ELSEVIER Scopus doi

Journal of Vibration Engineering

ISSN:1004-4523

Registered

- SCOPUS
- GOOGLE SCHOLAR
- DIGITAL OBJECT IDENTIFIER (DOI)
- IMPACT FACTOR 6.1

Our Website
www.jove.science

Exploring the Power of BART and BERT: Context Similarity andExtractionin NLP

By

HitakshiChellani
SymbiosisUniversityofAppliedSciences,IndoreE-mail:hitakshi.chellani@gmail.com
SauravSingh
SymbiosisUniversityofAppliedSciences,IndoreE-mail:sauravsingh9425@yahoo.com
ShaktiMourya
SymbiosisUniversityofAppliedSciences,IndoreE-mail:mouryashakti@gmail.com
DevendraChouhan
SymbiosisUniversityofAppliedSciences,IndoreE-mail:devendra.chouhan@suas.ac.in

Abstract

Inthispaper,weproposeanewpipelineforextractingcontextandanalyzingsemanticsimilarity from audio or video data. The pipeline consists of four main stages: first,video-to-audioconversionusingtheMoviepylibrary;second,speech-to-texttranscription using a pre-trained whisper model; third, text summarization using theBART-based "facebook/bart-large-cnn" model from the Hugging Face Transformerslibrary; and finally, similarity analysis using the BERT model from the same library.Ourpipelinehasnumerousapplicationsinfieldssuchasnaturallanguageprocessing ,information retrieval, and search engine algorithms. To validate the effectiveness ofour pipeline, we conducted experiments on a dataset of audio and video files anddemonstrated its high accuracy in context extraction and similarity analysis. Ourresearchprovidesausefulframeworkforfuturestudiesinnaturallanguageprocessingandinformationretrievalandhighlightsthepotentialofpre-trainedmodelsforsolvingcomplexNLPtasks.

Keywords: Whisper model; Semantic similarity; Meeting summary; BART; BERT;Textualdata extraction

1. Introduction

In today's fast-paced business environment, meetings are a common occurrence in organizationsof all sizes. Meetings are essential for effective communication, decision-making, and problem-solving. However, keeping track of the discussions and decisions made during meetings can bechallenging,especiallywhendealingwithlargevolumes ofinformation.

We provide a solution to this problem so that individuals won't have to listen to the entire tape orrisk missingtheimportantinformationwhilehearingitfromothers.Thisway,theycanobtainthesummary of the meeting they weren't able to attend. Our system involves a pipeline that employsvideo-to-audioconversion,audio-to-texttranscription,textsummarization,andtextcomparisontoextractimportantinformationfrommultimediadata.

The initial step of the pipeline involves converting video to audio, which allows for easierprocessing of the multimedia data. Next, an accurate whisper model is utilized for audio-to-texttranscription.Thefacebook/bart-large-cnnmodelofHuggingFaceisthenappliedfortextsummarizationinthethirdstep,whichcanquicklyandaccuratelygenerateasummaryofthetext.Finally, the fourth step involves comparing two texts using BERT, a state-of-the-art naturallanguageprocessingmodelthatcan identifysimilaritiesand differencesbetweenthetwo texts.

The objective of this research paper is to evaluate the effectiveness of this pipeline in extractingcrucial information from multimedia data. To test the accuracy of the pipeline, we will compareits output with manually generated summaries of the same data. Additionally, we will explore thepotential applications of this pipeline in various fields such as journalism, social media analysis,andmarketresearch.

In summary, this pipeline provides an automated and efficient approach to process multimediadata. This research paper aims to contribute to the development of more accurate and efficientmethodsforprocessingmultimediadata.

2. BasicDefinitions,PreliminariesandNotations

2.1. Natural Language Processing (NLP) [15]: This field of study focuses on the interactionbetween human language and computers. It involves the development of algorithms andmodelsforprocessing,understanding,and generatingnaturallanguage data.

