DiAna : DIALOGUE ANALYZER

Project Phase II Presentation : Second Review

Guided By : Ms. Aswathy James
Assistant Professor, Department of AD

Prepared By : Batch 2

Mr. George Joyal Vincent , SJC20AD032
Ms. Gouri S Govind , SJC20AD034
Mr. Judin Augustin , SJC20AD045
Mr. Noyal Joseph , SJC20AD049

Outline

Introduction

- In the digital age, multimedia content, including audio and video files, has become an integral part of our daily lives, covering a vast array of topics and purposes.
- As the volume of audio and video data continues to grow, there is a growing challenge in efficiently navigating and extracting meaningful insights from these diverse and expansive sources.
- Our project addresses this challenge through the integration of innovative technologies, such as diarization and summarization, designed to enhance the understanding and accessibility of multimedia content.
- Through accurate speaker identification, personalized summaries, and emotional insights, we aim to elevate the podcast experience, providing users with a more tailored and engaging way to interact with diverse content.

Literature Survey

- **J. Chen and H. Zhuge, ”Extractive Text-Image Summarization Using Multi-Modal RNN,” 2018 14th International Conference on Semantics, Knowledge and Grids (SKG), Guangzhou, China, 2018, pp. 245-248, doi: 10.1109/SKG.2018.00033.[1]**
  - The study aims to summarize documents with both text and images found on the internet, highlighting the importance of summarization beyond just text or images.
  - Created a smart summarization model called MRNN. It uses fancy tech (a type of neural network) to understand both the text and images in documents. This helps it decide which sentences are most important.
  - MRNN beats other advanced summarization methods when tested. The study concludes that including image information makes document summarization way better.

Literature Survey

- **W. Liu, Y. Gao, J. Li and Y. Yang, "A Combined Extractive With Abstractive Model for Summarization," in IEEE Access, vol. 9, pp. 43970-43980, 2021, doi: 10.1109/ACCESS.2021.3066484.[2]**
  - The study introduces a clever summarization model that combines two techniques – extractive and abstractive. It does this in two steps.
  - The model first picks out the most crucial sentences from the document using a neural network and attention mechanism. It pays attention to various things like sentence position, paragraph position, and keywords.
  - After selecting key sentences, the model rewrites and organizes them using a smart algorithm. This process repeats until it gets the best possible summary. The authors believe this method improves the quality of summaries.

Literature Survey

- **S. Horiguchi, Y. Fujita, S. Watanabe, Y. Xue and P. García, ”Encoder-Decoder Based Attractors for End-to-End Neural Diarization,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 1493-1507, 2022, doi: 10.1109/TASLP.2022.3162080.[3]**
  - The study introduces a clever method to organize audio by speakers, crucial for tasks like transcription.
  - EEND-EDA, uses a stacked Transformer neural network, handling various speaker situations during training.
  - This model is flexible. Trained for both fixed and unknown speakers, it fine-tunes on real data, making it versatile for different speech scenarios.

Literature Survey

- **S. Ghodratnama, A. Beheshti, M. Zakershahrak and F. Sobhanmanesh, "Extractive Document Summarization Based on Dynamic Feature Space Mapping," in IEEE Access, vol. 8, pp. 139084-139095, 2020, doi: 10.1109/ACCESS.2020.3012539.[4]**
  - The study introduces ExDoS, a unique method blending supervised and unsupervised techniques for multi-document summarization.
  - ExDoS employs dynamic feature space mapping, extracting key information from input documents to enhance summarization.
  - Evaluation against state-of-the-art methods reveals that ExDoS not only outperforms competitors in both automatic and human-preference assessments but also offers high interpretability for users to comprehend system decisions.

Literature Survey

- **P. K. Biswas and A. Iakubovich, ”Extractive Summarization of Call Transcripts,” in IEEE Access, vol. 10, pp. 119826-119840, 2022, doi: 10.1109/ACCESS.2022.3221404.[5]**
  - The study tailors an extractive summarization method specifically for call transcripts, enhancing readability and understanding of customer-agent conversations.
  - The proposed technique follows a 10-step process, including channel separation, LDA-based topic modeling, and sentence selection, finely tuned for call transcript nuances.
  - Leveraging LDA for topic modeling and advanced word embedding models like Word2Vec and GloVe, the study provides a nuanced approach to summarizing call transcripts.

