

Smart Meeting Assistant

¹Samir Kapure , ²Shreya Mangade, ³Apoorva Dhardhar, ⁴Geetha Chillarge

¹ UG Student, ² UG Student, ³ UG Student, ⁴ Assistant Professor

¹ Department of Computer Engg,

¹ Marathwada Mitra Mandal's College of Engineering Pune, India.

Abstract: It often happens that the people participating in a meeting, in the hurry of jotting down the key points tend to miss out on important points that are being discussed in the meeting. Smart Meeting Assistant is being developed to solve this problem. It will assist the people by making a summary of the meeting. It will be possible to review what happened in the meeting after it ends. With the smart meeting assistant, the participants of the meeting can worry less about taking the notes and can focus more on what is more important, like building client relationships, brainstorming ideas or taking business decisions. The smart meeting assistant comes under the domain of Artificial Intelligence and Machine Learning. Natural Language Processing, Voice Recognition API can be implemented to develop this assistant. Forming a summary of the meeting can be done by text mining. The data that will be generated will be unstructured. Hence, determining the definite patterns and trends to examine a textual data is the biggest issue in text mining. Text mining helps in obtaining fascinating patterns from unstructured texts. Information retrieval, summarization, categorization, information extraction, categorization, clustering are different techniques of text mining.

Index Terms – Artificial Intelligence, Machine Learning, Voice Recognition, Natural Language Processing, Text mining, Summarization.

I. INTRODUCTION

Meetings have become an important part of our life. Moreover, it may happen that people attend more than one meeting in a day. It becomes tedious and overhead to remember all the important points, topics discussed in meetings. Thus, the agenda of Smart Meeting Assistant is to help people analyze their meetings in summarized format and get ready document of the same. This application will help the people by providing a summary of important topics, discussion occurred during the meeting. The system will take the voice of the participants as the input. [1] As the meeting will be recorded, speaker diarization will be done. After speech to text conversion, summarizing the meeting for further analysis of the key points being discussed during the meeting, text mining will be done and suitable document will get ready.

Thus, the assistant will help the participants of the meeting by summarizing the meeting and providing the summary in the document format and save their time required in taking notes during the meeting.

II. EXISTING SYSTEM

In traditional video conference systems, we need to remember everything that we discussed during conference. This involves pen and paper work. Users attending video call need to write down important topics, decisions, ideas to have them recorded with them. As paper work is involved, further-more it can cause important data loss. Missing attention on such important topics causes problems while decision making. Though some system provides mechanism of "Conference Recording", it becomes time consuming to play the recording and review the meeting.

Highlighting important points become overhead. The disadvantages of current system can be overcome in an automated, Artificial Intelligence based tool which will help users to review what happened in the meeting after it's ended. So, you can worry less about taking notes for later, and focus on what's really important in the moment – whether that's building client relationships, brainstorming ideas, business decisions