2.2. Information Retrieval (IR): Refers to the process of retrieving information from a set ofdocumentsordatabasedon userqueries.Itemploystechniquessuch asindexing,searching,andrankingtoidentifythemostrelevantdocumentsfor agivenquery.

2.3. Video-to-Audio Conversion: The process of extracting audio content from video data,whichcanbe achievedusinglibrariessuchasMoviepyorFFMPEG.

2

2.4. Speech-to-Text Transcription [14]: The process of converting spoken language into text.This involves utilizing models such as deep neural networks to recognize and transcribespeech.

2.5. Text Summarization: The process of generating a summary of a document or text. Thiscanbedoneusingtechniquessuchasextractivesummarization,whereimportantsentences or phrases are selected from the original text, or abstractive summarization,whereanew summary isgeneratedusing naturallanguage generationtechniques.

2.6. SemanticSimilarity[10]:Ameasureofhowsimilartwopiecesoftextareintermsoftheirmeaning . This is often computed using models such as BERT or Word2Vec, whichgenerateembeddingsthatcapturethe semanticmeaningofwordsorphrases.

2.7. Paraphrase Identification: The task of determining whether two pieces of text convey thesame meaning, even if the words used are different. This is often accomplished usingmodelssuchas BERTorSiamesenetworks.

2.8. BERT:BidirectionalEncoderRepresentationsfromTransformersisapre-trainedlanguage model that has achieved state-of-the-art performance on a wide range of naturallanguage processingtasks.

2.9. HuggingFaceTransformers[5]:Apopularlibraryforworkingwithpre-trainedmodelsinnaturallanguageprocessing. Thelibraryprovides awiderangeofpre-trainedmodelsandtoolsforfine-tuningthemonspecifictasks.

2.10. Pipeline: A sequence of steps or operations performed in a specific order to achieve aparticular goal. In this paper, we propose a pipeline for context extraction and similarityanalysisfromvideooraudiodata.

2.11. Pre-trainedModel:Amodeltrainedonavastamountofdatathatcanbeusedasastartingpointforfurthertrainingorfine-tuningonspecifictasks

3. MethodologyandMainWork

3.1. **Data Collection:** The first step in the proposed methodology is data collection. Thisinvolves collecting video recordings that will be used for further processing. The videorecordings can be obtained from various sources, such as public archives or privatecollections.

3.2. **Video-to-audio conversion using the Moviepy library** – The first stage of our pipelineinvolves converting video to audio using the Moviepy library. It is a Python library thatfacilitates video editing, manipulation, and processing. It offers a user-friendly interfaceforworkingwithvideofilesandiscapableofperformingavarietyoftasks, rangingfrombasiceditingtointricatecompositingandspecialeffects.Thelibraryisbuiltontopof

3

NumPyandthePythonImagingLibrary(PIL),whichallowsittobeseamlesslyintegratedwithothers cientific computinglibraries.

3.3. Speech-to-text transcription using a pre-trained whisper model [17] – The secondstage of our pipeline involves using a pre-trained Whispermodel for speech-to-texttranscription. Whisper is an automatic speech recognition (ASR) system that has beentrained on a vast dataset of 680,000 hours of multilingual and multitask supervised datacollected from the internet. It is capable of transcribing speech in multiple languages andtranslating them into English. The Whisper architecture is built on a simple end-to-endapproachandimplementedasanencoder-decoderTransformer.Theaudioinputisdividedinto 30-second chunks, converted into a log-Mel spectrogram, and then fed into anencoder. A decoder is used to predict the corresponding text caption, interspersed withspecial tokens that guide the single model to perform various tasks, such as languageidentification, phrase-level timestamps, multilingual speech transcription, and speech-to-Englishtranslation.

