iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** II **Month of publication:** February 2024

DOI: https://doi.org/10.22214/ijraset.2024.58145

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue II Feb 2024- Available at www.ijraset.com

Meeting Summarization: A Survey

Manas Bidkar[1], Dr. Geeta Chillarge[2], Sahil Chaudhari[3], Nitin Choudhary[4], Chirantan Degloorkar[5]
Department Computer Science, MMCOE, Pune, India

Abstract: ***In an era characterized by an incessant influx of information, the demand for efficient knowledge extraction tools has become paramount. This research introduces a Meeting Sum- marizer, a cutting-edge system that amalgamates speech recog- nition and natural language processing (NLP) to autonomously distill salient information from recorded meetings. The primary objective of this research is to alleviate the cumbersome process of manual meeting summarization by harnessing the capabilities of advanced machine learning techniques tailored to audio data. The Meeting Summarizer leverages state-of-the-art speech recognition algorithms to transcribe spoken content into textual form, laying the foundation for subsequent NLP-based analysis. Through the integration of deep learning methodologies, the system discerns key discussions, identifies critical components, and extracts context from meeting transcripts. The synergy of speech recognition and NLP empowers the system to comprehend linguistic nuances, enabling it to adapt to diverse meeting contexts.***

The Meeting Summarizer promises not only time-saving ad- vantages but also heightened accuracy and reliability in the summarization process. As organizations grapple with the chal- lenge of managing vast amounts of meeting data, this research paves the way for a transformative tool that revolutionizes information extraction in professional settings. The paper delves into the technical intricacies of the speech recognition and NLP integration, the system's learning process, and the anticipated impact on meeting summarization. By the conclusion, we un- derscore the profound implications of the Meeting Summarizer in reshaping how organizations harness knowledge from their meetings, offering a glimpse into the future of autonomous information extraction.

Index Terms: Natural Language Processing (NLP), Meeting Summarization, Pattern Recognition, Context Extraction, Time-saving.

I. INTRODUCTION

In the contemporary landscape of information overload, the management and utilization of vast volumes of data have become critical challenges, particularly in professional settings where meetings serve as a fundamental means of communication and collaboration. Recognizing the need for efficient knowledge extraction from recorded meetings, this research endeavors to introduce a Meeting Summarizer – a sophisticated system designed to autonomously distill pivotal aspects and key discussions embedded within meeting record- ings.

The central objective of this research is to address the time- consuming and labor-intensive process of manually summariz- ing meeting content by leveraging advanced machine learning techniques tailored to audio data processing. By integrating modern approaches, including speech recongition and natural language processing, the Meeting Summarizer aims to discern critical components, identify significant phrases, and extract important context from recorded meetings.

Through a meticulous learning process, the system en- deavors to continuously refine its summarization capabilities over time. This evolution is driven by the utilization of ma- chine learning algorithms that enable the system to recognize patterns, understand language nuances, and adapt to diverse meeting contexts. By automating the summarization process, this research seeks to streamline workflow efficiency, allowing professionals to allocate their time more strategically and focus on higher-level tasks.

Moreover, the Meeting Summarizer is envisioned not only as a time-saving tool but also as a means to reduce the dependence on human intervention in the summarization pro- cess. The incorporation of machine learning techniques is expected to elevate the accuracy and reliability of meeting summarization, mitigating the likelihood of human-induced errors and providing a more robust and consistent output.

II. RELATED WORK

Researchers have recognised the necessity of video sum- marization in recent years and have determined that it is a very helpful and important tool with applications in a variety of fields, including business, law, medicine, education, and education, in addition to people's personal life. This section discusses relevant work in speech recognition and text summarising technologies as well as previous efforts to apply video summarization.

