximize ease of use - Ensure easy access to underlying system	Minimize cognitive effort - Minimize deliberate thinking
Maximize enjoyment - Maximize joy of the interaction itself	

Table 1.4.1

Means Objectives	
Ensure hands -free and eyes-free use - maximize multitasking	Minimize speech recognition errors - minimize repetitions of commands
Maximize naturalness of conversation - Minimize ‘command style’ interaction	Maximize system transparency - Maximize user understanding of the system
Ensure offline functionality - Minimize dependence on internet connection	Maximize compatibility - Maximize integration of different devices & apps
Maximize system adaptation - Maximize learning from users	Maximize trust - Ensure that the system does what is expected

2. METHODOLOGY

2.1 Methodology

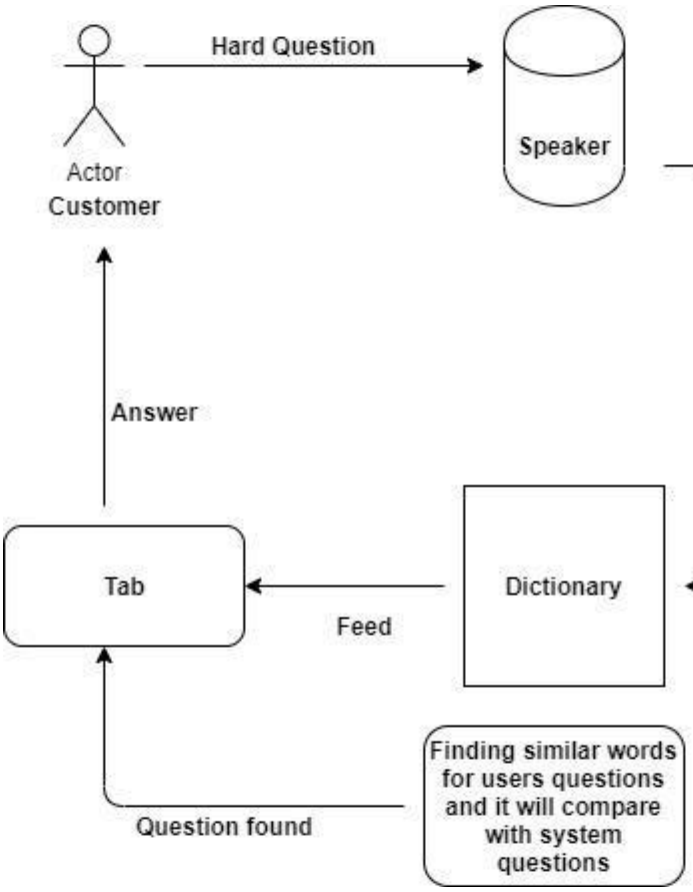


Figure: 2.1.1 methodology

First we will let the customer to ask question using microphone system will check the answer with the already implemented system code words in some case system will not understand some similar words (ex :place =location) .In that case it will search the dictionary and find the meaning of the word and then display the answer to the customer according to the meaning.

Speech to Text is a software. It is control pc features and dictates text by using voice. The system consists of two components, first component is for acoustic sign which is captured by using a microphone and second issue is to interpret the processed signal, then mapping of the signal to words. This may be executed by using growing voice reputation system: speech-to-text which lets in laptop to translate voice request and dictation into text. Voice popularity system: speech-to- text is the manner of changing an acoustic sign that is captured the use of a microphone to a hard and fast of words. The recorded facts may be used for document preparation.

Voice recognition system: Speech to text

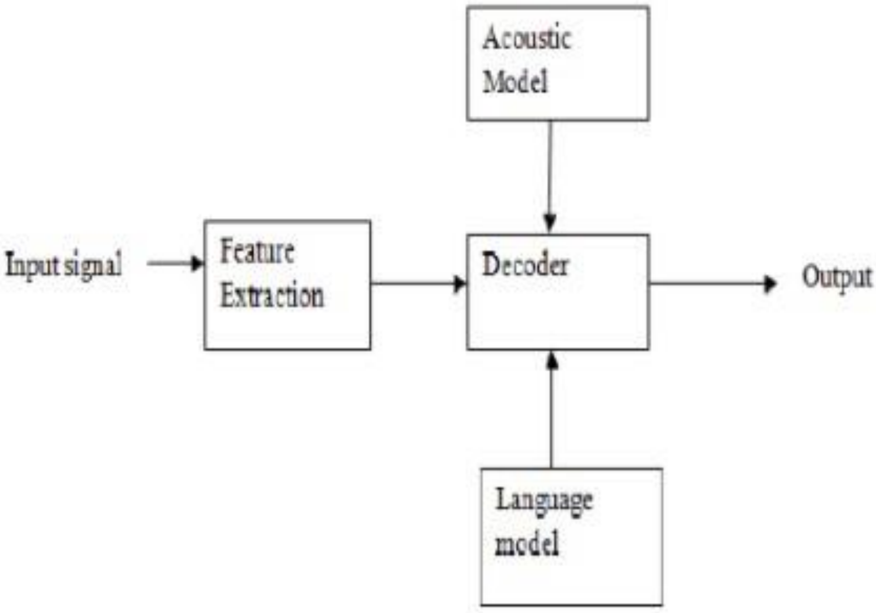
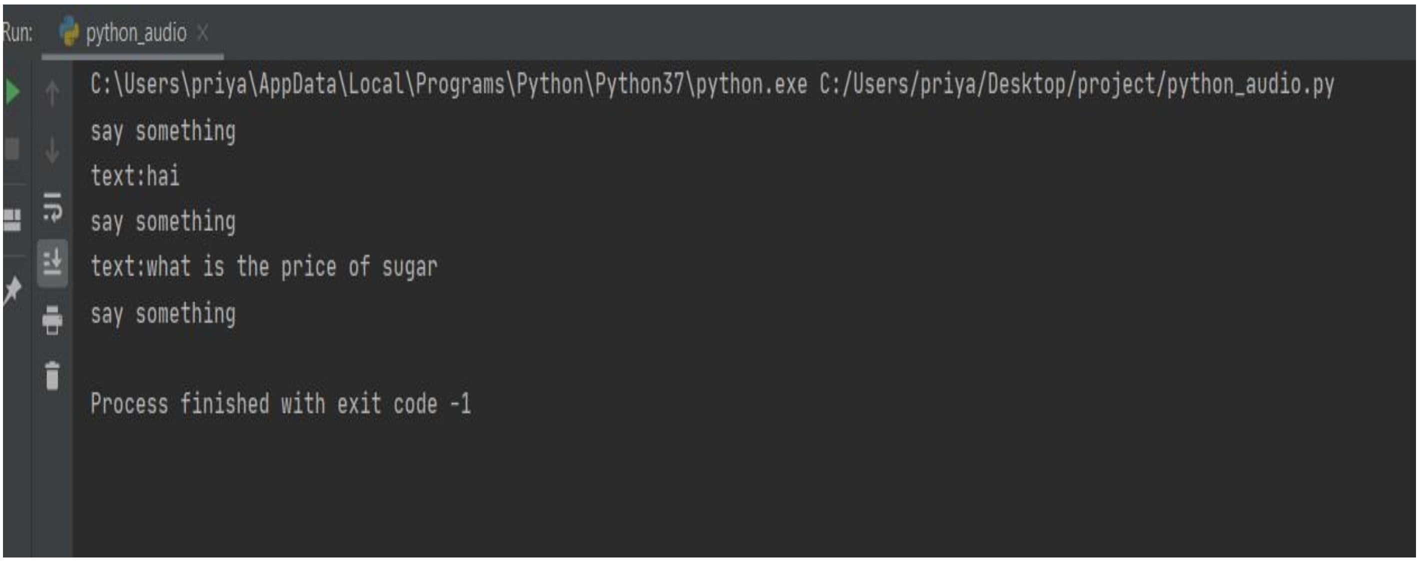