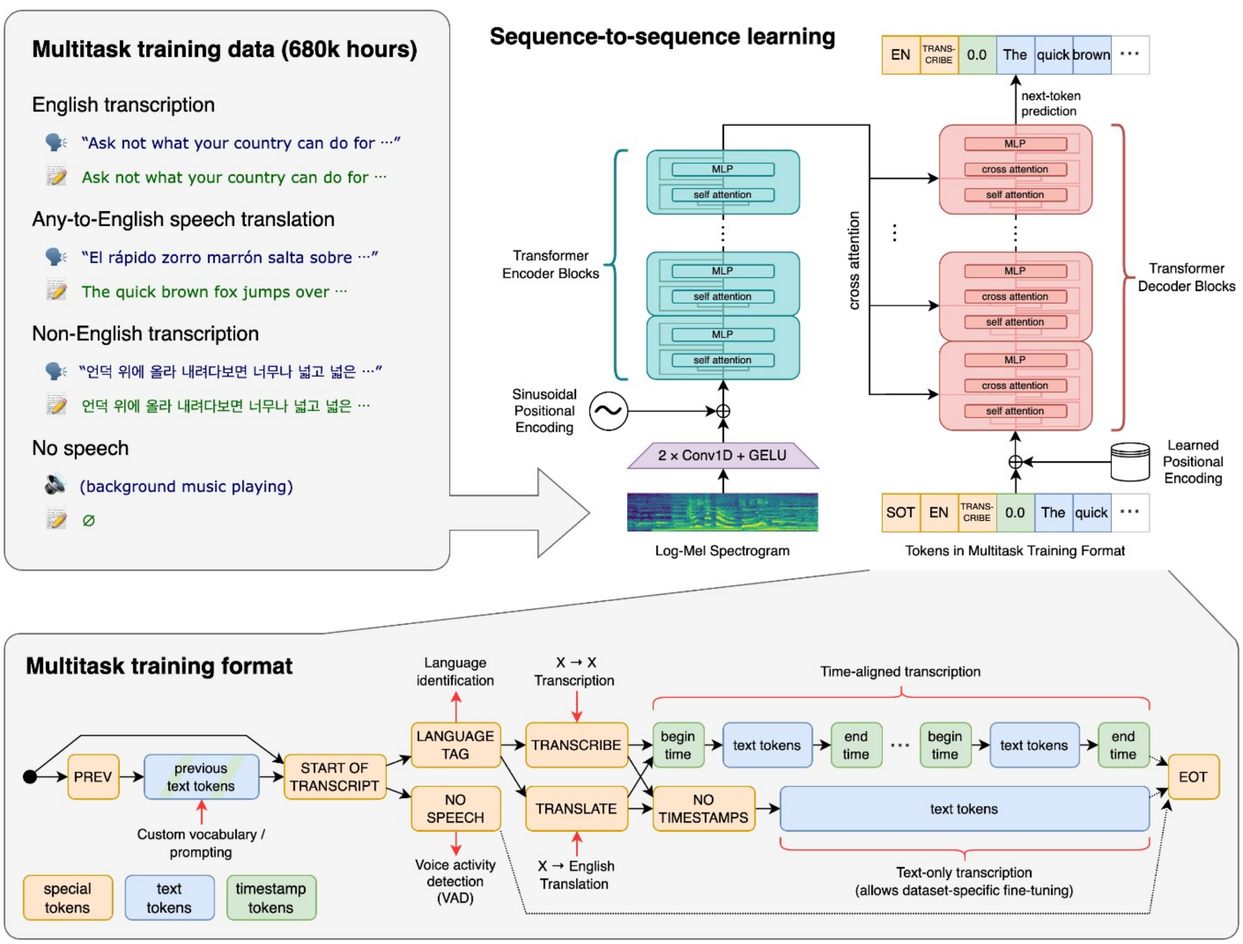


Figure1. Illustratesanoverviewofthetransformerapproach.

3.4. Text summarization using the BART-based "facebook/bart-large-cnn" model – Thethird phase of out pipeline involves using the facebook/bart-large-cnn summarizationmodelfromHuggingFaceTransformerslibrary'.BART[1,7,11](Bidirectional andAuto-RegressiveTransformers)isaversatilesequence-to-sequencemodelintroducedbyFacebookAIResearch.ItcombinesTransformerarchitecturewiththetechniquesfromautoregressive language modeling and denoisingautoencoders. BART excels in variousnaturallanguageprocessingtaskssuchastextgeneration,summarization,translation,andquestion answering. Its unique feature is the ability to handle both autoregressive andbidirectionaltasks,makingitadaptableandeffectiveforspecificapplications.

