ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 1, January 2023

A Study on Transpliing Speech into Python Code

[1] Anirudh Pandita, [2] Anusha V Salimath, [3] Anasuya Dutta

[1] [2] [3] Student, NHCE, Bengaluru, Karnataka, India.

Corresponding Author Email: [1] kppkanu@gmail.com, [2] anushasalimath@gmail.com, [3] duttaanasuya1412@gmail.com

Abstract— ***The world we live in today has seen major improvements in the field of technology. Most of the task which required maximum human effort is now being done with the help of machines. One such area is the processing of human languages. Recent studies have come up with solution which only requires one's voice to achieve the targeted output. But there is one such application which can be one of the many possible problems out there. As we know this entire thing is possible with the help of our developers and they too tend to get exhausted after some time. The concept of speech recognition can be used to propose a solution using natural language processing. This combined with a special type of parser to convert human transcription into a language such as python and many more popular languages out there used by everyone.***

Index Terms - Speech Recognition, Natural Language Processing, Recurrent Neural Network.

I. PROBLEM STATEMENT

An average developer tends to work for around 6-8 hours per day. It may be because of their job description or it's may be there project which is keeping them up. Often times, it becomes tedious to code for such a long time and can show effects such as carpel tunnel and so on. One possible solution to this problem is to have a helping assistant which is voice enabled and user friendly which can be used to ease out the work load on the developer. The basic idea is to have a desktop application which will listen to specific voice. In this case, it's going to be normal human speech. The idea is to convert the human speech into transcription which can be done used to achieve the desired output. The choice of conversion is python. Python, being a user friendly language, will be easy to convert the human speech to. Once the conversion takes place, we can automate the task of writing the code to any one of the existing editors such as VS Code or PyCharm. By having the ability to work with the existing editors, allows developer to not install separate ides and will also be able to use the existing features as well. Along with human transcription, there will be system where users will be sharing their codes to everyone such that it is accessible to all. This system can be one of the possible solutions out of the ones currently available ones. Let's look into some of the technologies used mainly for conversion of human speech.

II. TERMINOLOGIES

1. Natural Language Processing

Natural language Processing refers to the branch of AI that gives the machines the ability to read, which understands & derives the meaning from the human language. It combines the field of linguistics & computer science to decipher language structure guidelines. It can make models which can comprehend, break down and separate significant details from text to speech.

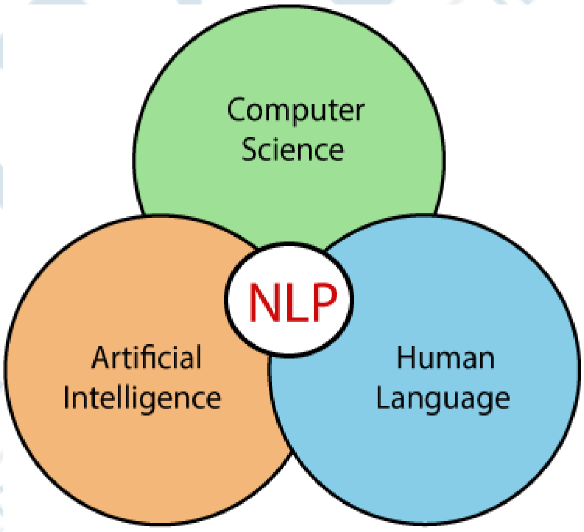

Fig 1. Natural language processing

As we know everyday people interact with each other through public media such as social networking sites, blogs and other social Medias which transfers huge quantity of free data that available to each other. This information are extremely useful in understanding the human behavior & habits. This data/information is utilized by the data analysts and machine learning experts to give the machines the ability to enact the human behavior & linguistic and it also helps in manpower & time as it does need to always have an individual person to be present to complete the other end.

NLP Tasks - It is extremely challenging to create software that reliably ascertains the intended meaning of text or voice data since human language is rife with ambiguity. NLP is used every day in seemingly normal & insignificant situations like:

Autocorrect: Helps in spelling a word correctly.