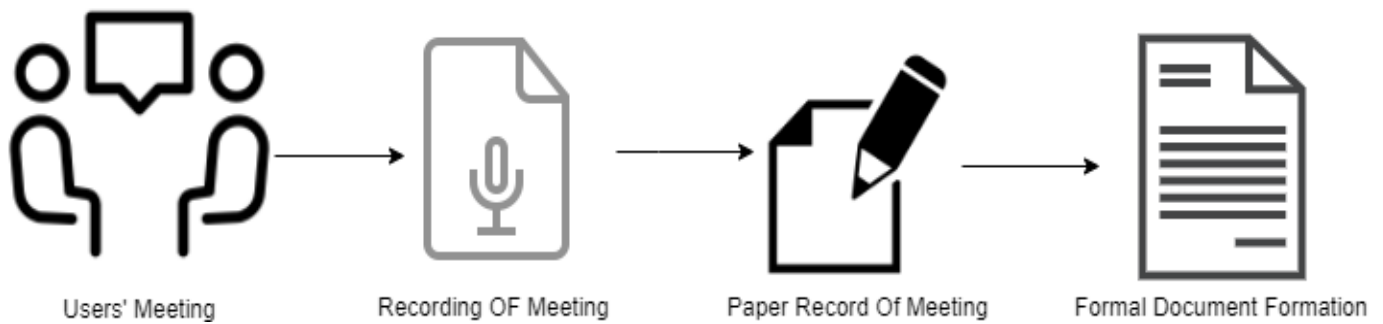


FIG. EXISTING SYSTEM

III.PROPOSED SYSTEM

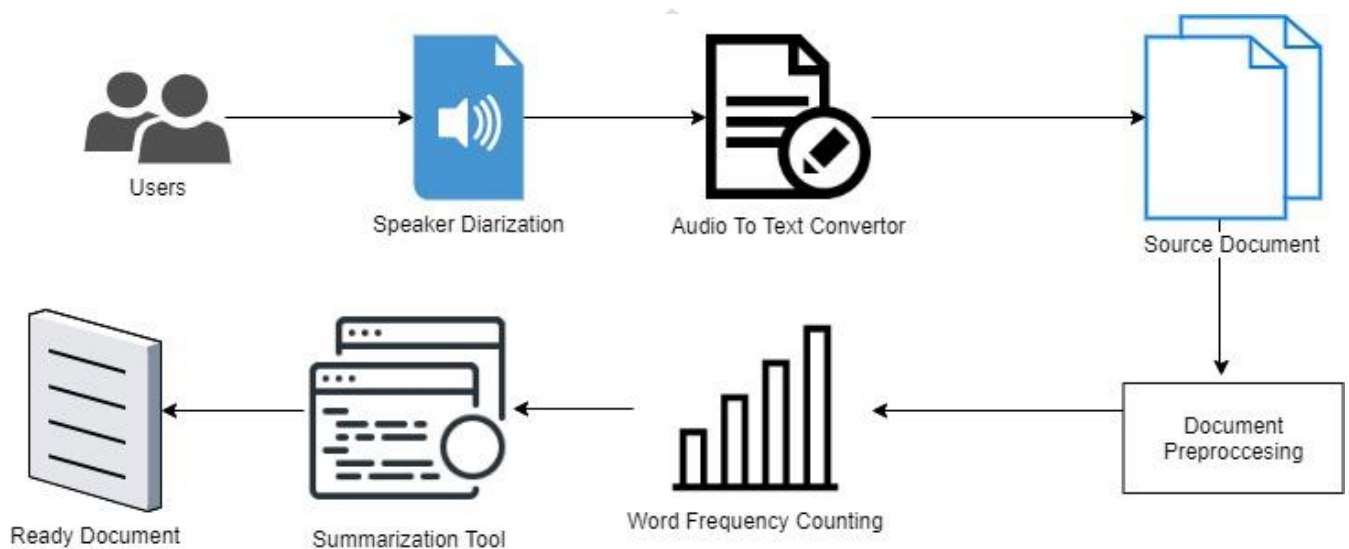


FIG.PROPOSED SYSTEM

Proposed System is an adept solution for problems faced in existing system. It is a Machine Learning/Artificial intelligence based system. It will help users to summarize the conversation occurred during meeting.

3.1 Working :

A. Voice To Text Conversion & Speaker Diarization :

System will start recording the conversation, once this action gets triggered by user. Voice API will play an important role. With the help mic, we can record the meeting. The main concern here is identifying users uniquely. API will help, to identify individual user by “speaker diarization” -. Speaker diarization is the process of identifying users using various voice parameters. Simultaneously voice input will be converted to text form. It will be available in form of text document.

B. Text Mining & Pre-processing of text document:

The following are the steps for text mining.

Step 1: Converting the unstructured or semi-structured data into structured data.

Step 2: Preprocessing and cleansing operations are implemented to detect and remove the anomalies. Cleansing operations remove the inconsistent and incorrect part if the data.

Step 3: Examining and inspecting the patterns using text mining techniques.

Step 4: Processing the data. Clustering and other data mining algorithms are applied.

Step 5: Extracting useful information from the text.

This text document will be pre-processed before further processing. In processing removal of stemming words, stopping words, expressions will be done.

- **Stopping words:**

These words are independent of situation.

For Example - “an”, “the”, “by” are considered as stopping words

- **Stemming words:**

It is the process for reducing words to their original or root or stem words.

Word “playing” can be reduced to “play”

- **Score Calculation:**

To determine importance of statement to consider it for summarization we need to calculate its score.

1. **Title Similarity:**

$$f = S \cap T / t \quad (3.1.1)$$

Where

S = Set of words in sentence

T = Set of words in title

$S \cap T$ = Similar words in sentence and title of document

By using this formula, we can calculate score of sentence. Words in the statement will be compared with words in words in title of summary, and frequency will be counted. Sentences having more frequency word count will be considered as an important statement.