Figure 2.1.2 Speech to Text



```
C:\Users\priya\AppData\Local\Programs\Python\Python37\python.exe C:/Users/priya/Desktop/project/python_audio.py
say something
text:hai
say something
text:what is the price of sugar
say something

Process finished with exit code -1
```

Figure 2.1.3 speech to text

In here I decided to build the questions and answers model. However, my intention isn't always to reach the state-of-the-art accuracy however to study exceptional NLP principles, put in force them and explore more solutions. I usually believed in starting with basic models to know the baseline and this has been my technique here as properly. This part will focus on introducing Facebook sentence embeddings and the way it is able to be used in building QA structures. In the future components, we can try and put into effect deep getting to know techniques, specially collection modeling for this problem.

Dataset

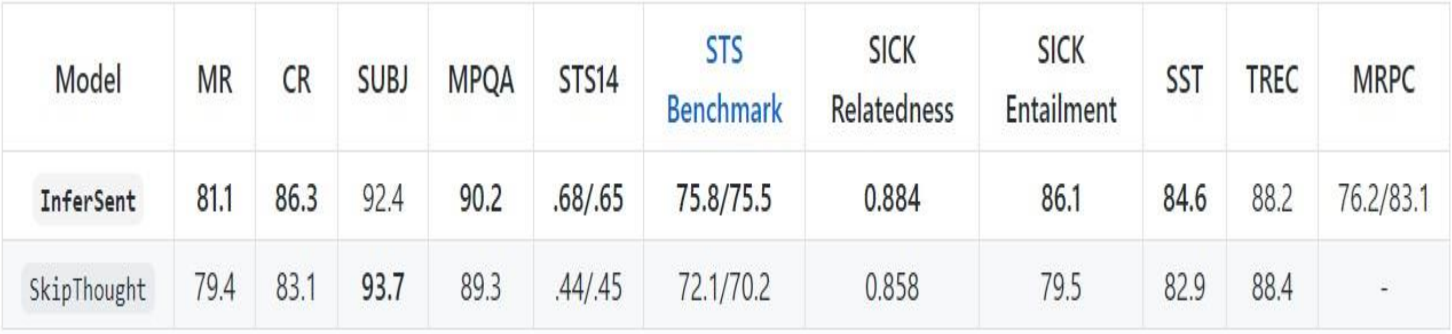
I have created the questions and answers samples datas by my own.

Introducing Infersent, facebook sentence embedding

InferSent is a sentence embeddings method that provides semantic representations for English sentences. It is trained on natural language inference information and generalizes nicely to many different tasks.

The first takes care of creating a dictionary of sentence embedding for all the sentences and questions of training dataset. Then unsupervised file calculates the distance between sentence & questions basis Euclidean & cosine similarity using sentence embeddings. It's accuracy is 45% &63% respectively.

I used the best model is infersent1.pkl.



Model	MR	CR	SUBJ	MPQA	STS14	STS Benchmark	SICK Relatedness	SICK Entailment	SST	TREC	MRPC
InferSent	81.1	86.3	92.4	90.2	.68/.65	75.8/75.5	0.884	86.1	84.6	88.2	76.2/83.1
SkipThought	79.4	83.1	93.7	89.3	.44/.45	72.1/70.2	0.858	79.5	82.9	88.4	-

Figure 2.1.4   Sample results

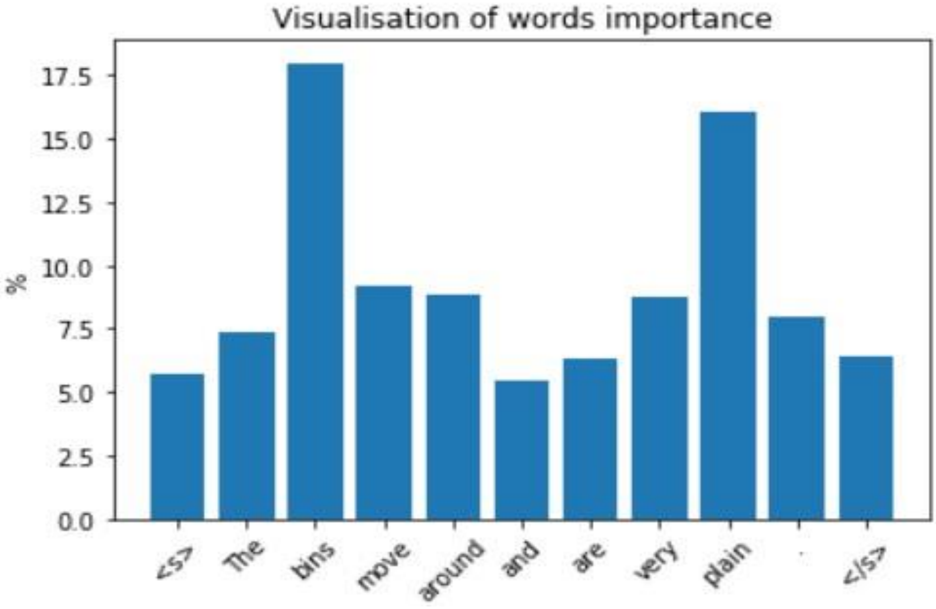
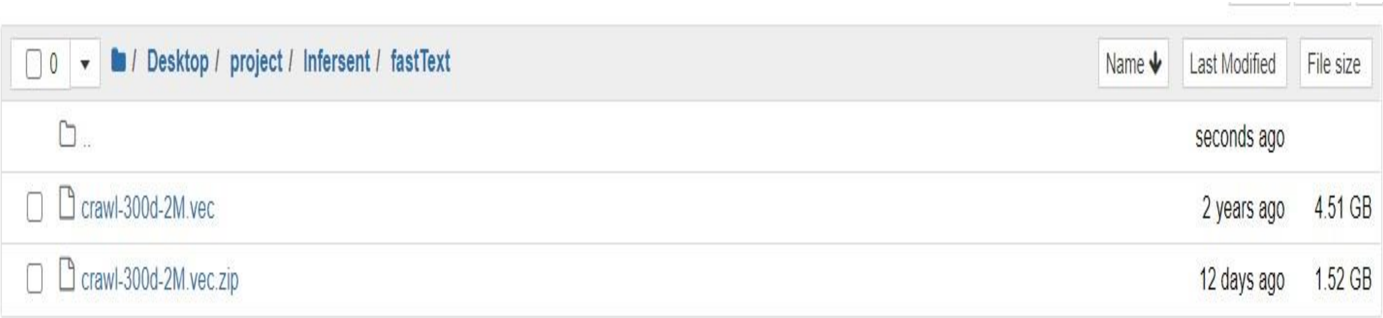