Plagiarism Checker: helps in detection of similar context in an article or thesis by searching through the web & find the published documents.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 1, January 2023

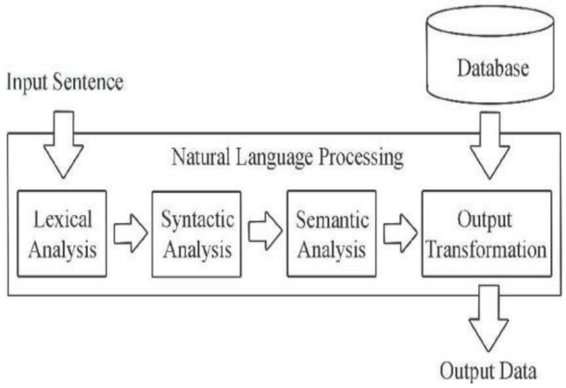

Fig 2. Sematic of Natural Language Processing

A General Question arise which is how does all work. The following shows the steps of NLP:

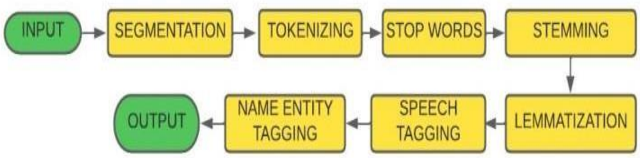

Fig 3. Working of NLP

Segmentation: It breaks the entire document into its constituent's sentences; this can be done segmenting the article along its punctuations like full stops and commas.

Tokenizing: It helps the algorithm to understand the sentences given in the document by breaking down the sentences given in the documents into tokens and store them.

Stop words: It makes the learning process faster by getting rid of non-essentials words like "are" , "and" , "the" , "was" , "in" and etc. which do not add meaning in the sentences given in the document.

Stemming: It helps the machine to understand the prefixes & suffixes which are added extra in the sentences.

Lemmatization: This help in identifying base words for different word tenses such as "mood" , "gender", etc.

Speech Tagging: It explains the concepts of "noun", "verb", "article", "prepositions" & other parts of the sentences to the machine.

Name Entity Tagging: It helps in introducing the machine with pop culture, famous personalities, & location.

Some of the commonly used application of natural language processing in the current time is: Google Assistant, Grammely, Text Summarization etc.

2. Recurrent Neural Network

Recurrent Neural network belong to a class of Artificial Neural Network which works on the simple concept of adding output to its input. This type of behavior makes the neural network to have some amount of memory stored in it. This allows the network to work on complex deep learning task such as handwriting recognition, speech recognition and so on. Due to its short term memory capacity, it is also used for sentence completion and language translation.

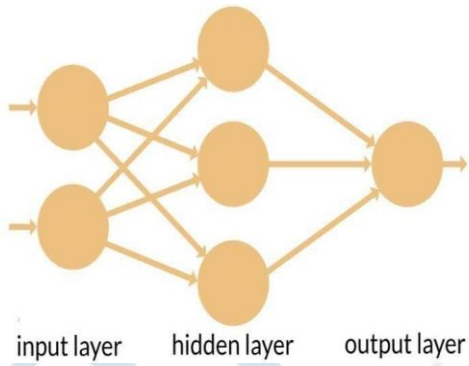

Fig 4. Feed Forward Neural Network Architecture

Based on Neural network design, recurrent neural network connects the output of all neurons to its input neuron. RNN has the ability to convert independent activations into dependent activations with the help of same weights and bias to all of its input layers. Therefore these three layers are joined such that the weights and the bias of its input layer be the same, hence forming the simple recurrent neural network. The calculation of the current state in which the recurrent neural network is present is given by this formula.

ht = f(ht-1,xt)

Where ht defines the current state, ht-1 defines the previous state; xt defines the input state of the hidden layer. RNN's generally use the tanh activation function whose range lies between [-1, 1].

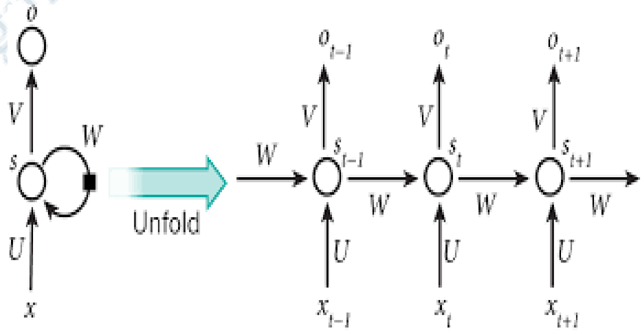

Fig 5. Unfolded view of a Recurrent neural network

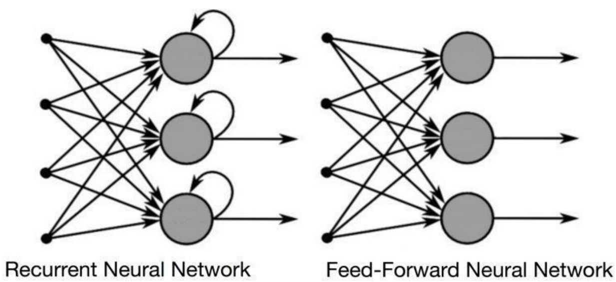

Fig 6. Recurrent Neural Network vs. Feed Forward Network

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 1, January 2023

Although having memory capacity, RNN's tend to forget the most critical sequence when learning for long sequence of data (also known as the vanishing gradient problem). Therefore making it ineffective for long series data set. To tackle this scenario, the concept of LSTM (Long Short term memory) was introduced.

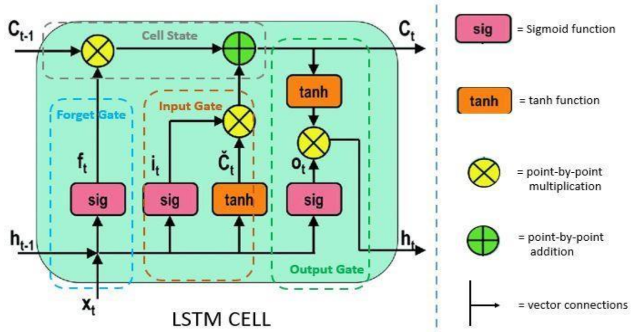

Fig 7. Long Short Term Memory Cell

Long Short Term Memory cell are used mainly to solve the vanishing gradient problem by using the concept of gates. This gating mechanism is used to determine whether a particular piece of information needs to be remembered or not. The opening and closing of the gates signifies the act of writing and reading. Activation Function such as Tanh and sigmoid are primarily used. Tanh function regulates the output value flowing through the network in the range of [-1, 1] whereas sigmoid function values ranges from 0 to 1. If the value turns out to be zero, then the information is considered forgotten. It mainly constitutes of three gates. Input Gate: This gate performs operations to update the status of the cell. Forget Gate: It decides whether the information from the previous iteration is off any use in the current state. Output Gate: This gate is used to calculate the output of the next hidden state.

III. LITERATURE SURVEY

Speech-to-Text Conversion:

Speech to text is the process of converting normal human spoken language into written text in the form of a string which can be interpreted by the computer. Earlier used as a synonym to speech recognition, it is now widely used for the process of speech understanding. One of the most popular models which is widely used is the Hidden Markov Model.

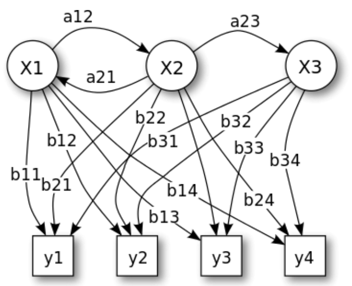

Fig 8. Probabilistic parameters of a HMM

The Hidden Markov model is a type of statistical model which follows the markov chain which states that the probability of each event depends on the state in which the model was in its previous iteration. This model has many applications in the field of thermodynamics, signal processing, speech tagging etc. This model depends on few of the parameters such as Recognition accuracy, Recognition Speed. The speech must undergo preprocessing of the speech signals which helps to remove unwanted waveforms from the speech signal to extract two kinds of acoustic features which are the Mel Frequency Cep- strum coefficients (MFCC) and Linear Predictive Coding Coefficients (LPCC)

Speech Recognition in Noisy Surrounding:

Due to the hustling nature of our surroundings, there are sounds such as noise's available along with normal sound. This makes the job of speech recognition little difficult as it is unable to differentiate from the both. Hence the need of an noise filter concept was introduced.