Tirath, Lakshaya, Yash, Renuka (2023)[1] suggests a two- pronged method for text summarising and speech detection in video summarization. Speech to text conversion and transcript generation for input movies are accomplished by Automatic Speech Recognition (ASR), which is based on a Convolutional Neural Network (CNN). The ASR model, which is appropriate for voice recognition, is a sequence-2-sequence (seq2seq) model based on Connectionist Temporal Classification (CTC). Word Error Rate (WER) is a metric used to assess the ASR model's accuracy. NLP-based algorithms for extractive text summarization use the output transcripts as input to produce succinct and enlightening summaries.

Neha Jain and Somya Rastogi (2019) [14] mentioned the use of interdisciplinary technologies such as Pattern Recognition, Signal Processing, and Natural Language Processing in imple- menting unified statistical frameworks for Speech Recognition Systems.

V Poliyev and O N Korsun (2020) [9] involves training the neural network on sets consisting of only hundreds or thousands of samples addressing the following four sce- narios: speaker-independent recognition with noise, speaker-dependent recognition with noise, speaker-dependent recog- nition with noise, and speaker-independent recognition with noise. With the selection of optimal CNN architecture achieved Recognition errors of 2.0 and 0.5 were achieved by using 7 and 20 samples of each word, respectively, in the training set.

Takatomo, Atsunori, Marc, and Shinji (2021) [22] proposes a cascade speech summarization model that combines auto- matic speech recognition (ASR) and text summarization (TS) using state-of-the-art models such as Transformer for ASR and Bidirectional Encoder Representations from Transformers (BERT) for TS. They explores another approach that replaces the input sub-word embedding of BERT with a sum of sub- word embedding vectors weighted by their ASR posterior values. This approach is referred to as "BERTSum Pos.fusion" and is evaluated on speech summarization datasets, including YouTube How2 video and TED Talk summarization

The two most common methods of summarising are ex- tractive and abstractive. Bryan Au-Yeung and Vasanth Ramani (2020) [9] applies the BERTSum model to generate abstractive summaries of narrated instructional videos, using transfer learning and pretraining on large cross-domain datasets in both written and spoken English. The transcripts of the instructional videos are preprocessed to restore sentence segmentation and punctuation in the output of an ASR system. The authors also fine-tune the BERT-based text summarization models on the auto-generated scripts from instructional videos.

Srikanth et al. (2020) [8] uses the BERT model for extractive summarization by clustering sentence embeddings using Kmeans clustering. It introduces a dynamic method to determine the suitable number of sentences to pick from clusters, also mentioned the use of transformers, specifically BERT, for identifying contextual relationships between words in a sentence.

Aswin et al. (2021) [6] proposed technique for automatic subtitle generation and semantic video summarization utilizes speech recognition and NLP-based text summarization algo- rithms. Performanced is enhanced using two ensemble method the Intersection method and Weight-based learning method.

Vinnarasu A. and Deepa V. Jose (2019) [10] proposed method which involves converting speech to text using Google API and producing a summarized version of the text using python NLTK.The method eliminates words with less im- portance by setting a minimum and maximum range for the occurrence of specific words.

Sarah S. Alrumiah and Amal A. Al-Shargabi (2021) [5] proposed atent Dirichlet Allocation (LDA) for subtitles sum- marization, which involves three phases: preparing the subtitle file, training the LDA model on subtitles, and generating a summary based on the keywords list. The authors also propose a length enhancement method to improve the precision rates of the generated summaries.

Shah et al. (2022) [20] discuss the value of ignoring a video's unnecessary portions and the time lost by watching only the most pertinent portions segments of a video. The discussion also covers a logical method using Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for comparing different frames and generating end results. Various scores were calculated, including superframe cut scores, motion scores, uniqueness scores, and colorfulness scores, to rank and select important frames.

III. DOMAIN BASED SUMMARIZATION

A. *Dialogue Summarization*

Dialogue summarization, covering various domains and organizing the current works in the field. The field of dialogue summarization is becoming more and more popular as a result of the abundance of conversational data that is available when individuals use digital platforms and cellphones to com- municate. Emails, conferences, online chats, customer sup- port exchanges, doctor-patient medical discussions, podcasts, and other communications can all be found in the dialogue domain.[21].Section IV will go into more detail about the meeting-based summarising sub-domain within the dialogue summarization domain.

