# Bilingual Minutes of the Meet Generator

**Aarushi Sharan [1], Nandika Rathore [2], Yashveer Tiwari [3] & S. S. Sonawane [4]**

[1]Student; Pune Institute of Computer Technology, (CE), Pune, Maharashtra, India, aarushirsharan@gmail.com

[2]Student; Pune Institute of Computer Technology, (CE), Pune, Maharashtra, India, nandikarathore@gmail.com

[3]Student; Pune Institute of Computer Technology, (CE), Pune, Maharashtra, India, yashveertiwari2001@gmail.com

[4]Associate Professor; Pune Institute of Computer Technology, (CE), Pune, Maharashtra, India, sssonawane@pict.edu

**Abstract:**

A Minutes of Meeting generator plays a pivotal role in addressing critical needs for organizational efficiency and communication. By automating the meticulous process of documenting meeting proceedings, this tool ensures not only accuracy, consistency, and standardization in the minutes but also saves valuable time previously spent on manual transcription and formatting. The generator establishes a clear, standardized format that facilitates easy reference, contributing to a more streamlined and efficient documentation process.

Moreover, the digitized Minutes of the Meeting offer enhanced accessibility, searchability, and the capability to be tagged for categorization, significantly boosting their usability. Beyond these advantages, the Minutes of the Meet generator supports compliance efforts by ensuring that meetings and decisions are documented in alignment with legal or industry standards.

In addition to its role in compliance, the generator facilitates real-time updates and seamless integration with other organizational tools. In essence, a Minutes of Meeting generator proves indispensable for efficient administrative processes, particularly beneficial for large enterprises or teams with frequent meetings, as well as in regulated environments where meticulous documentation is paramount to success.

**Keywords:** Text Summarization, Speech Recognitions, NLP, Minutes of the Meet, Transcription. **1**

## Introduction

Meetings are an essential part of professional life as they help in collaboration, decision-making, and information exchange. However, the process of recording meeting minutes can be time-consuming and challenging. Manual minute-taking can consume resources, introduce inaccuracies, and hamper efficiency. Natural Language Processing (NLP), a subset of artificial intelligence, has emerged as a solution to address this issue. With the advancements in NLP, automated systems, known as NLP minutes of the meeting generators, have revolutionized meeting documentation.

This paper aims to explore the development, capabilities, and implications of NLP meeting generators. It delves into the technologies, applications, and benefits of NLP and analyses the challenges, concerns, and ethical considerations in adopting NLP. The paper fosters a discussion on the role of NLP in business communication.

A solution has been proposed to address the challenge of manual meeting documentation. The proposed solution is technologically driven and automates the Minutes of the Meeting (MoM) process. The solution utilizes various technologies, including ASR, NER, Sentiment Analysis, and Simulation Algorithms to ensure a comprehensive approach. The solution takes into consideration linguistic and cultural contexts, efficiency, accuracy, as well as ethical and privacy concerns. This approach utilizes the latest developments in natural language processing (NLP) and other relevant technologies to automate the generation of meeting minutes.

The field of Automatic Speech Recognition (ASR) has experienced a significant transformation thanks to the advancements in deep learning, specifically neural networks [1]. Complex feature engineering is no longer necessary for End-to-End ASR systems, like "Listen, Attend, and Spell"[2]. Additionally, the demand for multilingual support is met by Multilingual ASR, as explored in "Massively Multilingual ASR and TTS"[3]. Furthermore, "Listen, Attend and Walk"[4] addresses the need for reliable performance in noisy environments, ensuring robust ASR in challenging conditions.

The field of Named Entity Recognition (NER) has evolved from rule-based approaches to machine learning and deep learning techniques. Recently, transformer-based models such as BERT have shown remarkable performance in NER, as demonstrated in the study "BERT for Named Entity Recognition"[5]. Another area of development is multilingual NER, which aims to extract named entities from various languages. The XLM-R model, introduced in "XLM-R: Multilingual NER with Cross-Lingual Pretraining"[6], is one such model. Additionally, there are domain-specific NER models that cater to specialized domains like biomedical or legal texts. These models are still in development and are continuously being improved.