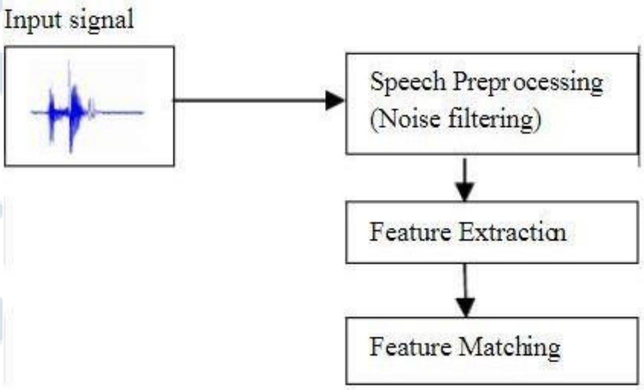

Fig 9. Sample of Features Extraction with noise filter

The traditional Approach of noise removal algorithms mainly used to find the frequencies that have higher levels of background noise and simply deleting it. This approach uses static filter and best suited for deterministic signals, making it ineffective in varying conditions. The industry standard used to remove noise is the Wiener filter which is used widely in technologies such as hearing aids, noise cancelling microphones in smartphones and etc. Further improvements were done by combining two complementary models i.e.: Repeating Weiner Filter and characteristics abandon model.

Speech to text with effective understanding and summarization:

During long hours of transcription of human speech, It is necessary that the output of the model should actually make sense when read. Certain measure have to be taken care of such that model is able to make an effective transcription which is easy to understand.

IFERP connecting engineers...developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 1, January 2023

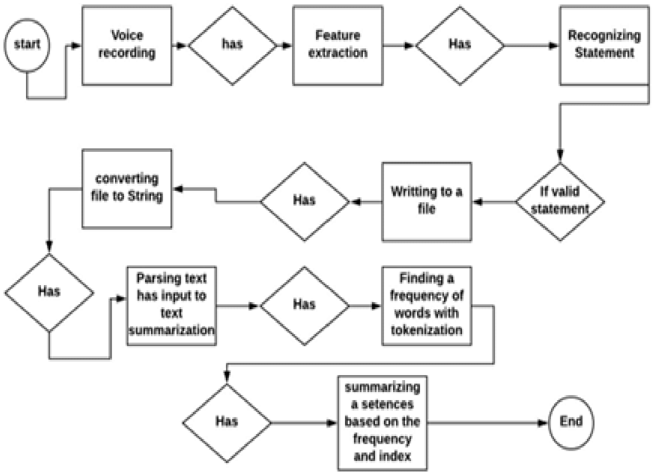

Fig 10. Proposed work for effective transcription and summarization

The above proposed system has followed the normal process of speech recognition and feature extraction. The basic idea is to introduce a pause delay of 2e+6 microseconds or more. This delay allows adding punctuation marks such as period (.), question mark (?) and so on. If the time period exceeds the system wait for the second input for validation.

The type of punctuation used depends upon the type of conjunction the model is predicting. For example Wh-Questions have to be terminated with the question mark (?) Once the formatting is finish, the output transcription can be done used for summarization. But the problem of synonyms still poses an issue. To overcome this, words with less significant is removed. This is done by keeping the words having good number of repeated occurrence which is done used to process the important sentences using a ranked model.

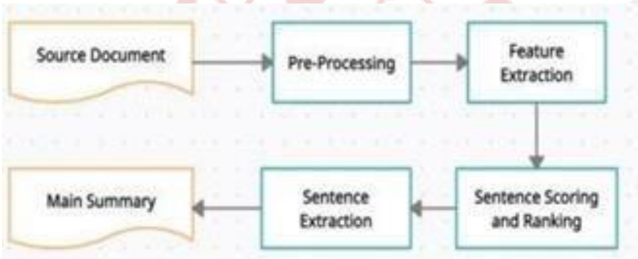

Fig 11. General Working of summarization

Automatic summarization, co-reference analysis, among other activities, serve as subtasks that help solve bigger problems. NLP is a topic of discussion today due to its many uses and recent advancements, despite the fact that the phrase wasn't even coined until the late 1940s.

Therefore, learning about the development of NLP, its history, and some current initiatives utilizing it would be intriguing. These features are the subject of this paper's second goal. The third goal of this essay is to discuss datasets, methodologies, assessment standards, and other NLP problems. The remainder of this essay is structured as follows. It discusses the first goal while referencing numerous key NLP and NLG concepts. It covers the background of NLP, its applications, and a rundown of current advancements in an unnumbered footnote on the first page.

IV. EXISTING SYSTEM

The existing system developed uses keyword identification to scan a collection of ready-made code from a database. This system comes along with a workspace which is plain and unlike the IDEs of today.

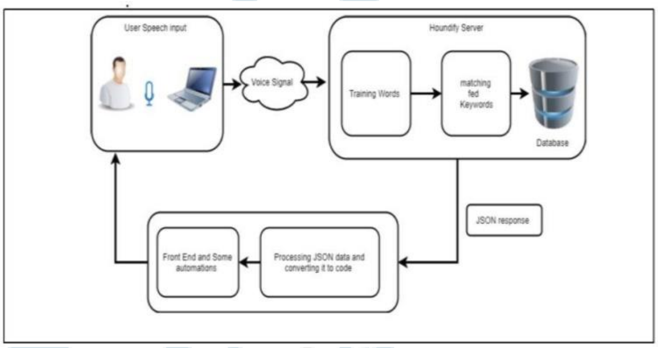

Fig 12. Existing System Architecture

The system consists of following modules:

Designing the User Workspace

Key word recognition is built into editors and notepads. The subsequent step is to retrieve the relevant code using these keywords. Through user voice commands, the resultant codecan be compiled and run.

Recognition of voice

Through the use of a microphone, the system will identify the user's speech and print the associated keywords on the editor and terminal.

Using Editor and Matching Keywords

If the comparison yields a result of zero, the recognized keywords will be printed on the editor. The recognized keywords can be compared with the keywords that are stored in the database.

Assemble and run the code

Through user voice commands, the printed code can be assembled and run.

The editor which comes along with this system has lack of features and functionalities compared to the existing Integrated Development Environment/ Code-Editors. So, the developers don’t prefer to use this system.

V. PROPOSED SYSTEM

The proposed system involves developing a desktop application, which will not only retrieve the code snippets for certain a keyword. But, also will be able to parse single line transcription from the voice input. A dedicated transcription parser will be developed which will be able to parse sentences into equivalent python code.

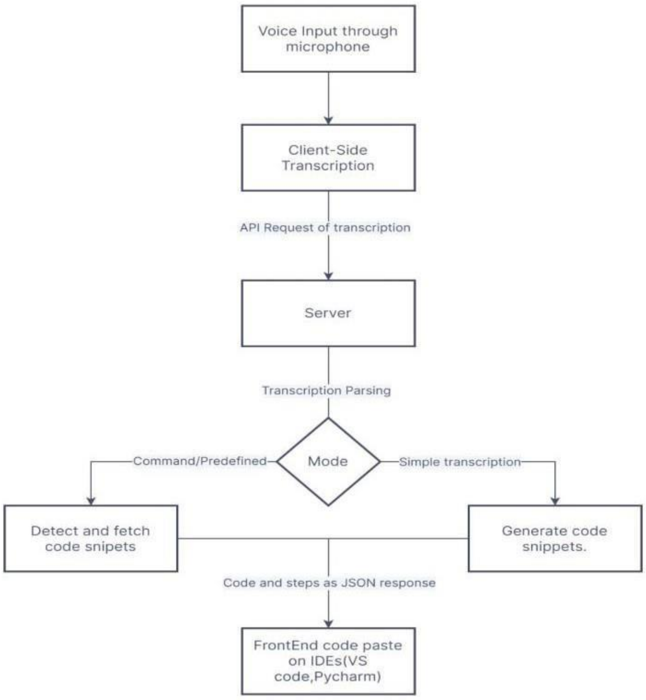

Fig.13. Proposed System Flowchart

The desktop application will not act as an editor but will be a helping assistant to the existing IDEs such as PyCharm and Code-Editors such as VS Code. The Speech-To-Text model will be trained to quickly identify some of the commonly used python keywords such as: input, def., for, while, if, else-if, class and so on. A particular syntax will be followed for each keyword so that the parser can parse the sentence correctly. The client and the server will be connected using web socket as it allows bi-directional communication between the two bodies. The server's main goal is to properly identify which sentences are to be parsed and which sentences to be scanned for code snippets in the database. The result generated by the server will consist of python code and steps for the Client side, which can be used to paste the python code on the desired IDEs/ Code-Editor.

VI. CONCLUSION

We have seen the existing techniques used to convert human speech into transcription and have also come up with an optimal solution over the existing system using natural language processing and web sockets. This solution can be used to help in lowering the overall workload on a developer and can boost productivity as well.

VII. ACKNOWLEDGEMENTS

We would like to thank Ms. Pramilarani for guiding us through this paper and assisting us in developing a workable solution by challenging us to go outside our comfort zones. We appreciate the opportunity to work on this paper and bring it to life.

REFERENCES

[1] M.A Jawale, A.B. Pawar, D.N Kyatanavar, Smart Python Coding through Voice Recongnitipn, International Journel of Innovative Technology and Exploring Engineering (2019)
[2] Vinnarasu A, Deepa V. Josin, Speech To Text Conversion and Summarization For Effective Understanding, International Journal of Electrical and Computer Engineering (IJECE) (2019)
[3] Chethan S, Amrutha C., Automatic Speech Recognition, International Research Journal of Engineering and Technology (IRJET) (2020)
[4] Shivangi Nagdewani, Ashika Jain , Methods For Speech-To-Text, International Research Journal of Engineering and Technology (IRJET) 5 May 2020
[5] Santosh Kumar Behera, Mitali M Nayak, Natural Language Processing For Text and Speech Processing, International Journal of Advanced Research in Engineering and Technology (IJARET) (2020)
[6] Chunling Tang 1*, Min Li 1 , Speech Recognition in High Noise Environment, Ekoloji 28(107) (2019)
[7] R.Thiruvengatanadhan, Speech Recognition using AANN, International Journal of Innovations in Engineering and Technology (IJIET) (2019)
[8] Aditya Amberkar, Gaurav Deshmukh, Speech Recognition using Recurrent Neural Networks, IEEE International Conference (2018)
[9] Babu Pandipati, Dr. R.Praveen Sam, Speech to text Conversion using Deep Learning Neural Net Methods, Turkish Journal of Computer and Mathematics Education (2021)
[10] Vineet Vashisht, Aditya Kumar Pandey, and Satya Prakash Yadav, Speech Recognition using Machine Learning, IEIE Transactions on Smart Processing and Computing (2021)
[11] Amritpreet Kaur, Rohit Sachdeva, Amitoj Singh, Classification approaches for automatic speech recognition system, CRC Press (2021)
[12] Pete Warden , Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition published by Google.inc (2018)
[13] G. Vijay Kumar, Arvind Yadav, B. Vishnupriya, Text Summarizing Using NLP, IOS Press (2021)
[14] Ujwalla Gawande,*, Speech Recognition in Noisy Environment, Issues and Challenges:A Review, International Conference on Soft-Computing and Network Security (2015)