Gap Identification

- Some studies lack explicit discussions on how their models might handle diverse content types or noise in real-world scenarios, especially in internet documents containing both text and images.
- Multiple studies recognize the challenge of balancing extractive and abstractive summarization techniques, requiring further exploration to optimize the trade-off for improved summary quality.
- Several studies highlight potential challenges related to real-world scalability, computational expense, and latency, suggesting the need for more efficient processing, especially in scenarios with large datasets or real-time requirements.
- Speaker diarization studies reveal challenges related to handling speaker variability, emphasizing the need for adaptability in scenarios with varying numbers of speakers and speaker overlaps.

Objectives

1. To develop a reliable diarization system to precisely identify and segment speakers in podcasts.
2. To implement summarization techniques to create customized, speaker-specific summaries highlighting key contributions.
3. To integrate sentiment analysis to capture and visualize the emotional tone of the media, enhancing the listening experience.
4. To design an intuitive interface for users to explore speaker-specific summaries and sentiment analysis results, ensuring accessibility and usability.

Objectives

- **Beneficiaries of the work :**
  - Listeners/Consumers
  - Content Creators/Producers
  - Researchers/Analysts
  - Content Platforms/Providers
  - Educational Institutions

Materials and Methods

- **pyannote.audio**
  - An open-source Python toolkit for speaker diarization.
  - Built on PyTorch, it employs trainable neural building blocks for diarization.
  - Comes with pretrained models and pipelines, excelling in tasks like voice activity detection and speaker segmentation.
  - Designed to achieve top-tier performance in various speaker diarization domains.

Materials and Methods

- **Open AI whisper**
  - Whisper, an ASR system, is trained on a diverse dataset of 680,000 hours, making it robust to accents, noise, and technical language.
  - It adopts a simple end-to-end Transformer model, splitting audio into 30-second chunks for efficient processing.
  - Despite not specializing in benchmarks like LibriSpeech, Whisper shows remarkable zero-shot performance, making 50% fewer errors across diverse datasets.
  - Whisper excels in transcribing and translating multiple languages, particularly in speech-to-text translation.

Materials and Methods

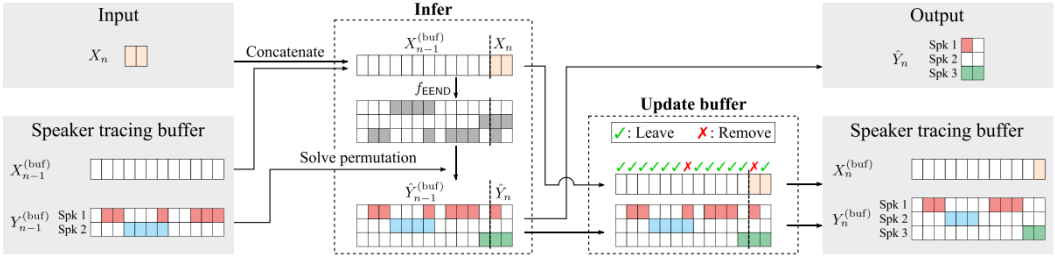

Figure: Online diarization using speaker-tracing buffer proposed in **[8]**

Proposed Study

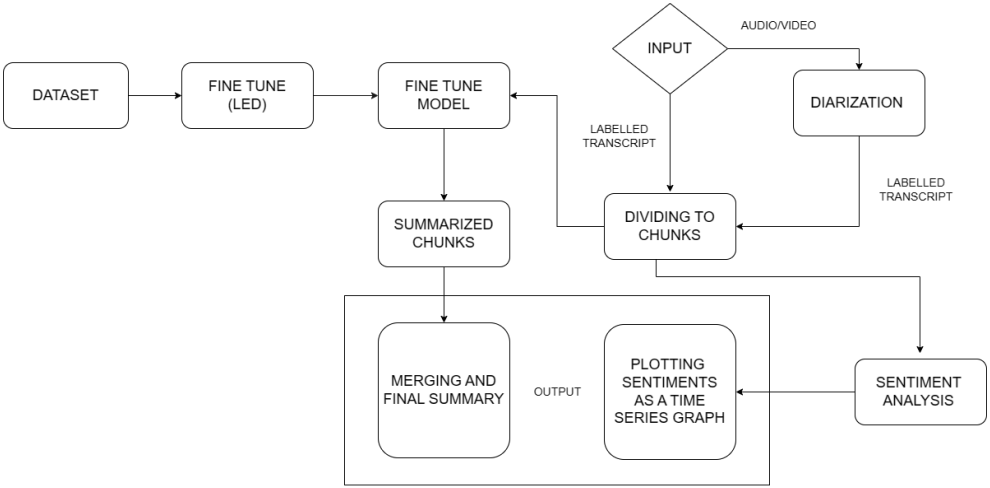

Figure: Proposed System

RESULT AND DISCUSSION

Code : Diarization

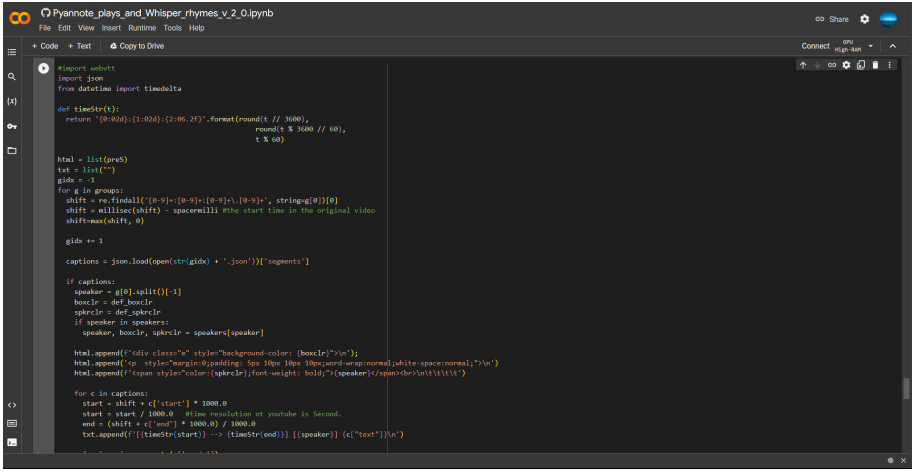

Figure: Code for Diarization

Code : Diarization

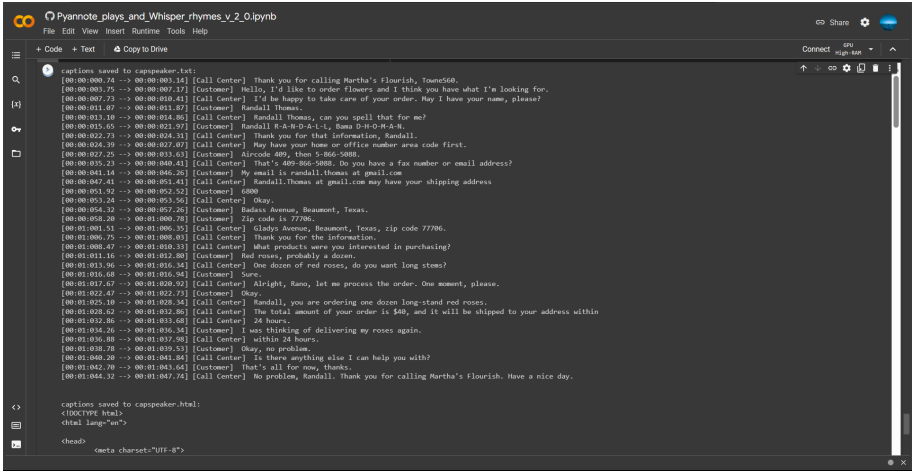

Figure: Code for Diarization

Code : Summarization

```
# Your input text
input_text = """
"Hey, how's it going? Getting into the Christmas spirit yet?"
"Wow, you're really ahead of the game! I'm slowly getting there. Just put up the tree yesterday. It's beginning to look a lot like Christmas."
"Definitely! I'm organizing a Secret Santa gift exchange with my friends, and we're also volunteering at a local soup kitchen on Christmas Eve."
"Sounds like a blast! Oh, speaking of treats, do you have any favorite Christmas recipes?"
"I'm a sucker for peppermint bark. I make a big batch every year and give it out as gifts. It's always a hit!"
"So many! But one that stands out is when my family would gather around the fireplace on Christmas Eve to read 'Twas the Night Before Christmas' before bed. It was so cozy and magical."
"Absolutely. It's not just about the gifts or the decorations, but about the joy of being together and spreading love and kindness."
"""

# Generate summary
summary = summarizer(input_text, max_length=150, min_length=30, do_sample=False)[0]['summary_text']

print(summary)
```

```
"It's not just about the gifts or the decorations, but about the joy of being together and spreading love and kindness" "I'm a sucker for peppermint bark. It's always a hit!"
```

Figure: Code for Summarization

Code : Summarization

```
# Your input text
input_text = """
"Oh, absolutely! I've already decorated the entire house with lights, wreaths, and a big, beautiful tree. How about you?"
"That's great! There's something so magical about the holiday season. Are you planning any special activities this year?"
"That's wonderful! We're hosting a Christmas party for our neighbors, complete with festive treats, games, and maybe even a visit from Santa Claus for the kids."
"Too many to count! But my absolute favorite is my grandma's recipe for gingerbread cookies. They're soft, chewy, and oh-so-delicious. What about you?"
"Yum, that sounds amazing! Hey, do you have any cherished Christmas memories from your childhood?"
"That sounds absolutely enchanting. Christmas really is a time for creating special memories with loved ones."
"Well said! Anyway, I should probably get back to wrapping presents. It was great chatting with you!"
"""

# Generate summary
summary = summarizer(input_text, max_length=150, min_length=30, do_sample=False)[0]['summary_text']

print(summary)
```

```
any special activities this year?" "Oh, absolutely! I've already decorated the entire house with lights, wreaths, and a big, beautiful tree" "My absolute favorite is my grandma's recipe for gingerbread cookies," she say
```

Figure: Code for Summarization

Output : Diarization

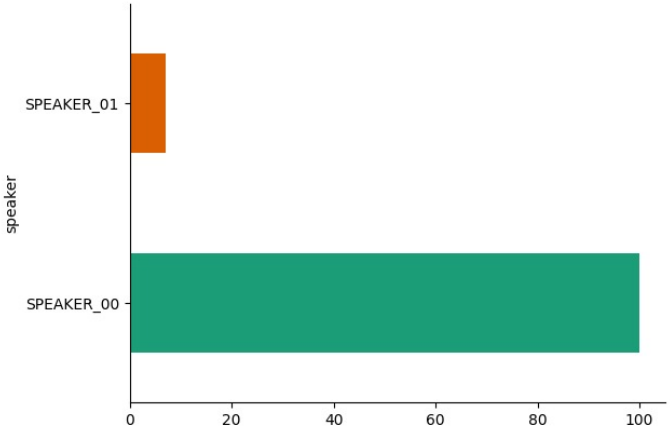

Figure: Graph of Diarization

Output : Sentiment Analysis

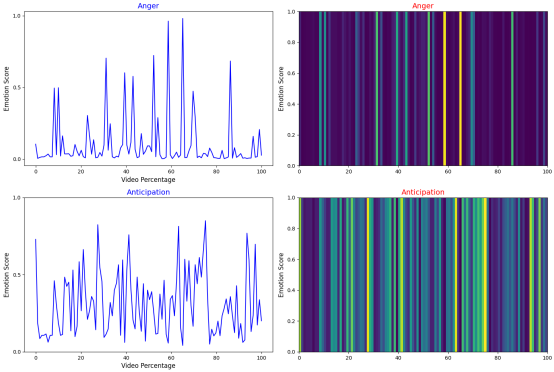

Figure: Speaker Level Sentiment Analysis of Anger and Anticipation

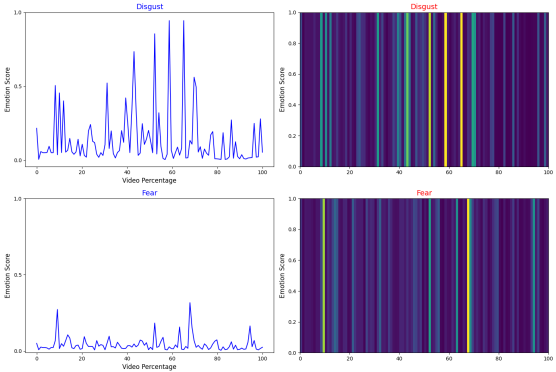

Figure: Speaker Level Sentiment Analysis of Disgust and Fear

Output : Sentiment Analysis

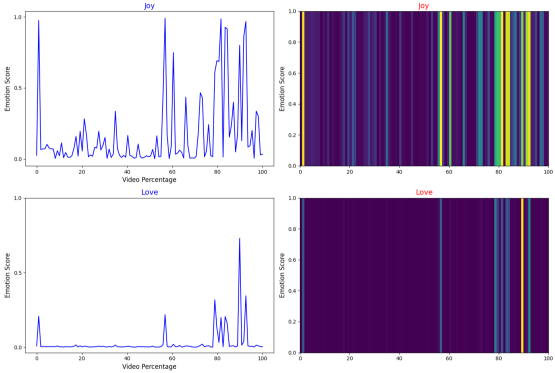

Figure: Speaker Level Sentiment Analysis of Joy and Love

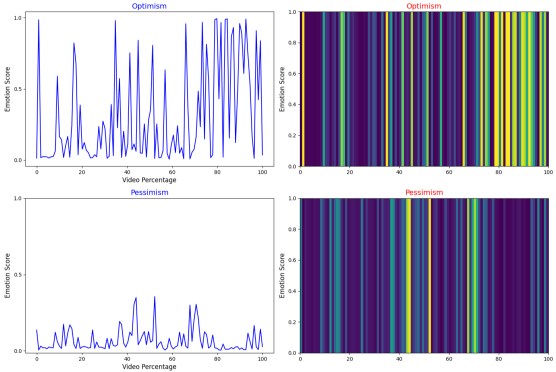