Figure 2.1.5 Sample word result

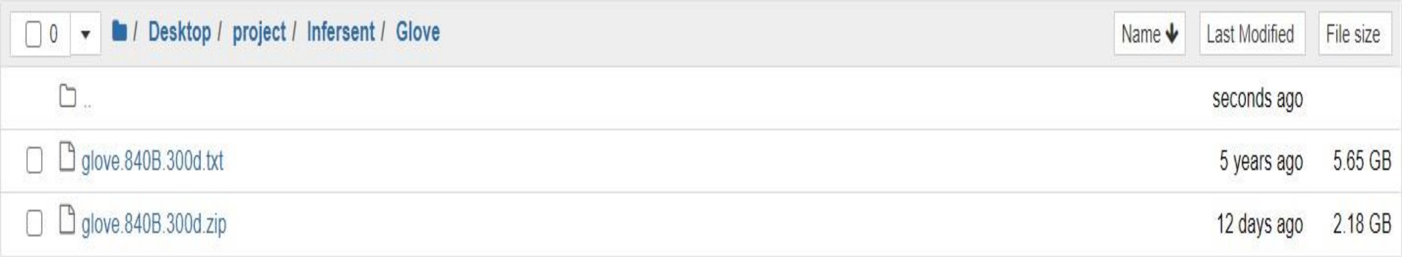


□ 0 ▾ / Desktop / project / Infersent / fastText

Name ↓	Last Modified	File size
..	seconds ago	
□ crawl-300d-2M.vec	2 years ago	4.51 GB
□ crawl-300d-2M.vec.zip	12 days ago	1.52 GB

Figure 2.1.6 FastText

An extension to Word2Vec suggested by Facebook in 2016 is FastText. FastText splits words into many n-grammes (sub-words) instead of feeding individual words into the Neural Network. For example, app, ppl, and ple (ignoring the beginning and end of word boundaries) are the tri-grams for the word apple. The sum of all these n-grams will be the word embedding vector for the apple. After training the Neural Network, provided the training dataset, we will have word embedding for all the n-grams. Rarity, rare. It is now possible to accurately describe rare terms since it is extremely likely that some of their n-grams will also occur in other terms. In the following section, I am going to show you how to use FastText with Gensim.



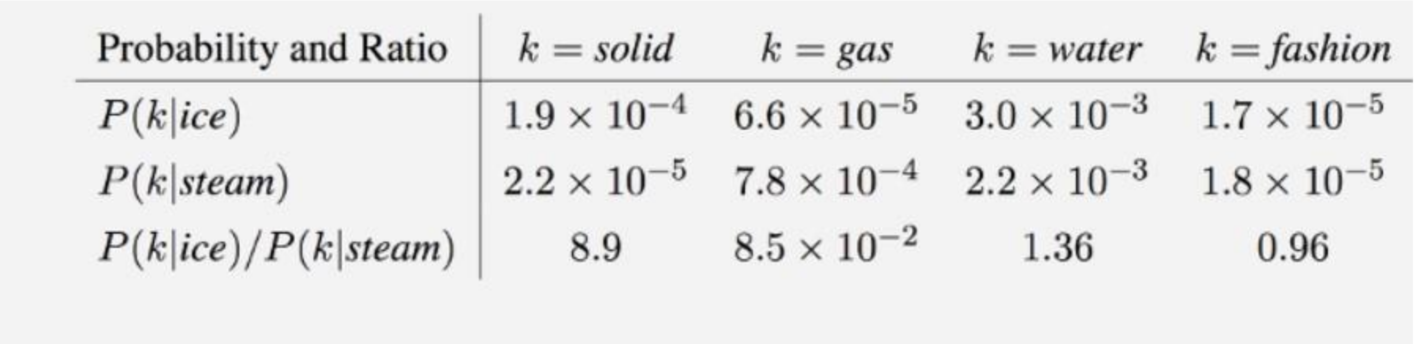
□ 0 ▾ / Desktop / project / Infersent / Glove

Name ↓	Last Modified	File size
..	seconds ago	
□ glove.840B.300d.txt	5 years ago	5.65 GB
□ glove.840B.300d.zip	12 days ago	2.18 GB

Figure 2.1.7 Glove

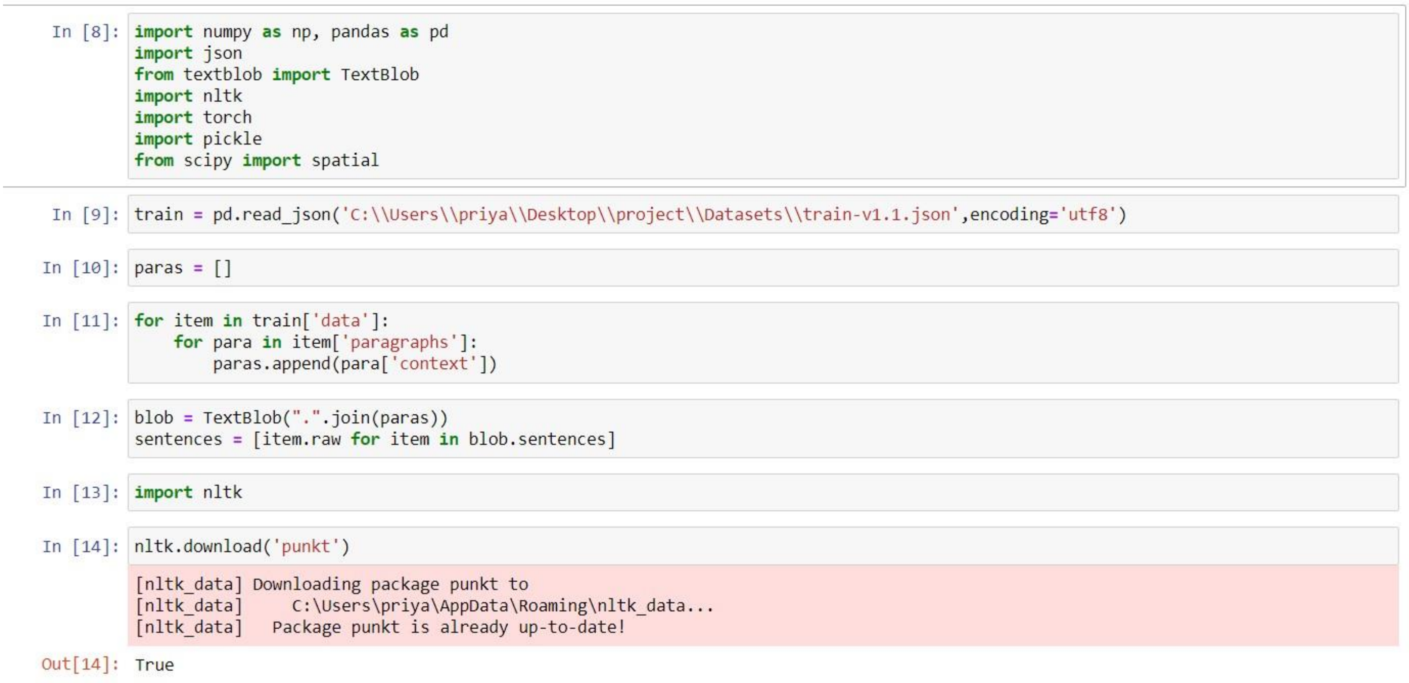
GloVe is an algorithm for unsupervised learning to obtain vector representations for terms. Training is carried out on aggregated global word-word co-occurrence statistics from a corpus, and interesting linear substructures of the word vector space are shown by the resulting representations.