The input text was first converted into tokens, and then the first chunk of tokens wasselected. If the length of the chunk exceeded the maximum input length of the model, itwas appended to a list of chunks that would be processed separately. To ensure that eachchunk remained coherent and meaningful, we trimmed the chunk to a length of 1024tokens and located the last full stop before this point. The index of the last full stop wasstored in a variable, and the chunk appended in the list was replaced with the new chunkthatendedatthispoint.This processwasrepeateduntiltheendoftheinputtext.

Oncetheinputtextwassplitintochunks,weconvertedeachchunkback intoastringandpassed it to the model to generate a summary. By looping through each chunk in the listand processing them separately, we were able to generate summaries for longer texts thatexceeded the maximum input length of the model. It's worth noting that in cases wherethe last full stop in a chunk occurred after the 1024th token, we adjusted our strategy forfinding the last full stop or split the chunk at a different point to ensure that the input tothe modelwas nottruncatedinthemiddleofa sentence.

Overall, our chunking strategy allowed us to generate summaries for longer texts thatexceededthemaximuminputlengthofthefacebook/bart-large-cnnmodel,whilemaintaining coherence andrelevanceinthe generatedsummaries.

3.5. Similarity analysis using the BERT [2,3,8] model – The last stage of the pipelineinvolvesusingtheBERTmodelfromthesamelibrarytodeterminethesemanticsimilaritybetween two given sentences. The code utilizes the BERT Tokenizer to encode text, andweusethebase-base-uncasedpre-trainedmodel.TheBert SemanticDataGeneratorclassgenerates batches of data with an option to include or exclude labels based on whetherthey are used for training/validation or testing. After encoding the sentence pairs with thetokenizer, the model predicts the similarity between the two sentences by generatingprobabilities for each of the three labels: contradiction, entailment, and neutral. Thepredict() function produces the predicted probability for each label, which can then beused to determine the semantic similarity between the two sentences. This stage of ourpipeline is critical in determining the similarity between two given sentences, and itsapplicationscanextendtovariousnaturallanguageprocessingdomains,suchasinformationretrievalandquestion-answeringsystems.

4. NumericalExamplesandResults

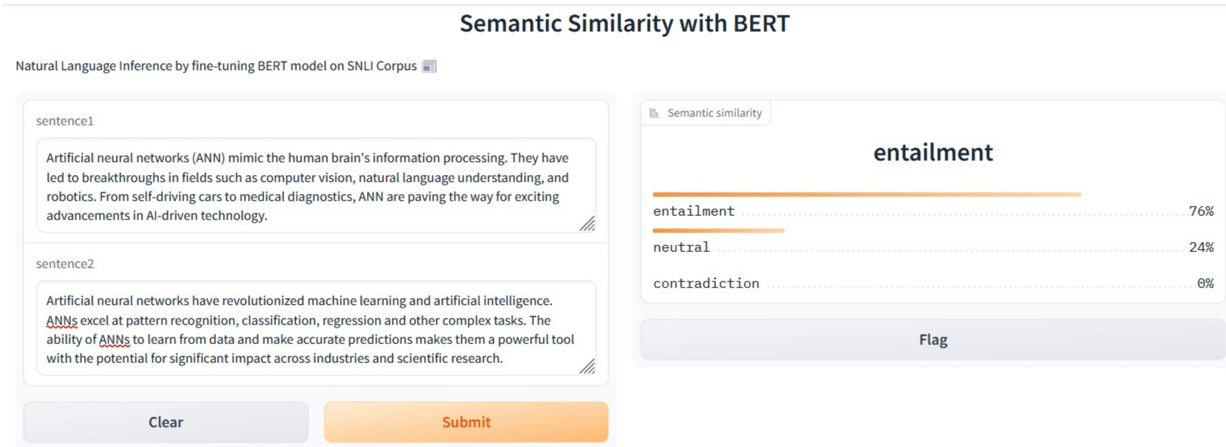
4.1.Audio-to-textrouge[16]scoretable- Thetablebelowshowshowwellthewhispermodelperformed while converting audio to text. The output generated by the whisper model iscomparedwiththegroundtruthofthemeetingandthewellnessofthemodelisdeterminedby rouge score which is a value between [0,1]. We get an average of 0.8-0.9 using thewhispermodel,whichisconsideredquietgood.