2. **Term Weight (Term frequency):**

To know importance of statement term weight is used. It calculates “tf” and “idf”. “idf” represents whether the term is evenly appeared in document or it is rarely present in document. Following formula helps to calculate importance score of word “wi” ,

$$w_i = tf_i \cdot idf_i = tf_i \cdot \log(N/n_i) \quad (3.1.2)$$

where,

tf_i = the term frequency of word i in the document

N = the total number of sentences

n_i = the number of sentences in which word i occurs

This feature can be calculated as follows.

$$f = \sum_{k=1}^k W_i(S) \cdot \max(\sum_{k=1}^k W_i(S_i/N)) \quad (3.1.3)$$

where

k = number of words in sentence

3. **Sentence Length:**

Length of sentence is a massive factor while deciding importance of sentence. Sentence which are too short can be eliminated. Score is calculated by,

$$f = \text{Number of Words in Sentence} / \text{Number of words occurring in longest sentence} \quad (3.1.4)$$

4. **Thematic Word:**

It is related to how relevance of sentence to domain or topic of conversation. Frequency of words in sentences is checked for subject of domain. If sentence has more thematic words then, it has more importance. This score is calculated by

$$f = \text{Number of thematic word in } S / \text{Maximum(Number of Thematic Word)} \quad (3.1.5)$$

5. **Proper Nouns:**

As people gets involved in conversation, referencing to each other, delegation of work mostly happens. During this, proper nouns are used. Thus proper nouns are most essential and should be considered while calculating importance score of sentence.

$$f = \text{Number of Proper nouns in } S / \text{Sentence Length} \quad (3.1.6)$$

6. **Sentence to Sentence Similarity:**

Similarity between sentences and relation between them decides importance score of conversation.

7. Numerical Data:

During meeting numerical terms are highly expected. Sentence having numerical values should be considered while summarizing text document. It is essential when conversation is leading into financial, business context. Score is calculated by

$$f = \text{Number of Numerical Data in } S / \text{Sentence Leng} \quad (3.1.7)$$

IV.CONCLUSION

The proposed system effectively addresses the issue of lengthy and manual systems for conference meeting. This will help to serve the timing related issues as well as encourage to avoid use of papers. Use of this system gives the advantage of better analysis and administration of data.

REFERENCES

- [1] R. Sagayam, A survey of text mining: Retrieval, extraction and indexing techniques, International Journal of Computational Engineering Research, vol. 2, no. 5, 2012.
- [2] N. Padhy, D. Mishra, R. Panigrahi et al., "The survey of data mining applications and feature scope," arXiv preprint arXiv:1211.5723, 2012.
- [3] W. Fan, L. Wallace, S. Rich, and Z. Zhang, "Tapping the power of text mining," Communications of the ACM, vol. 49, no. 9, pp. 76–82, 2006.
- [4] S. M. Weiss, N. Indurkha, T. Zhang, and F. Damerou, Text mining: predictive methods for analyzing unstructured information. Springer Science and Business Media, 2010.
- [5] S.-H. Liao, P.-H. Chu, and P.-Y. Hsiao, "Data mining techniques and applications—a decade review from 2000 to 2011," Expert Systems with Applications, vol. 39, no. 12, pp. 11 303–11 311, 2012.
- [6] W. He, "Examining students online interaction in a live video streaming environment using data mining and text mining," Computers in Human Behavior, vol. 29, no. 1, pp. 90–102, 2013.
- [7] G. King, P. Lam, and M. Roberts, "Computer-assisted keyword and document set discovery from unstructured text," Copy at <http://j.mp/1qdVqhx> Download Citation BibTex Tagged XML Download Paper, vol. 456, 2014.
- [8] N. Zhong, Y. Li, and S.-T. Wu, "Effective pattern discovery for text mining," IEEE transactions on knowledge and data engineering, vol. 24, no. 1, pp. 30–44, 2012.
- [9] A. Henriksson, H. Moen, M. Skeppstedt, V. Daudaravicius, and "M. Duneld, "Synonym extraction and abbreviation expansion with ensembles of semantic spaces," Journal of biomedical semantics, vol. 5, no. 1, p. 1, 2014.
- [10] B. Laxman and D. Sujatha, "Improved method for pattern discovery in text mining," International Journal of Research in Engineering and Technology, vol. 2, no. 1, pp. 2321–2328, 2013.