GloVe is basically a log-bilinear model with a minimum-squares target that is weighted. The simple observation that ratios of word-word co-occurrence probabilities have the capacity to encode some sort of meaning is the key intuition underlying the model. For example, with various probe words from the vocabulary, consider the co-occurrence probabilities for target words ice and steam. Here are some real probabilities from a corpus of 6 billion words:



Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36	0.96

Figure 2.1.8 glove model overview



```
In [8]:  import numpy as np, pandas as pd
         import json
         from textblob import TextBlob
         import nltk
         import torch
         import pickle
         from scipy import spatial

In [9]:  train = pd.read_json('C:\\Users\\priya\\Desktop\\project\\Datasets\\train-v1.1.json',encoding='utf8')

In [10]: paras = []

In [11]: for item in train['data']:
             for para in item['paragraphs']:
                 paras.append(para['context'])

In [12]: blob = TextBlob(".".join(paras))
         sentences = [item.raw for item in blob.sentences]

In [13]: import nltk

In [14]: nltk.download('punkt')

         [nltk_data] Downloading package punkt to
         [nltk_data]     C:\Users\priya\AppData\Roaming\nltk_data...
         [nltk_data]   Package punkt is already up-to-date!

Out[14]: True
```

Figure 2.1.8 ipynb_checkpoints

2.2 Commercialization of the Product

The term voice assistant refers to conversational dealers who perform responsibilities with or for a character, whether practical or social in nature or not or not and have the ability to enhance their comprehension of the interlocutor and meaning by themselves. Embedded in smart objects, this app utilizes a combination of AI techniques such as Automated speech recognition, text-to-speech system and natural language knowledge to communicate with individuals in herbal conversation encounters. Such class of IoT goes underneath diverse names that encompass however are not confined to smart speaker, AI assistant, shrewd personal assistant, non-public digital assistant, voice-controlled smart assistant, voice-activated intelligent assistant, and conversational agent. The proposed approach of an open domain question answering gadget the use of NLP strategies and web snippets is done by collecting questions from exclusive domains and verified the correct consequences. Each query is exceeded through the Google seek engine for retrieving the net pages after pre-processing. From the ones net pages snippets have been extracted using the indexing approach and is made to save inside the local disk. Usually the snippet contains the statements which give brief creation approximately that document. Sometimes those snippets might not be related to the given query. So NLP technique is used, to discover the accurate solution for the given question.

As voice technology sneaks its manner into all components of our lives, it will fast come to be our trusted assistant when we want to search for information, ask a query, command a device to do something and, most significantly, buy something. Enter voice and retail.

Figure 2.2.1 Pros & cons voice commerce

Benefits of google Speech to Text: -

- Recognizes more than 120 languages
- For improved precision, multiple machine learning models.
- Recognition of automated language
- Transcripting Text
- Proper Identification of nouns
- Privacy of data
- Noise cancellation for audio from video and phone cells

Disadvantages of google Speech to Text: -

- Costs money
- Small builder of custom vocabulary

Why should we be paying attention to this fad of voices? Isn’t all of this only about these smart speakers?

As a fast-developing tech mash-up, we need to grasp Artificial Intelligence-enabled Speech (AI-voice). This is about software, not hardware, software that allows you to talk from any digital device to the internet, smart stuff, and business apps-smartphones, personal computers, remote controls, appliances(from refrigerators to coffee makers), and vehicles, as well as home-based smart speakers. Soon, it will be the most commonly used way to communicate with the modern world. In 2019 some 3.25 billion voice assistants will be in use, rising to approximately 8 billion by 2023, a compounded annual growth rate of more than25 percent, according to Juniper Research’s recent report. Moreover, in 2018, approximately one in eight google searches were conducted via speech, which is about 250 billion voice searches.

These days’s AI-voice environment is just like the early, pre-requirements days of the net. There’s no identified registry of vacation spot or dispatch names, no assured way for purchasers to discover businesses or manufacturers. There are significant troubles of personal and industrial data use and privacy. Market leaders have created proprietary, gated AI-voice structures and ecosystems. In such an environment, there’s a vital want for leaders – organization users and developers alike – to sign up for together and broaden the requirements with a purpose to unharness this transformative generation. Standards to be able to inspire use and boom. Standards on the way to allow competitive differentiation.

Why voice commerce is the E-Commerce future

Businesses will continue to find methods of controlling their fees and maximizing their sales through the strategic use of voice search structures.

Depending on the effects experienced via the brands which have embraced this sort of generation, services consisting of Google voice purchasing might be accelerated to new frontiers. Competition for having the first-class era will even upward thrust.

For instance, we will foresee a destiny of virtual concierge for lodges where a number of the traditional features are now undertaken with the aid of technology. Consumers becomes used to era as a manner of life.

Services inclusive of voice buying with Alexa and Echo becomes the norm.

Given a number of the issues that have been skilled in the enterprise, inclusive of privateness and security, it would no longer be surprising if the state corporations began to monitor the sports of merchandise which include the voice-activated shopping listing app.

Figure 2.2.2 voice activated app

The aim of such monitoring will be to resolve consumer concerns. More items may come to the market, such as the voice-activated shopping list organizer.

Before the public completely embraces them at large, some of these items may have an extended embedding span. Traditional shops and supermarkets can find a way to customize their offerings in the age of voice-assisted shopping.

This generation will remain cell with merchandise including the voice purchasing list app iPhone. At the pinnacle level of the patron, the pyramid might be expert voice buying, which includes very technical factors. Hence, a product together with a voice buying list app will ought to be used in a different way to seize the imagination of the consumer.

How brands may benefit from the e-commerce voice

The first step is to ensure that sufficient software and hardware acquisitions are made. The website hosting a voice-activated shopping list or even the support systems for a voice concierge in a modern hotel is a case in point.

Secondly, the organization in query will should educate its personnel in order that they may be cozy with the new technology. It takes effort from both the consumer and manufacturer for voice seek buying to be successful.

Thirdly, since this will strengthen their policies about the industry as a whole, these organizations will have to dig into data about voice shopping statistics.

How Voice Commerce Performs

Some guidelines for voice recognition software best practices have been developed by the industry to guide those who bring these products on the market.

The software and hardware will usually use algorithms to work. The devices of the voice assistant then become command units that follow orders set by the user. The mobile system needs to be able to hear the human voice and this is done by an in-built microphone.

The digital concierge hotel app, which relies on pre-installed software to process the instruction, is a case in point and becomes an activity the completes the task set by the user for those with smartphones. To access the related features, they can use a voice assistant app.

The voice commerce platform can have a list of results after the command while working in the sense of a search engine like Google. Such websites would then appear at the top that are either well optimized or have otherwise paid a listing fee.

2.3 Testing and Implementation

2.3.1 Testing

Software Testing is a process to evaluate the capability of a software program software with a purpose to locate whether the advanced software program met the required necessities or not and to discover the defects to ensure that the product in order-free if you want to produce the high-quality product.
Types of tests that were conducted on the system are listed below.

1. Unit Testing
2. Integration Testing
3. System Testing
4. Component Testing

Unit Testing - The scope is to validate that every unit of the software program code performs as expected. Unit Testing is completed at some stage in the improvement (coding segment) of a software by way of the builders. Unit Tests isolate a segment of code and verify its correctness. A unit can be a person feature, approach, system, module, or object.

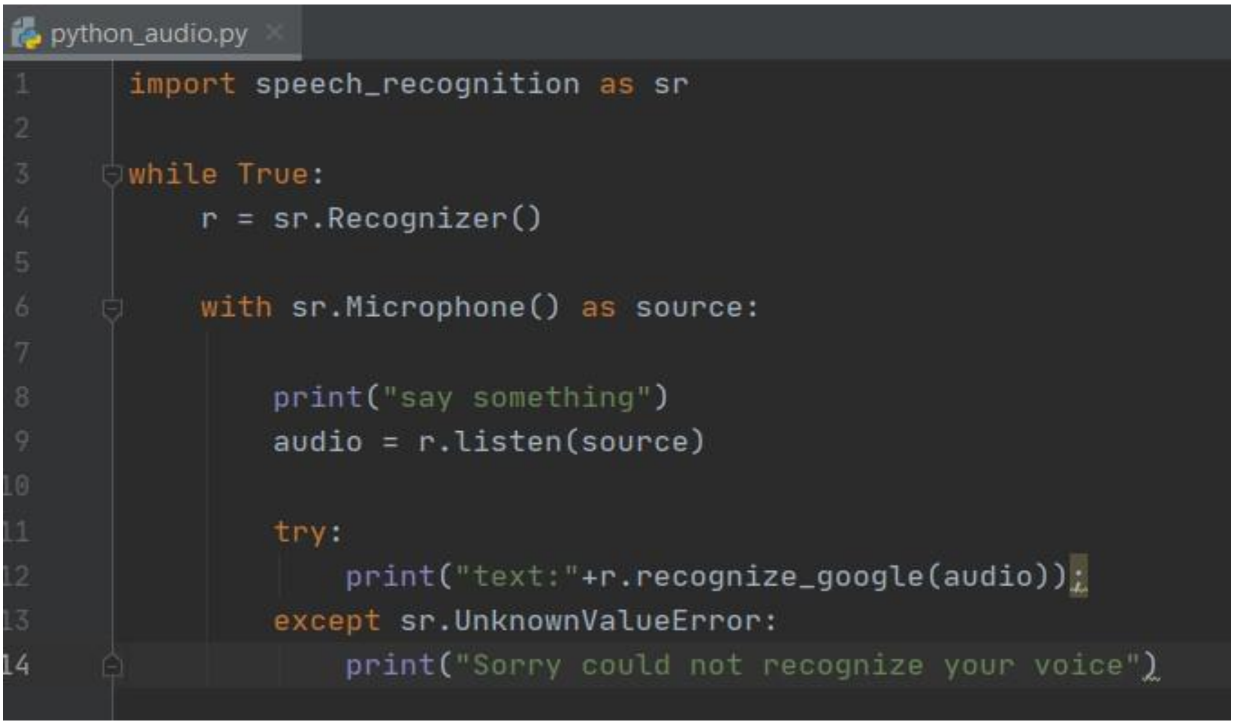
Integration Testing- It is Described as a form of checking out wherein software program modules are incorporated logically and examined as a set. A regular software program task consists of multiple software program modules, coded via distinct programmers. The purpose of this level of checking out is to show defects in the interaction between these software modules whilst they may be integrated.

System Testing – It is a stage of testing that validates the entire and absolutely integrated software product. The purpose of a system takes a look at is to evaluate the cease-to-quit gadget specs. Usually, the software program is handiest one element of a bigger pc-based totally machine. Ultimately, the software program is interfaced with other software program/hardware structures. System Testing is actually a series of various assessments whose sole cause is to work out the entire pc-based device.

Component Testing - Component is defined as a software testing type, wherein the trying out is completed on each individual element separately without integrating with different additives. It's additionally known as Module Testing whilst it's miles viewed from an structure perspective. Component Testing is likewise referred to as Unit Testing, Program Testing or Module Testing.

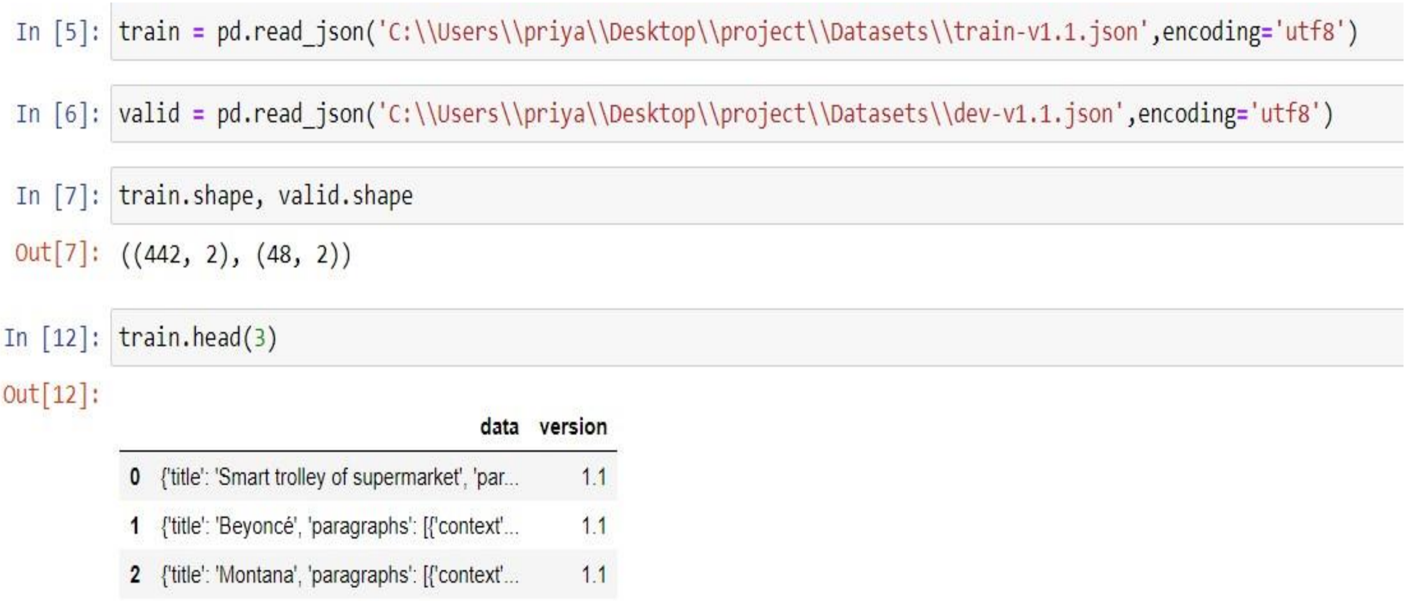
2.3.2 Implementation

Implement the speech to text



```python
import speech_recognition as sr

while True:
    r = sr.Recognizer()

    with sr.Microphone() as source:

        print("say something")
        audio = r.listen(source)

        try:
            print("text:"+r.recognize_google(audio));
        except sr.UnknownValueError:
            print("Sorry could not recognize your voice")
```