	Generated_transcript	Ground_truth	rouge1	rouge2	rougeL	rougeLsum
0	Hello. Hello. Hi, Akshay. How are you? I'm good. So today we're going to talk. Yes, yes, continue. Today we're going to talk about artificial neural networks. That is ANN. ANN has become the cornerstone of the world. The ability to mimic the human brain's information processing has lead to breakthroughs in fields such as computer vision, natural language understanding, and robotics. ANN consists of interconnected layers of neurons that connect the brain to the brain. The brain is a complex structure that is connected to the brain. The brain is a complex structure that is connected to the brain. ANN consists of interconnected layers of neurons that perform calculations on input data, gradually extracting meaningful features, and generating output predictions. The strength of ANN lies in their ability to learn and adapt through a process called training, where they adjust their internal parameters based on labeled data set. This enables ANN to generalize from examples and make accurate predictions on unseen data. From self-driving cars to medical diagnostics, ANN are transforming industries and paving the way for exciting advancements in AI-driven technology. That was a wonderful session. Thank you so much. you	Hello, Hello. Hi, Takshi, how are you? Hi, Saurav. I'm good. So today we're going to talk, Yes, yes, continue. Today we're going to talk about artificial neural networks. That is Ann. Ann have become the cornerstone of modern AI applications. Their ability to mimic the human brains information processing has lead to breakthroughs in fields such as computer vision, natural language understanding, and robotics. Ann consists of interconnected layers of neurons that perform calculations on input data, gradually extracting meaningful features and generating output predictions. The strength of Ann lies in their ability to learn and adapt through a process called training, where they adjust their internal parameters based on labeled data set. This enables Ann to generalize from examples and make accurate predictions on unseen data. From self driving cars to medical diagnostics, Ann are transforming industries and paving the way for, Exciting advancements in AI driven technology. That was a wonderful session. Thank you so much.	0.847645	0.818942	0.847645	0.847645
1	Hello, Takshya. Good morning. Hello. Good morning, Sourav. So, as last time we were talking about artificial neural networks, we're going to continue with that discussion. So, artificial neural networks have revolutionized machine learning and artificial intelligence. ANNs excel at pattern recognition, classification, regression and other complex tasks. During the training process, ANNs learn to recognize patterns and make predictions by adjusting the weights and biases associated with connections between neurons. This process, known as backpropagation, involves iteratively feeding the network with input data and comparing its output with desired output. With advances in deep learning, ANNs with numerous layers known as deep neural networks have achieved remarkable success in various domains such as image and speech recognition, natural language processing and autonomous vehicles. The ability of ANNs to learn from data and make accurate predictions makes them a powerful tool with the potential for significant impact across industries and scientific research. Okay. So, thank you so much for this session. Bye.	Hello, Takshay. Good morning. Hello, Good morning, Sourav. So as last time we were talking about artificial neural networks, we're going to continue with that discussion. OK. So artificial neural networks have revolutionized machine learning and artificial intelligence. Ann's excel at pattern recognition, classification, regression and other complex tasks. During the training process, Ann learn to recognize patterns and make predictions by adjusting the weights and biases associated with connections between neuron. This process, known as backpropagation, involves iteratively feeding the network with input data and comparing its output with desired output. With advances in deep learning, Ann with numerous layers known as deep neural networks, have achieved remarkable success in various domains such as image and speech recognition, natural language processing and autonomous vehicle. The ability of Anns to learn from data and make accurate predictions makes them powerful tool with the potential for significant impact across industries and scientific research. