

## TRANSCRIPTION OF CONFERENCES

**Anika Babu<sup>1</sup>, Indhu P<sup>2</sup>, Revathy Surendran<sup>3</sup>, Dr Surekha Mariam Varghese<sup>4</sup>**

*<sup>1,2,3</sup>Graduate Student, Dept of Computer Science and Engineering, Mar Athanasius College of Engineering  
Kothamangalam, Kerala, India*

*<sup>4</sup>Head of Department, Dept of Computer Science and Engineering, Mar Athanasius College of Engineering,  
Kothamangalam, Kerala, India*

-----\*\*\*-----

**Abstract** - A conference can be defined as any exchange of knowledge, information or a discussion in a prearranged meeting with formal agenda. The formal, professional nature of conferences makes it important to maintain an accurate record of the discussions that take place during its course. This can either be done by someone in the conference who is dedicated to registering every valuable point expressed by any one of the participants or by creating an audio recording of the conference and getting it professionally transcribed. Audio recording and transcription has inherent advantage over manual registering as it reduces the chances of human errors. Intelligent word to word transcription has become an integral part of conference transcription. The techniques for transcription of conference begin with audio recording followed by speech to text conversion. The transcribed speech is then summarized using naïve summarizer that is based on word frequency score. The final output contains the summarized text along with Wikipedia URL links to relevant words.

**Key Words:** NLTK, NLP, Data mining, NLU, RAKE, Frequency scoring

### 1. INTRODUCTION

In the era of internet, plethoras of online information are freely available for readers in the form of e-Newspapers, journal articles, technical reports, transcription dialogues etc. There are huge number of documents available in above digital media and extracting only relevant information from all these media is a tedious job for the individuals in stipulated time. There is a need for an automated system that can extract only relevant information from these data sources.

It is vital for business and employers to keep minutes of their meetings and conferences. Conference transcription provides support in transcribing the minutes of the meetings. As long as conferences and meetings serve as powerful means of communications, the importance of these services will always be high. Keeping minutes for meetings is thus important. It doesn't take much to figure out that keeping minutes gives a verifiable record of the meeting where crucial information have been shared. Conference transcription has the following benefits: Accurate records of the meeting can be maintained, and this strategy is much better than recording just the minutes of the meetings. This transcripts help to review and sift out the key points, Transcribing the minutes of the meeting helps to quickly obtain the actionable data of a meeting or a conference. The transcribed report can be shared among those who were not able to attend the meeting.

#### 1.1 Problem Statement

The transcription of conferences has become very essential criterion for the growth of an organization since many crucial decisions are taken at those important meetings. These vital decisions are transcribed and used for future references. Since its a very significant task the service provider who handles the transcription service should be well aware of the nuances and should be capable of presenting all visions, exploits and verdicts in the precise way like the original and replica of the innovative meeting. The transcriptionist who transcribes the outcomes of meeting should be smart enough and full of ears to hear out and transcribe even the slightest of the slightest details that take place in the conversation. As far as the turnaround time, attentions to particulars, attitude and work ethics, meeting deadlines are concerned most of the transcription service providers offer fully committed transcription service. Since the discussions and slangs utilized in meetings are of intricate type and bears immense financial propositions in transcriptions. Hence the transcriptionist who is assigned with the job of transcribing meeting must be extremely proficient and tremendously knowledgeable in communication terms.

#### 1.2 Problem Justification

A conference can be defined as any exchange of knowledge, information or a discussion in a prearranged meeting with formal agenda. The formal, professional nature of conferences makes it important to maintain an accurate record of the discussions that take place during its course.

### 1.3 Main Objectives

Develop a transcription service which does not require a transcriptionist dedicated to transcript minutes of the conference. By recording the audio of conference, transcribe it professionally. Audio recording and transcription has inherent advantage over manual transcription as it reduces the chance of human errors.

## 2. PROPOSED SYSTEM

The proposed system use techniques for transcription of conference includes speech to text conversion, text summarization for extracting relevant sentences, include wikilink for further references. Google Cloud speech-to-text enables developers to convert audio to text by applying powerful neural network models in an easy-to use API. It can process real-time streaming audio using Google's machine learning technology. It can accurately transcribe proper nouns. Text summarization can be done using fuzzy inference systems. Summarization results are presented by text. It has advantages such as: a) the documents can be easily looked through; b) the part of the document that is interesting for users can be easily extracted; c) information extraction and retrieval techniques can be easily applied to the documents. This paper presents an automatic text summarization where the most appropriate sentences, words or phrases are automatically extracted from text to produce a summary. The document consisting of summarized contents will include Wikilink for further references. There would be some terms or words which are not familiar to everyone. Such terms can be technical words which have a definite definition and Wikipedia pages would be present. So Wikilinks are provided for easy understanding of summarized document.