Sentiment analysis has developed over time, moving from rule-based to machine learning and deep learning methods. With the recent advancements in transformer-based models such as BERT, the accuracy and efficiency of sentiment analysis have significantly improved.[7] Aspect-Based Sentiment Analysis is a specialized technique that focuses on identifying sentiment toward specific aspects of a text [8]. Multilingual Sentiment Analysis is a technique that deals with sentiment analysis across multiple languages. The paper titled "Multilingual Sentiment Analysis with Transformers"[10] explores this topic in detail. There are certain challenges to this sentiment analysis as well. FineGrained Sentiment Analysis: Improving sentiment analysis to recognize fine-grained sentiment categories beyond just positive and negative. Emotion Detection: Advancing sentiment analysis to detect and classify emotions such as joy, anger, sadness, and more. Multimodal Sentiment Analysis: Integrating visual and audio information with text for more comprehensive sentiment analysis. Cross-Domain and Cross-Lingual Sentiment Analysis: Extending sentiment analysis models to work well across different domains and languages.

Simulation algorithms, including Monte Carlo Simulation, Discrete Event Simulation, Agent-Based Modelling, and Machine Learning-driven simulation, have seen significant developments.[9]

Quantum Simulation leverages quantum computers for modelling quantum systems.

Text Rank is a graph-based approach that takes inspiration from Google's PageRank algorithm, which assesses the significance of web pages in search engine results. In the domain of keyword extraction, Text Rank builds a graph representation of the text, where words or phrases are nodes, and the connections between them are edges. The algorithm assigns scores to each node based on the graph structure, where higher scores indicate greater importance. Text Rank has been proven effective in identifying keywords by considering the contextual relationships between words and phrases in a document. The method is particularly helpful for capturing key terms that are crucial to the overall meaning of the text.

Frank et al. (1999) introduced a supervised algorithm in their paper titled "A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text." This approach to keyword extraction relies on machine learning techniques and operates by learning from annotated data. In a supervised setting, the algorithm is trained on a dataset where examples of keywords are labelled, indicating their presence or absence in the text. The model then generalizes from this labelled training data to identify keywords in unseen or new texts.

In the context of the mentioned paper, the algorithm likely involves the use of features derived from the text, such as word frequency, context, or syntactic patterns, to train a machine-learning model. The model learns to recognize patterns associated with the presence of keywords, particularly focusing on abbreviations and their definitions in biomedical texts. This method is beneficial when a reliable labelled dataset is available for training, enabling the algorithm to learn the characteristics of keywords specific to the biomedical domain.

Turney and Littman delved into unsupervised keyword extraction in their work "Measuring Praise and Criticism: Inference of Semantic Orientation from Association."[11] Unlike supervised methods, unsupervised approaches do not rely on labelled training data. Instead, they aim to identify keywords by exploring inherent patterns, relationships, or properties within the text itself.

In the specified paper, Turney and Littman likely used unsupervised techniques based on the semantic orientation of words. Semantic orientation refers to the polarity or sentiment associated with words, and by measuring the associations between words in a corpus, this method infers the semantic orientation of a term. Terms with strong associations are considered potential keywords, as their relationships with other terms contribute to the overall meaning of the text. Unsupervised methods are advantageous in scenarios where obtaining labelled data for training is challenging or impractical, allowing for broader applicability across various domains and types of texts.

Lin introduced the ROUGE metric in the paper titled "ROUGE: A Package for Automatic Evaluation of Summaries."[12] ROUGE is a set of metrics designed for the automatic evaluation of machine-generated summaries. The primary focus of ROUGE is on assessing the quality of summaries based on recall, which measures the ability of the system to capture important information present in the reference summaries.

ROUGE evaluates the overlap between the n-grams (contiguous sequences of n items, usually words or characters) in the machine-generated summary and the reference summary. It includes various measures such as ROUGE-N (unigrams, bigrams, trigrams, etc.), ROUGE-L (longest common subsequence), and ROUGE-W (weighted overlap) to capture different aspects of summary quality. ROUGE has become a standard metric in the field of natural language processing particularly in tasks like keyword extraction and summarization.

Cross-lingual key phrase extraction [13] addresses the challenges associated with extracting keywords from documents that span multiple languages. Cross-lingual key phrase extraction is particularly relevant in scenarios where documents are available in different languages, and there is a need to identify key terms that convey important information across language barriers. The study likely explores techniques that leverage bilingual dictionaries or other cross-lingual resources to enhance the accuracy of key phrase extraction in a multilingual context.

Guo (2020) proposed "BERT-MEE: Multilingual End-to-End Key Phrase Extraction,"[14] introducing a method that leverages multilingual BERT (Bidirectional Encoder Representations from Transformers) for cross-language key phrase extraction. BERT, a powerful transformer-based model, is known for its contextualized word representations and has been widely adopted in various natural language processing tasks.
In the context of key phrase extraction, BERT-MEE likely extends the capabilities of BERT to handle documents in multiple languages seamlessly. The use of multilingual BERT allows the model to capture contextual information and semantic relationships across diverse languages, enabling more accurate and context-aware key phrase extraction. Multilingual models like BERT-MEE contribute to breaking down language barriers in information extraction tasks, making it possible to obtain meaningful keywords from documents irrespective of the language in which they are written. This advancement is crucial for applications dealing with multilingual content, such as global information retrieval and cross-cultural knowledge discovery.

The collective impact of these technological advancements is poised to reshape communication and information processing in professional settings. The proposed solution not only addresses existing challenges but also sets the stage for a future where automated, intelligent systems play a central role in ensuring efficiency, accuracy, and ethical considerations. As these technologies continue to evolve, the envisioned transformation holds the potential to elevate professional practices, fostering a new era of streamlined, informed, and responsible communication and decisionmaking.

In summary, the exploration of Natural Language Processing (NLP) and its related technologies in this paper reveals a paradigm shift in addressing critical challenges across various domains, including meeting documentation, keyword extraction, sentiment analysis, and simulation. The proposed solution presented for meeting documentation stands out as a testament to the transformative potential of cutting-edge technologies in streamlining and automating traditionally cumbersome processes.

## 2 Proposed Methods

The system has four main steps to translate and summarize meeting minutes in a source language to a target language. The steps are as follows:

**1. Data Collection:**

Collect a diverse range of meeting minutes in the source language and their corresponding translations in the target language. Include various meeting types, topics, and domains to ensure dataset diversity.

**2. Data Preprocessing:**

Clean and preprocess the collected data to remove any irrelevant content, formatting inconsistencies, and noise. Tokenize the text and create aligned bilingual sentence pairs for training.

**3. Model Selection:**

Use BART (Bidirectional and Auto-Regressive Transformers) as the AI model for translating and summarizing meeting minutes. BART is chosen because of its sequence-to-sequence capabilities and its translation and summarization capabilities.

**4. Model Training:**

Fine-tune the pre-trained BART AI model using the collected bilingual dataset. Train the model to perform two primary tasks: translation (translate meeting minutes from the source language to the target language) and summarization (generate concise summaries of the meeting minutes in the target language).

### 2.1 Algorithm:

1. Start
2. Import Bart and Speech Recognition;
3. Listen to the user's speech using the listen() function;
4. Transcribe the speech using the recognize_google() function; 5. Tokenize the text;

6. Generate the summary and then de-tokenize it; 7.
Print the summarized text;
8. Stop.

### 2.2 Speech to Text:

- Importing the Speech Recognition [15] library in Python.
- Then utilize the imported library to set up the device microphone as the source.
- After that employ the `listen()` function to actively capture and record the speaker's voice through the microphone.
- Then apply the `recognize_google()` function to transcribe the recorded audio into textual format, hence leveraging Google's speech recognition capabilities to convert spoken words into written text accurately.

### 2.3 Summarization:

- Summarization starts by importing the transformers library in Python, which facilitated access to various natural language processing models and tools.
- Facebook's Bidirectional and Auto-Regressive Transformers as the model for summarization were used for this. This model is known for its effectiveness in capturing contextual information from both directions and generating coherent summaries.

- Tokenization of the entire input speech by leveraging the pre-trained transformer model. Tokenization involves breaking down the input speech into smaller units (tokens) for further processing.
- Later, the selected transformer model generates a concise summary of the tokenized speech. The model's architecture enables it to understand the contextual nuances and key information, ensuring a meaningful summarization.
- Then, the tokens are then de-tokenized in the generated summary, reconstructing the text into a coherent and readable format.
- This de-tokenized output represents the final summarized version of the input speech, capturing the essential information while maintaining readability and coherence.

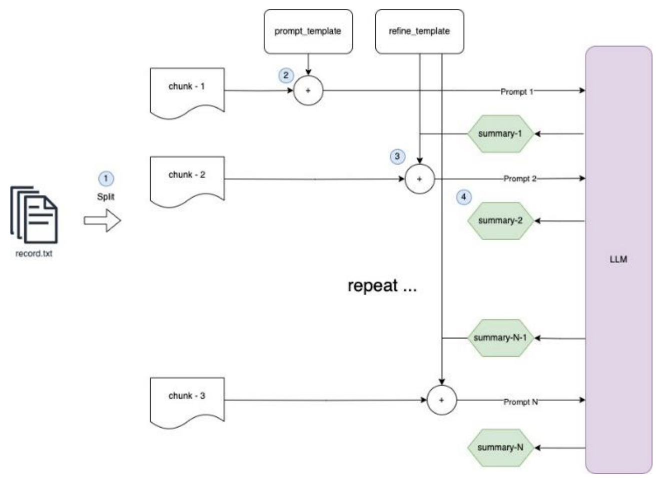


Fig.1: Document Summary Refine Workflow [16]

### 2.4 BART's architecture [18]:

- BART is built upon the Transformer architecture, which consists of an encoder-decoder framework with self-attention mechanisms.
- Transformers use self-attention mechanisms to weigh the importance of different words in a sequence when processing each word, enabling the model to capture long-range dependencies.
- BART incorporates bidirectional pretraining, which means it is trained to predict both the previous and next words in a sentence. This bidirectional training helps the model capture contextual information from both directions.
- The auto-regressive pretraining aspect involves training the model to generate a target sequence one token at a time, conditioned on the preceding tokens. This prepares the model for tasks like language generation. • BART is pre-trained using a masked language modelling objective, like BERT (Bidirectional Encoder Representations from Transformers). During training, random tokens in the input are masked, and the model is trained to predict these masked tokens.

- After pretraining, BART is fine-tuned for specific tasks using a sequence-to-sequence training setup. This involves using a pair of input and target sequences, where the model is trained to generate the target sequence from the input sequence.
- In the context of text summarization, BART can be fine-tuned using pairs of sources (original text) and target (summary) sequences. The model learns to generate concise and coherent summaries of input texts.

## 3. Results & Discussion

### 3.1 Equations [18]

First, the model is pre-trained on tokens "t" looking back to "k" tokens in the past to compute the current token. This is done unsupervised on a vast text corpus to allow the model to "learn the language."

$$\boldsymbol{L_1(T)} = \sum_i \log P(\boldsymbol{t_i} | \boldsymbol{t_{i-k}}, \ldots, \boldsymbol{t_{i-1}}; \boldsymbol{\theta}) \qquad \textbf{(1)}$$

Next, to make the model robust on a specific task, it is fine-tuned in a supervised manner to maximize the likelihood of label "y" given feature vectors x1…xn.

$$\boldsymbol{L_2(C)} = \sum_{x,y} \log P(\boldsymbol{y} | \boldsymbol{x_1}, \ldots, \boldsymbol{x_n}) \qquad \textbf{(2)}$$

Combining 1 and 2, we get the objective in 3. Lambda represents a learned weight parameter to control the influence of language modeling.

$$\boldsymbol{L_3(C)} = \boldsymbol{L_2(C)} + \boldsymbol{\lambda L_1(C)} \qquad \textbf{(3)}$$

### 3.2 Tables and Figures

Table 1 Evaluation Metrics

| **Test** | **Score (0 - 1)** |
|---|---|
| BLEU score | 0.209025 |
| ROUGE-1 score | 0.507937 |
| ROUGE-2 score | 0.287582 |
| ROUGE-L score | 0.47619 |

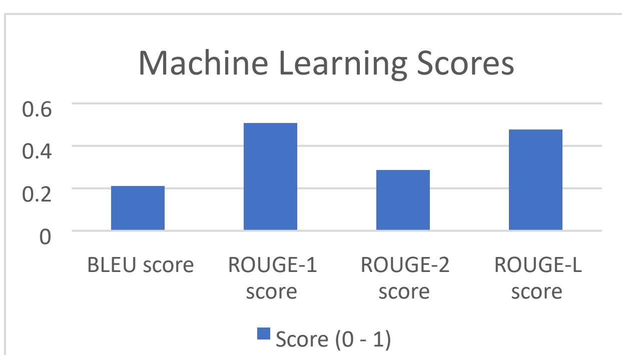


Fig. 2: Chart 1: Histogram of the Evaluation metrics Obtained.

CNN daily mail dataset [17] is used for analysis purposes. Evaluation metrics - BLEU (Bilingual Evaluation Understudy and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is used for analysis purpose. The analysis is shown in table 1. The Rouge L score indicates a marginally good overlapping score of system summary with reference summary.

## 4 Conclusion

In summary, the development of a Bilingual Minutes of the Meeting (MoM) generator represents a notable advancement in the domain of natural language processing and meeting management. This innovative tool has the potential to transform the way organizations conduct and document their meetings, particularly in the context of international communication.

While the Bilingual MoM generator shows promise in addressing language barriers and enhancing meeting documentation, it is not without its challenges. Language pair dependencies, accuracy issues, and the handling of cultural nuances pose hurdles that require attention. Additionally, privacy and post-editing concerns remain valid.

Looking ahead, there are significant opportunities for improvement and expansion. Multilingual support, improved domain adaptation, and real-time translation hold the promise of enhanced accuracy and user-friendliness. Privacy measures and seamless integration with existing meeting tools are vital for broader adoption.

In an era of globalization, the Bilingual MoM generator stands as a valuable tool for promoting cross-cultural understanding and effective communication in diverse contexts. As it continues to evolve and address its limitations, this technology is poised to be indispensable for organizations navigating the complexities of a globalized world. However, the limitation of the research work is the system's effectiveness heavily depends on language pairs, with less common pairs or those with substantial linguistic differences posing challenges, Capturing Cultural nuances, idiomatic expressions, and humor in the source language can be difficult, potentially affecting output quality, resolving ambiguity in the source text, such as homonyms or polysemy, can be intricate and may result in inaccurate translations or summaries and handling sensitive or confidential information in meeting minutes may raise privacy and security concerns, particularly when using cloud-based translation services.

The future scope is to Expand the system to support multiple languages and dialects and enhance its applicability in diverse international contexts, develop techniques for better domain adaptation to handle specialized jargon and terminology in meeting content, and enable real-time translation and summarization during live meetings for instant access to bilingual minutes, develop AI models that can self-assess the quality of translations and summaries and provide feedback for refinement and leveraging the latest advancements in AI models, such as more powerful transformer-based models, to improve translation and summarization quality.

## References

[1] Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., & Ng, A. Y. Deep Speech: Scaling up end-to-end speech recognition. (2014).

[2] Chan, W., Zhang, Y., Le, Q., & Jaitly, N. Latent Sequence Decompositions. (2016).

[3] Baevski, A., Schneider, S., & Auli, M. Vq-wav2vec: Self-Supervised Learning of Discrete Speech Representations. (2019).

[4] Liu, H., Simonyan, K., & Yang, Y. DARTS: Differentiable Architecture Search. (2018).

[5] Alsentzer, E., Murphy, J. R., Boag, W., Weng, W., Jin, D., Naumann, T., & McDermott, M. B. Publicly Available Clinical BERT Embeddings. (2019).

[6] Du, J., Grave, E., Gunel, B., Chaudhary, V., Celebi, O., Auli, M., Stoyanov, V., & Conneau, A. Self-training Improves Pre-training for Natural Language Understanding. (2020).

[7] Biesialska, M., Biesialska, K., and Rybinski, H. Leveraging contextual embeddings and self-attention neural networks with bi-attention for sentiment analysis. Journal of Intelligent Information Systems, 57(3), 601-626. (2021).

[8] Transfer Learning in Natural Language Processing (Ruder et al., NAACL 2019)

[9] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*. (2018).

[10] Barriere, V., & Balahur, A. Improving Sentiment Analysis over non-English Tweets using Multilingual Transformers and Automatic Translation for Data-Augmentation. (2020).

[11] Turney, P. D., Littman, M. L. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. Canada: National Research Council of Canada. (2003).

[12] ROUGE: A Package for Automatic Evaluation of Summaries (Lin, 2004)

[13] R. Jungnickel, A. Pomp, A. Kirmse, X. Li, V. Samsonov and T. Meisen, "Evaluation and Comparison of Cross-lingual Text Processing Pipelines," 2019 IEEE Symposium Series on Computational Intelligence (SSCI), Xiamen, China, 2019

[14] Weiwei Guo, Xiaowei Liu, Sida Wang, Huiji Gao, Ananth Sankar, Zimeng Yang, Qi Guo, Liang Zhang, Bo Long, Bee-Chung Chen, and Deepak Agarwal. 2020. DeText: A Deep Text Ranking Framework with BERT. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20). Association for Computing Machinery, New York, NY, USA, 2509–2516.

[15] G. Hinton et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," in IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82-97, Nov. 2012

[16] Kevin Du "Summarizing your meeting with ChatGpt and LangChain" dxiaochuan.medium.com 8th June 2023 [17] Dataset: cnn_dailymail https://huggingface.co/datasets/cnn_dailymail 2023

[17] Param Raval "Transformers BART Model Explained for Text Summarization" www.projectpro.io 12th Oct 2023