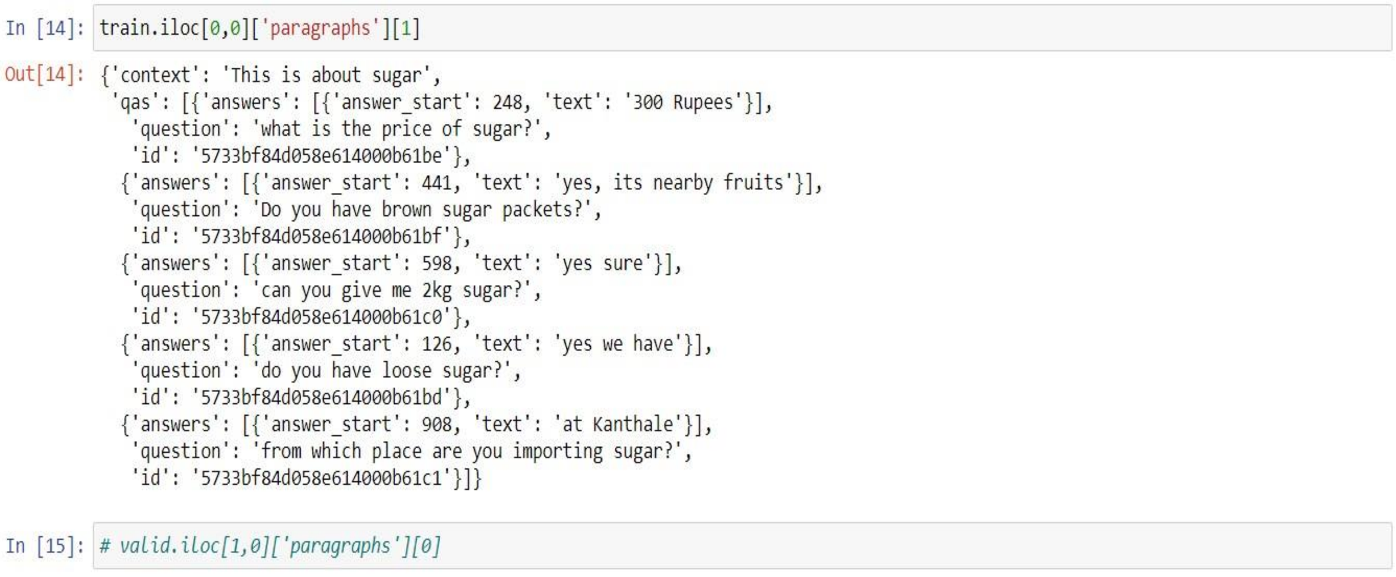
Figure 2.3.2.1 speech to text



```
In [5]: train = pd.read_json('C:\\Users\\priya\\Desktop\\project\\Datasets\\train-v1.1.json',encoding='utf8')

In [6]: valid = pd.read_json('C:\\Users\\priya\\Desktop\\project\\Datasets\\dev-v1.1.json',encoding='utf8')

In [7]: train.shape, valid.shape

Out[7]: ((442, 2), (48, 2))

In [12]: train.head(3)

Out[12]:
```

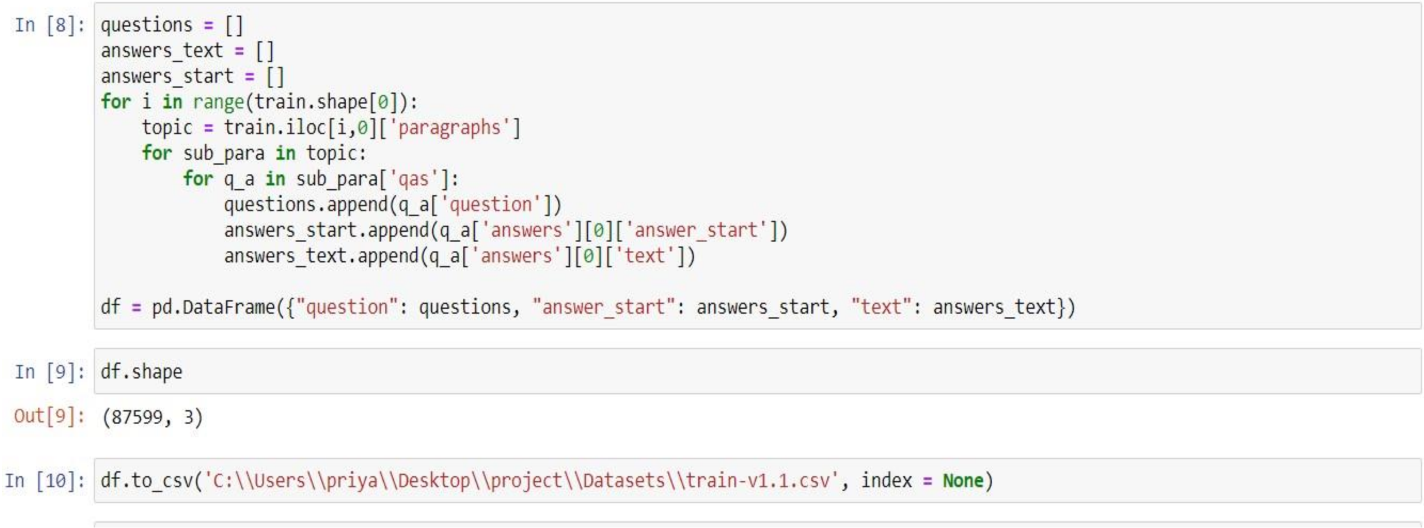
	data	version
0	{'title': 'Smart trolley of supermarket', 'par...	1.1
1	{'title': 'Beyoncé', 'paragraphs': [{'context'...	1.1
2	{'title': 'Montana', 'paragraphs': [{'context'...	1.1

Figure 2.3.2.2 convert json to pandas dataframe



```
In [14]: train.iloc[0,0]['paragraphs'][1]

Out[14]: {'context': 'This is about sugar',
          'qas': [{'answers': [{'answer_start': 248, 'text': '300 Rupees'}],
            'question': 'what is the price of sugar?',
            'id': '5733bf84d058e614000b61be'},
           {'answers': [{'answer_start': 441, 'text': 'yes, its nearby fruits'}],
            'question': 'Do you have brown sugar packets?',
            'id': '5733bf84d058e614000b61bf'},
           {'answers': [{'answer_start': 598, 'text': 'yes sure'}],
            'question': 'can you give me 2kg sugar?',
            'id': '5733bf84d058e614000b61c0'},
           {'answers': [{'answer_start': 126, 'text': 'yes we have'}],
            'question': 'do you have loose sugar?',
            'id': '5733bf84d058e614000b61bd'},
           {'answers': [{'answer_start': 908, 'text': 'at Kanthale'}],
            'question': 'from which place are you importing sugar?',
            'id': '5733bf84d058e614000b61c1'}]}

In [15]: # valid.iloc[1,0]['paragraphs'][0]
```

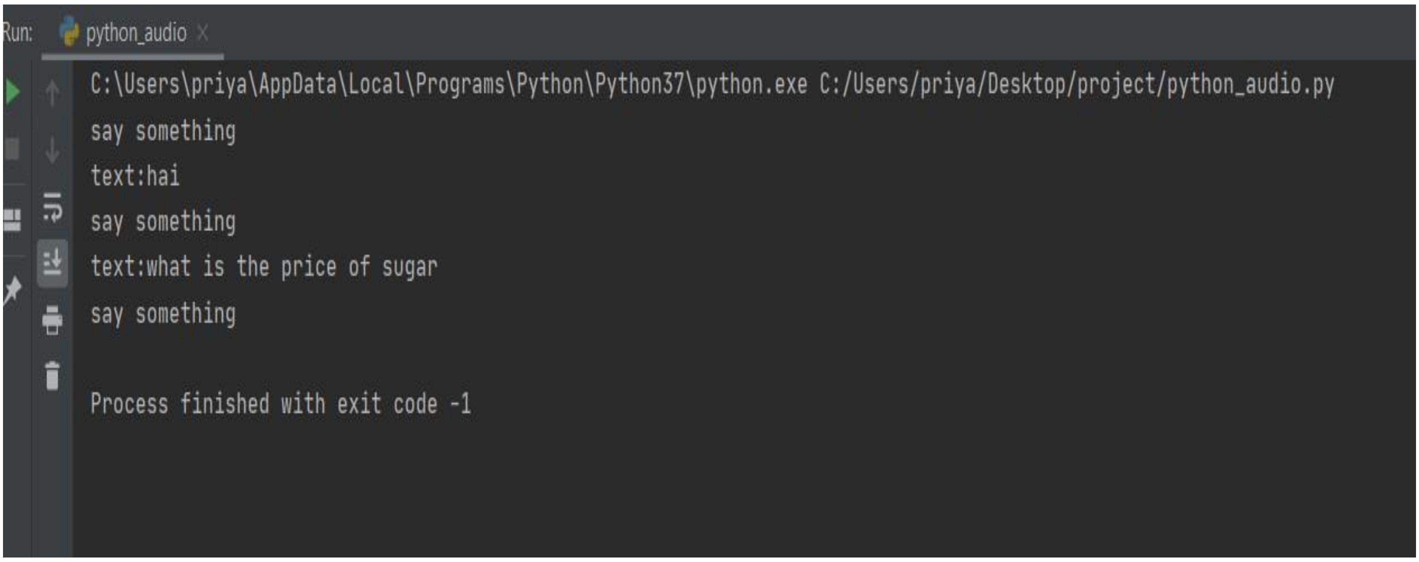
Figure 2.3.2.3   convert json to pandas dataframe



```
In [8]: questions = []
        answers_text = []
        answers_start = []
        for i in range(train.shape[0]):
            topic = train.iloc[i,0]['paragraphs']
            for sub_para in topic:
                for q_a in sub_para['qas']:
                    questions.append(q_a['question'])
                    answers_start.append(q_a['answers'][0]['answer_start'])
                    answers_text.append(q_a['answers'][0]['text'])

        df = pd.DataFrame({"question": questions, "answer_start": answers_start, "text": answers_text})

In [9]: df.shape

Out[9]: (87599, 3)

In [10]: df.to_csv('C:\\Users\\priya\\Desktop\\project\\Datasets\\train-v1.1.csv', index = None)
```

Figure 2.3.2.4   convert json to pandas dataframe

3. Results and Discussion



```
C:\Users\priya\AppData\Local\Programs\Python\Python37\python.exe C:/Users/priya/Desktop/project/python_audio.py
say something
text:hai
say something
text:what is the price of sugar
say something

Process finished with exit code -1
```

Figure 3.1 speech to text

This figure about speech to text

4.DESCRIPTION OF PERSONAL AND FACILITIES

Member	Component	Task
R.Priyanka	Voice Assistant	• Customer asks the questions using microphone • It will check the queries with answers already implemented in the database • If the customer ask hard questions in hard words. It will checked in dictionary • Then it will replied in Text.

Table: 4.1 work break-down table

Conclusion

The speech to text conversion may additionally seem effective and efficient to its users if it produces natural speech and by making several modifications to it. This device is beneficial for deaf and dumb humans to Interact with the other peoples from society. Speech to Text synthesis is important research and alertness location within the discipline of multimedia interfaces. A database has been comprised of the numerous domain phrases and syllables. The desired speech is produced with the aid of the Concatenative speech synthesis method. Speech synthesis is nice for individuals who are visually handicapped. The system offers the input facts from mice in the shape of voice, then preprocessed that statistics & converted into text format displayed on PC. The increasing diffusion of smart speakers illustrates the adoption of speech interaction based on voice assistants in private in addition to organizational contexts. However, while voice assistants allow users to perform tasks in a one-of-a-kind manner, those obligations may want to typically also be solved the usage of conventional user interfaces. This raises the question what drives users to choose speech interaction over other modes of interplay with information systems. We examined unmonitored cross-language representations of speech learned from the raw the waveshape. We show that pretraining in multiple languages on data enhances both over monolingual data pre-training as well as prior work, with the greatest changes to languages with low resources.

By gathering questions from various domains and checking the correct answers, the proposed solution of an open domain question answering framework utilizing NLP techniques and web snippets is done. After pre-processing, each query is passed through the Google search engine to retrieve the web pages. Snippets were collected from those web pages using the indexing technique and are rendered for local disk storage. The snippet usually contains statements about that document that provide a brief introduction. Often the fragments may not be relevant to the topic in topic. NLPtechnique is then used to describe the precise answer to the given question. The following table shows the results of our opendomain answering system for questions.