## 3. SYSTEM DESIGN

The implementation of the proposed system is divided into mainly three phases. The first phase involves listening to the audio and converting it into a suitable audio file. The second phase receives the output of the first phase as input i.e., the audio converted-text file which is further processed to retrieve the summary. The final phase returns the summarized text along with the relevant keywords and their Wikipedia links.

### 3.1 Speech Recognition

The first and foremost step in speech recognition is to feed sound waves into a computer. Everybody knows, the sound is transmitted as waves but the computer knows only numbers. So it needs to be converted to numbers. Sound waves are one-dimensional. At every moment in time, they have a single value based on the height of the wave. To turn this sound wave into numbers, record the height of the wave at equally spaced points. This is called sampling. It takes a reading of so many words a second, and recording a number representing the height of the sound wave at that point in time.

### 3.2 Summarization

The machines have become smarter than people and can help them with every aspect of life. The technologies have reached to an extent where they can do all the tasks of human beings like household tasks, controlling home devices, making appointments etc. The field which makes these things happen is Machine Learning. Machine Learning train the machines with some data which makes it capable of acting when tested by the similar type of data. The machines have become capable of understanding human languages using Natural Language Processing. Natural language processing is a large subject matter ranging from generating summaries or detecting intent or sentiment to generating content from data which is been utilized in the coming sections.

At this phase the text file generated from conversion of speech is summarized .That is shorten the text document, in order to create a summary of the major points of the original document. The main idea of summarization is to find a subset of data which contains the "information" of the entire set. Such techniques are widely used in industry today. Search engines are an example; others include the summarization of documents, image collections, and videos. Document summarization tries to create a representative summary or abstract of the entire document, by finding the most informative sentences, while in image summarization the system finds the most representative and important images. For surveillance videos, one might want to extract the important events from the uneventful context.

A naive summarizer that works based on word frequency scoring is build using NLTK Python library isto b NLTK is natural language toolkit library. It is a platform for building Python programs to work with human languages. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum. In text summarizer, this library is used to remove stop words in English vocabulary and to convert these words to root forms.

### 3.3 Keyword Selection and Wikipedia URL Link

Once summary is obtained, pick out the relevant words spoken in the conference. In addition to that provide links to Wikipedia pages of the keywords which enable users to find out more on the keyword.

## 4. METHODOLOGY

Speech recognition is the ability of a machine or a program to identify words and phrases in spoken language and convert them to a machine-readable format. The following are the steps used for this purpose:

**Configure microphone:** For external microphones it is advisable to specify the microphone during the program to avoid any glitches.

**Set chunk size:** This involves specifying how many bytes of data has to be read at once.

**Set device ID to the selected microphone:** the device ID of the microphone that is used is specified in order to avoid ambiguity in case there are multiple microphones.

**Allow adjusting for ambient noise:** Since the surrounding noise varies adjust the energy threshold of recording of program a second or too so that it is adjusted according to external noise level.

**Speech to text translation:** This is done with the help of google speech recognition. It requires an active internet connection to work.

The audio converted text file is then further processed to retrieve the summary of major points of the original document. The naïve summarizer that works based on word frequency scoring is built. NLTK Python library is used to build this pipeline. The frequency scoring system can be implemented as follows:

**Reading from file:** This step involves reading from the file. Here it accepts a file path and returns the entire contents in memory, in order to pick up the relevant terms.

**Data cleanup:** Data cleaning is the process of cleaning and standardizing the data to make it ready for analysis. Most of times, there will be discrepancies in the captured data such as incorrect data formats, missing data, errors while capturing the data. This is an important step in any given data science project because the accuracy of the results depends heavily on the data used. Researches says that 80% of the time is spent working on data sanitization, collection, and normalization. Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data that is needed to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonymized or removed from the data entirely. In order to sanitize input, the different white space that includes tab, new line characters has to be stripped away. The goal is to replace any extra whitespace characters besides the one space after ending punctuation.

Stop word is a commonly used word (such as “the”, “a”, “an”, “in”) that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query. These words must be ignored as it takes up space in the database, or takes up valuable processing time. NLTK (Natural Language Toolkit) in Python has a list of stopwords stored in 16 different languages that could be found in the NLTK data directory. In the process of tokenization a list of all sentences and a unique list of all tokens are obtained.

**Scoring:** A list of unique sentences as well as a list of unique words is used to score the frequency of each word occurring in the story and use that to grade the sentences. Ranking is a central part of many information retrieval problems, such as document retrieval, collaborative filtering, sentiment analysis, and online advertising. Scoring and ranking is mainly done to see how important or relevant a particular sentence or word is in the context. Each element would be given a score followed by ranking the elements according to the score. The FreqDist function is another function that can be imported from NLTK. It accepts a list of tokens, like filtered word list, and returns a structure where each key is the word and each value is the number of times that word occurred. Then initialize a structure for making frequency maps, iterate over the sentences and increase their score based on the frequency of that particular word. The value of ranking will then contain key values of the sentence’s numeric position, and their score.

**Selection:** When scoring is complete, construct summary from the N highest scoring sentences where N is the desired length. Pass the sentences ranked that are generated with scoring function and the length of the summary requested which can be any default value if a value is not provided when launched. Make sure the length requested is not longer than the total number of sentences available. If it is, error it out. Sentence ranking data is analyzed and turned into a list of the numeric positions of the sentences based on rank given. Use this list of indexes in a list comprehension to put each sentence from the tokenized list into the final summary. Then sort the list of indexes so that sentences will appear in their organic order. The sentences are ordered according to their relevance in the context. Finally these values are joined together into a string and returned.

Extracting the relevant keyword is implemented using the Python library RAKE that stands for Rapid Automatic Keyword Extraction algorithm. It is a domain independent keyword extraction algorithm which tries to determine key phrases and words in a body of text by analysing the frequency of word appearance and its co-occurrence with other words in the text. Links to the Wikipedia pages of the keywords is also provided that helps users to get detailed description on it. This is done using the Python library *Wikipedia* that makes it easy to access and pass data from Wikipedia pages. Wikipedia-API is easy to use Python wrapper for Wikipedia's API. It supports extracting texts, sections, links, categories, translation etc. from Wikipedia. Thus for all the relevant words a hyperlink to its existing Wikipedia page is provided.

## 5. CONCLUSION

The proposed method would capture the audio of the conference or the meeting and it produces a text that contains the relevant sentences and phrases along with Wikipedia link. Data mining techniques in speech recognition helps in the areas of prediction, search, explanation, learning, and language understanding. A new class of learning systems can be created that can infer knowledge automatically from data. Effective techniques for mining speech, audio, and dialog data can impact numerous business and government applications. Such a system would be less time consuming when compared to the tedious task of someone listening and picking out the sentences. It does not require the aid of another person as the system summarizes on its own. The proposed method is operationally efficient comparing to the existing systems.

## REFERENCES

- [1] Napoletani D, Struppa DC, Sauer T, Berenstein CA, Walnut D, " Delay coordinates embeddings as a data mining tool for denoising speech signals".
- [2] Lee Begeja, Harris Drucker, David Gibbon<sup>1</sup>, Patrick Haffner, Zhu Liu, Bernard Renger, Behzad Shahraray, "Semantic Data Mining of Short Utterances", IEEE Trans. On Speech Audio Proc.: Special Issue on Data Mining of Speech, Audio and Dialog July 1, 2005.
- [3] Park, A.S. Glass, J.R., " Unsupervised Pattern Discovery in Speech", IEEE transaction on Audio, Speech, and Language processing- Jan. 2008.
- [4] Ying Shi; Weihua Song, "Speech emotion recognition based on data mining technology", IEEE transaction on Audio, Speech, and Language processing Jan. 2014
- [5] Dai Sheng-Hui, Lin Gang-Yong, Zhou Hua-Qing, " A Novel Method for Speech Data Mining", Journal of Software, Vol. 6, No. 1, January 2011.
- [6] Surendra Shetty, K.K Achar, "Audio Data Mining Using Multi-perceptron ANN" International Journal of Computer Science and Network Security, Vol. 8, No.10 October 2008.