B. Chat Summarization

As we navigate through the landscape of chat-based com- munication, the implications of Chat Summarization extend beyond data condensation. It has the potential to revolutionize how we distill insights from online discussions, enhancing information retrieval, decision-making processes, and user experience in various domains. One of the earliest work is a personalized summarization based on user profiles from group chats, like Slack channels by a chat assistant service, Collabot dive more into the issue of abstractive dialogue summarization in the current paper.[15] Summarising discussions between a small number of participants is an intriguing new area of research in summarization, given the increasing popularity of online talks through apps like WeChat, Messenger, and WhatsApp. We have developed the SAMSum Corpus1, which includes over 16,000 chat conversations with manually anno- tated summaries, specifically for this purpose. The scientific community can access the dataset without charge.[11]

C. Email Summarization

Email thread summarization was first approached using a variety of techniques, such as extracting topic phrases (Mure- san, 2001) [18], the extraction of shorter overviews and longer summaries for each group, providing a way to summarize and condense the information in the discussions. creating message clusters based on topic groups (Newman and Blitzer, 2003) [18], extracting significant sentences based on particular email- related characteristics (Rambow, 2004) [17], etc.

IV. MEETING SUMMARIZATION

This section outlines different methods for summarising meeting transcripts in order to produce a brief summary. The two types of summarization techniques covered in this section are extractive and abstractive based.

A. Extractive Meeting Summarization

Extractive meeting summarization is a technique that in- volves selecting and extracting key sentences or phrases di- rectly from a meeting transcript to form a concise summary. Instead of generating new sentences, this approach takes a key phrase or piece of information and extracts it from the source text file. Using linguistic or statistical characteristics, an extractive text summarising method chooses pertinent, educational sentences. One of the earliest work in extractive meeting summarization is by (Luhn, 1985)[20]. He conducted his research on the extractive summarization . From his studies, he gleaned significant phrase by figuring the word frequency and frequency that provides a helpful indicator of its importance. (Baxendale) conducted his first summarization research at IBM in 1958.[21] Using the text's position, he was able to extract a significant sentence. The writer has examined 200 paragraphs in search of his objective to discover that in 85 percent of the sentences in which the author has chosen the primary topic. Topic sentences and final sentences accounted for 7 percent. The majority of precise phrases would be chosen from these two phrases. In 1969, (Edmundson) conducted research on extracted summarization . He used two features—word frequency and importance—that were derived from earlier works to extract key sentences. The skeleton of the document and the use of cue words are the two new addi- tions made by the author. In 2019, (J.N.Madhuri and Ganesh Kumar.R) [12] proposed an approach involves tokenizing the input text file, calculating the weighted frequency of keywords, and ranking sentences based on their weighted frequency. The high-ranked sentences are then extracted and converted into audio form.

B. Abstractive Meeting Summarization

Abstractive meeting summarization, which automatically provides the condensed summary encompassing the key topics in the meeting dialogue, is a tough challenge in natural language comprehension. But because they don't model the unstructured long-form conversational contents, the current abstractive summary efforts may be inapplicable to the task of summarising meetings. Instead, they concentrate mostly on structured text documents. (Zhou Zhao) propose the hierar- chical adaptive encoder to learn the high-level semantic rep- resentation of meeting conversation contents, and then devise the reinforced decoder network to generate the summaries for abstractive meeting summarization. He conducted extensive experiments on the popular AMI meeting conversation dataset to verify the efficiency of suggested technique[15]. Author uses transformers instead of RNN-based models for training data parallel in abstractive summarization. The paper high- lights the benefits of using multi-head attention and positional encoding in transformer models, which outperform LSTM- based models.The abstractive method of summarization is employed to train the model for dialogue systems, which involves generating new vocabulary words while acquiring extractive information from the text corpus[16].