Figure: Speaker Level Sentiment Analysis of Optimism and Pessimism

Output : Sentiment Analysis

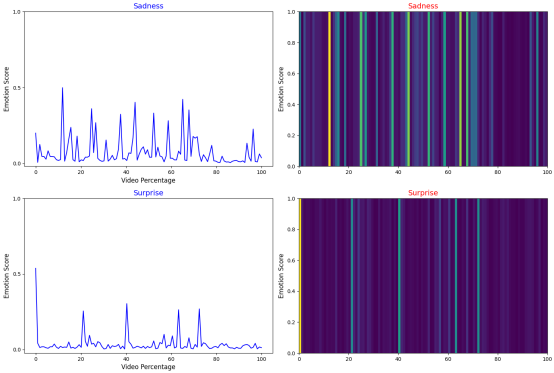

Figure: Speaker Level Sentiment Analysis of Sadness and Surprise

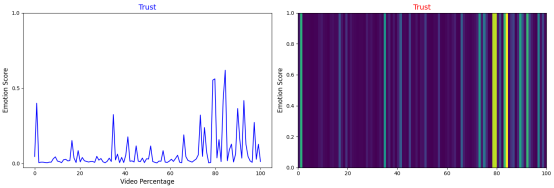

Figure: Speaker Level Sentiment Analysis of Trust

Future Scope

- Investigate ethical considerations in summarization, especially in sensitive domains, and develop frameworks that prioritize fairness, transparency, and unbiased representation in generated summaries.
- Develop diarization models capable of online learning to adapt dynamically to changing speakers and acoustic conditions, allowing continuous improvement over time.
- Address privacy concerns by exploring diarization techniques that respect individual privacy, ensuring that personal information is not compromised during the speaker identification process.
- Extend summarization models to support multiple languages, considering the challenges of language-specific nuances and variations in content structure.

Conclusion

- Our diarization technique effectively distinguishes between two speakers in media, ensuring clear identification.
- The proposed summarization model bridges the gap between human summarization techniques and automated systems, offering a more accurate and efficient approach.
- Sentiment analysis adds emotional depth, allowing users to grasp the overall emotional tone of the media and each speaker's contributions.
- The sentiment graph visually illustrates speaker emotion levels, providing users with a simple yet insightful representation of emotional dynamics throughout the media.

References

[1] J. Chen and H. Zhuge, ”Extractive Text-Image Summarization Using Multi-Modal RNN,”2018 14th International Conference on Semantics, Knowledge and Grids (SKG), Guangzhou, China, 2018, pp. 245-248, doi: 10.1109/SKG.2018.00033.

[2] W. Liu, Y. Gao, J. Li and Y. Yang, ”A Combined Extractive With Abstractive Model for Summarization,” in IEEE Access, vol. 9, pp. 43970-43980, 2021, doi: 10.1109/ACCESS.2021.3066484.

[3] S. Horiguchi, Y. Fujita, S. Watanabe, Y. Xue and P. García, ”Encoder-Decoder Based Attractors for End-to-End Neural Diarization,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 1493-1507, 2022, doi: 10.1109/TASLP.2022.3162080.

References

[4] S. Ghodratnama, A. Beheshti, M. Zakershahrak and F.Sobhanmanesh, ”Extractive Document Summarization Based on Dynamic Feature Space Mapping,” in IEEE Access, vol. 8, pp. 139084-139095, 2020, doi: 10.1109/ACCESS.2020.3012539.

[5] P. K. Biswas and A. Iakubovich, ”Extractive Summarization of Call Transcripts,” in IEEE Access, vol. 10, pp. 119826-119840, 2022, doi: 10.1109/ACCESS.2022.3221404.

[6] M. Ferràs, S. Madikeri, P. Motlicek and H. Bourlard, ”System fusion and speaker linking for longitudinal diarization of TV shows,” 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016, pp. 5495-5499, doi: 10.1109/ICASSP.2016.7472728.

References

[7] A. Dilawari, M. U. G. Khan, S. Saleem, Zahoor-Ur-Rehman and F. S. Shaikh, ”Neural Attention Model for Abstractive Text Summarization Using Linguistic Feature Space,” in IEEE Access, vol. 11, pp. 23557-23564, 2023, doi: 10.1109/ACCESS.2023.3249783.

[8] S. Horiguchi, S. Watanabe, P. García, Y. Takashima and Y. Kawaguchi, ”Online Neural Diarization of Unlimited Numbers of Speakers Using Global and Local Attractors,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 706-720, 2023, doi: 10.1109/TASLP.2022.3233237

[9] A. Phaphuangwittayakul, Y. Guo, F. Ying, W. Xu and Z. Zheng, ”Self-Attention Recurrent Summarization Network with Reinforcement Learning for Video Summarization Task,” 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 2021, pp. 1-6, doi: 10.1109/ICME51207.2021.9428142.

References

[10] H. Suda, D. Saito, S. Fukayama, T. Nakano and M. Goto, ”Singer Diarization for Polyphonic Music With Unison Singing,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 1531-1545, 2022, doi: 10.1109/TASLP.2022.3166262.

[11] 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.

[12] Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu and Xuanjing Huang, ”Extractive summarization as text matching”, 58th Annual Meeting of the Association for Computational Linguistics, July 5-10 2020.

References

[13] G. Soldi, C. Beaugeant and N. Evans, ”Adaptive and online speaker diarization for meeting data,” 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, France, 2015, pp. 2112-2116, doi: 10.1109/EUSIPCO.2015.7362757.

[14] T. Yokota, Q. Zhao and A. Cichocki, ”Smooth PARAFAC Decomposition for Tensor Completion,” in IEEE Transactions on Signal Processing, vol. 64, no. 20, pp. 5423-5436, 15 Oct.15, 2016, doi: 10.1109/TSP.2016.2586759.

[15] M. -H. Su, C. -H. Wu and H. -T. Cheng, ”A Two-Stage Transformer-Based Approach for Variable-Length Abstractive Summarization,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2061-2072, 2020, doi: 10.1109/TASLP.2020.3006731.

References

[16] Q. Wang, C. Downey, L. Wan, P. A. Mansfield and I. L. Moreno, ”Speaker Diarization with LSTM,” 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 5239-5243, doi: 10.1109/ICASSP.2018.8462628.

[17] G. Yapinus, A. Erwin, M. Galinium and W. Muliady, ”Automatic multi-document summarization for Indonesian documents using hybrid abstractive-extractive summarization technique,” 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 2014, pp. 1-5, doi: 10.1109/ICITEED.2014.7007896.

[18] Y. Fujita, S. Watanabe, S. Horiguchi, Y. Xue and K. Nagamatsu, ”End-to-end neu- ral diarization: Reformulating speaker diarization as simple multi-label classification”, ArXiv, 2020.

Thank You!

Questions?