5. References

[1] Emi Moriuchi, "*OKAY, Google !: An empirical study on voice assistants on consumer engagement and loyalty*", January 2019. Available: https://www.researchgate.net/publication/330419587_Okay_Google_An_empirical_study_on_voice_assistants_on_consumer_engagement_and_loyalty

[2] thanyaphorn Lerlerdthaiyanupap, "Speech-based-dictionary" ,june 2008,Available: https://pdfs.semanticscholar.org/35ea/45dbd416576378bf709b517589896ef8e238.pdf

[3] Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad, "*Voice Recognition System: Speech-to-text*", November 2015. Available: https://www.researchgate.net/publication/304651244_VOICE_RECOGNITION_SYSTEM_SPEECH-TO-TEXT

[4] Ahmad Bisker Tarakji,Jian Xu and Juan A. Colmenares Iqbal Mohomed, "*Voice enabling mobile applications with UIVoice*", June 2018. Available: https://www.researchgate.net/publication/325436368_Voice_enabling_mobile_applications_with_UIVoice

[5] Aravind pai, "learn how to build your own speech-to-text model (using python)", Available: https://www.analyticsvidhya.com/blog/2019/07/learn-build-first-speech-to- text-model-python/, [Accessed: June 15 2019]

[6] Shruti Joshi, Aarti Kumari, Pooja Pai, and Saiesh Sangaonkar, "Voice Recognition System", Available: https://www.academia.edu/33497914/Voice_Recognition_System

[7] Pankaj Pathak, "*Speech Recognition Technology:* Application & future", International Journal of Adavnced Computer Research, December 2010, Available: https://www.researchgate.net/publication/289614337_Speech_Recognition_Technology_Application_future

[8] Ayushi Trivedi,Navya Pant, Pinal Shah,Simran Sonik and Supriya Agrawal , "*Speech to text and text to speech recognition systems-Areview*", PP 36-43,Available: https://www.iosrjournals.org/iosr-jce/papers/Vol20-issue2/Version- 1/E2002013643.pdf

[9] Eric hal schwartz, "The Decade of Voice Assistant Revolution " ,Available: https://voicebot.ai/2019/12/31/the-decade-of-voice-assistant-revolution/, [Accessed: December 31,2019]

[10] Jimcymol James , "*A Mobile Application for Voice Enabled Virtual Bot*", International Journal of Applied Engineering Research, Available: https://www.researchgate.net/publication/333893369_A_Mobile_Application_for_Voice_Enabled_Virtual_Bot

[11] **J SIMPON, "5 Best Speech-to-text APIs",Available:** https://nordicapis.com/5-best-speech-to-text-apis/, [Accessed:February 20th 2020]

[12] Venkatesh C.R, "How to add voice search to your Mobile app" Available: https://www.business2community.com/seo/how-to-add-voice-search-to-your-mobile- app-02250123 [Accessed: October 17,2019]

[13] Sentiance, "What voice assistants can learn from motion", Available: https://www.sentiance.com/2018/07/11/motion-voice-assistants/, [Accessed: July 11 2018]

[14] Gaikwad Vijayendra Sanjay, Kundur Ajinkya Sham, K, amble Sanket Mohan, Hulbutti Akash Huchcheshwar, Thorve Shubham Prakash, "DICTIONARY APPLICATION WITH SPEECH RECOGNITION AND SPEECH SYNTHESIS", January-February 2018, Available: https://www.ijarcs.info/index.php/Ijarcs/article/viewFile/5155/4446

[15] Andreas, J., Rohrbach, M., Darrell, T., and Klein, D. Neural module networks. CVPR, pp. 39–48, 2015. URL https: //arxiv.org/abs/1511.02799.

[16] Jacob Andreas, Marcus Rohrbach ,Trevor Darrell and Dan Klein, "Learning to Compose Neural Networks for Question Answering" https://www.aclweb.org/anthology/N16-1181.pdf

[17] Alexis Conneau, Douwe Kiela, Holger Schwenk, Lo¨ıc Barrault, Antoine Bordes, "Supervised Learning of Universal Sentence Representations from Natural Language Inference Data" https://arxiv.org/pdf/1705.02364.pdf

[18] Ramon Sanabria, Ozan Caglayan, Shruti Palaskar, Loïc Barrault, "How2: A Large-scale Dataset for Multimodal Language Understanding" https://arxiv.org/pdf/1811.00347.pdf

[19] Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, Noah A. Smith," Recurrent Neural Network Grammars" https://www.aclweb.org/anthology/N16-1024.pdf

[20] Thuy Vu, D. Stott Parker, "K-Embeddings: Learning Conceptual Embeddings for Words using Context" https://www.aclweb.org/anthology/N16-1151.pdf

[21] Mercedes Garc ´ıa-Mart ´ınez, Lo ¨ıc Barrault, and Fethi Bougares, "Neural Machine Translation By Generating Multiple Linguistic Factors" https://arxiv.org/pdf/1712.01821.pdf

[22] Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, Michael Auli, “Unsupervised Cross-lingual Representation Learning for Speech Recognition” https://arxiv.org/pdf/2006.13979.pdf

[23] Alexis Conneau, German Kruszewski, Guillaume Lample, Loïc Barrault, Marco Baroni, “What you can cram into a single $&!#* vector: Probing sentence embeddings for linguistic properties” https://arxiv.org/pdf/1805.01070.pdf

[24] Dr. Kshama V. Kulhalli, Dr.Kotrappa Sirbi, Mr. Abhijit J. Patankar, “Personal Assistant with Voice Recognition Intelligence” https://www.ripublication.com/irph/ijert_spl17/ijertv10n1spl_80.pdf

[25] Alex Mari, “Voice Commerce: Understanding shopping-related voice assistants and their effect on brands” https://www.researchgate.net/publication/336363485_Voice_Commerce_Understanding_shopping-related_voice_assistants_and_their_effect_on_brands

[26] Jun Suzuki and Masaaki Nagata, “Right-truncatable Neural Word Embeddings” https://www.aclweb.org/anthology/N16-1135.pdf

[27] Ryan Lowe , Abhinav Gupta , Jakob Foerster, Douwe Kiela, Joelle Pineau, “ON THE INTERACTION BETWEEN SUPERVISION AND SELF-PLAY IN EMERGENT COMMUNICATION” https://arxiv.org/pdf/2002.01093.pdf

[28] Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad, “Voice Recognition system: speech to text”, https://www.researchgate.net/publication/304651244_VOICE_RECOGNITION_SYSTEM_SPEECH-TO-TEXT

[29] Miss.Prachi Khilari, Prof. Bhope V. P.2, “A REVIEW ON SPEECH TO TEXT CONVERSION METHODS” http://ijarcet.org/wp-content/uploads/IJARCET-VOL-4-ISSUE-7-3067-3072.pdf

[30] Dhanush Kumar S, Lavanya S, Madhumita G, Mercy Rajaselvi V, ” Journal of Speech to Text Conversion” https://www.ijariit.com/manuscripts/v4i2/V4I2-1429.pdf

[31] Christine Rzepka , “Examining the use of Voice Assistants: A Value- focused Thinking Approach” https://www.researchgate.net/publication/335557789_Examining_the_Use_of_Voice_Assistants_A_Value-Focused_Thinking_Approach

[32] Maheshwari K B , Nandhinipriya J, Shantha Lakshmi M, Menaha R, “Question answering system for voice based search using NLP techniques and web snippets” https://www.academia.edu/38234298/QUESTION_ANSWERING_SYSTEM_FOR_VOICE_BASED_SEARCH_USING_NLP_TECHNIQUES_AND_WEB_SNIPPETS_pdf