V. METRICES FOR SUMMARY EVALUATION

Rouge (Recall-Oriented Understudy for Gisting assessment), an assessment metric commonly used for text sum- marization, is an automated method for assessing the calibre of produced synopses. This metric counts the amount of overlapping ngrams between a generated summary and a set of human-authored summaries in order to compare them citation summaries. Some of the often reported metrics in most articles include F1 scores for Rouge-L (longest common sequence), Rouge-1 (unigram overlap), and Rouge-2 (bicogram overlap). Another artificial metric called BERTScore uses contextual embeddings from pretrained BERT to determine semantic similarity.

VI. CHALLENGES

One of the big problems faced by Abstractive Meeting Summarizers is making sure the summary stays true to the original meeting and doesn't make up things. Meetings can be complex, with lots of different ways people talk, so it's hard for the summarization system to get the exact meaning right. Furthermore, meeting summarizers frequently face diffi- culties when dealing with long meeting transcripts. A meeting transcript usually contains more tokens than a transformer architecture is able to manage. In an effort to efficiently condense long inputs, this problem has been the subject of numerous recent efforts. Still, this is a problem that requires further focus. In addition, the majority of meetings held in the industry are confidential in nature, hence there aren't enough annotated datasets available for meetings. A meeting involves several people, and information is shared through speaker-to- speaker conversations. The conversations are more talkative in style and also experience a deficiency in consistency.

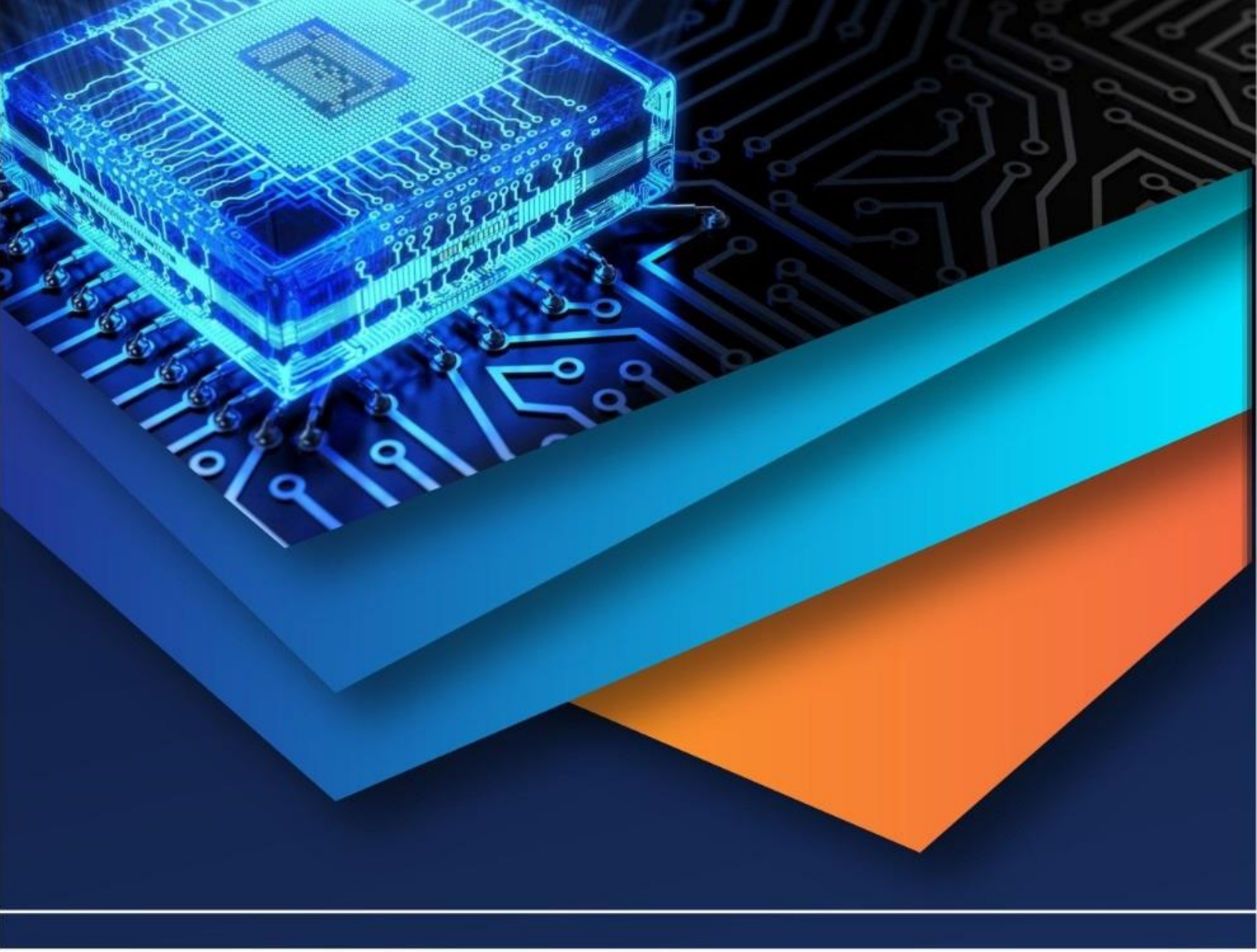
VII. CONCLUSION

In conclusion, leveraging Transformer models and Speech Recognition for meeting summarization can significantly en- hance efficiency and productivity. By automating the process of capturing and condensing information, we can empower participants to focus on meaningful discussions and make in- formed decisions. Embracing these technologies opens up new possibilities for more effective and streamlined meetings in the digital age. This document provides an extensive overview of the work that has been done on meeting summaries thus far. The most recent abstractive and extensive summarization models for meetings are covered in this literature. We also talk about the difficulties that researchers encounter and provide insight into upcoming studies. Future research may concentrate on the issue of factual discrepancy and the distillation of extensive meeting transcripts.

REFERENCES

[1] T. Tyagi, L. Dhari, Y. Nigam and R. Nagpal, "Video Summarization using Speech Recognition and Text Summarization," 2023 4th Interna- tional Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-7, doi: 10.1109/INCET57972.2023.10169901.
[2] D. Shah, M. Dedhia, R. Desai, U. Namdev and P. Kanani, "Video to Text Summarisation and Timestamp Generation to Detect Impor- tant Events," 2022 2nd Asian Conference on Innovation in Technol- ogy (ASIANCON), Ravet, India, 2022, pp. 1-7, doi: 10.1109/ASIAN- CON55314.2022.9909008.
[3] Kumar, Lakshmi and Kabiri, Arman. "Meeting Summarization: A Sur- vey of the State of the Art", 2022 10.48550/arXiv.2212.08206.
[4] T. Kano, A. Ogawa, M. Delcroix and S. Watanabe, "Attention- Based Multi-Hypothesis Fusion for Speech Summarization," 2021 IEEE Automatic Speech Recognition and Understanding Work- shop (ASRU), Cartagena, Colombia, 2021, pp. 487-494, doi: 10.1109/ASRU51503.2021.9687977.
[5] Alrumiah, Sarah and Al-Shargabi, Amal. "Educational Videos Sub- titles' Summarization Using Latent Dirichlet Allocation and Length Enhancement", 2021 Computers, Materials and Continua. 70. 6205-6221. 10.32604/cmc.2022.021780.
[6] V. B. Aswin, Mohammed Javed, Parag Parihar, K. Aswanth, C. R. Druval, Anupam Dagar, C. V. Aravinda, "NLP-Driven Ensemble-Based Automatic Subtitle Generation and Semantic Video Summarization Technique", 2021 Advances in Artificial Intelligence and Data Engi- neering, Volume 1133 ISBN : 978-981-15-3513-0
[7] D. Singhal, K. Khatter, T. A and J. R, "Abstractive Summarization of Meeting Conversations," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298305.
[8] A. Srikanth, A. S. Umasankar, S. Thanu and S. J. Nirmala, "Extractive Text Summarization using Dynamic Clustering and Co-Reference on BERT," 2020 5th International Conference on Computing, Communica- tion and Security (ICCCS), Patna, India, 2020, pp. 1-5, doi: 10.1109/IC- CCS49678.2020.9277220.
[9] Alexandra Savelieva and Bryan Au-Yeung and Vasanth Ramani, "Ab- stractive Summarization of Spoken and Written Instructions with BERT", 2020 10.48550/arXiv.2008.09676.
[10] A V Poliyev and O N Korsun, "Speech Recognition Using Convolutional Neural Networks on Small Training Sets", 2020 IOP Conf. Ser.: Mater. Sci. Eng. 714 012024
[11] A, Vinnarasu and Jose, Deepa. "Speech to text conversion and summa- rization for effective understanding and documentation", 2019 Interna- tional Journal of Electrical and Computer Engineering (IJECE). 9. 3642. 10.11591/ijece.v9i5.pp3642-3648.
[12] Bogdan Gliwa, Iwona Mochol, Maciej Biesek, and Aleksander Wawer. "SAMSum Corpus: A Human-annotated Dialogue Dataset for Abstrac- tive Summarization", 2019 In Proceedings of the 2nd Workshop on New Frontiers in Summarization, pages 70–79, Hong Kong, China. Association for Computational Linguistics.
[13] J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," 2019 International Conference on Data Sci- ence and Communication (IconDSC), Bangalore, India, 2019, pp. 1-3, doi: 10.1109/IconDSC.2019.8817040.

[14] Zhou Zhao, Haojie Pan, Changjie Fan, Yan Liu, Linlin Li, Min Yang, and Deng Cai, "Abstractive Meeting Summarization via Hierarchical Adaptive Segmental Network Learning", 2019 In The World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, NY, USA, 3455–3461. https://doi.org/10.1145/3308558.3313619
[15] Jain, Neha and Rastogi, Somya. "SPEECH RECOGNITION SYSTEMS – A COMPREHENSIVE STUDY OF CONCEPTS AND MECHANISM", 2019 Acta Informatica Malaysia. 3. 01-03. 10.26480/aim.01.2019.01.03.
[16] Tepper, Naama, Hashavit, Anat, Barnea, M.Ronen, Inbal, Leiba and Lior, "Collabot: Personalized Group Chat Summarization", 2018 WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 771-774. 10.1145/3159652.3160588.
[17] Chen, Hongshen, Liu, Xiaorui, Yin, Dawei and Tang, Jiliang. "A Survey on Dialogue Systems: Recent Advances and New Frontiers", 2017 ACM SIGKDD Explorations Newsletter. 19. 10.1145/3166054.3166058.
[18] Rambow, Owen, Shrestha, Lokesh, Chen, John, Lauridsen and Chirsty, "Summarizing Email Threads", 2004 10.3115/1613984.1614011.
[19] Paula S. Newman and John C. Blitzer. "Summarizing archived dis- cussions: a beginning", 2003 In Proceedings of the 8th interna- tional conference on Intelligent user interfaces (IUI '03). Associ- ation for Computing Machinery, New York, NY, USA, 273–276. https://doi.org/10.1145/604045.604097
[20] Muresan, Smaranda, Tzoukermann, Evelyne and Klavans, Judith, "Com- bining Linguistic and Machine Learning Techniques for Email", 2001 10.3115/1117822.1117837.
[21] H. P. Luhn, "The Automatic Creation of Literature Abstracts," in IBM Journal of Research and Development, vol. 2, no. 2, pp. 159-165, Apr. 1958, doi: 10.1147/rd.22.0159.
[22] P. B. Baxendale, "Machine-Made Index for Technical Literature—An Experiment," in IBM Journal of Research and Development, vol. 2, no. 4, pp. 354-361, Oct. 1958, doi: 10.1147/rd.24.0354.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY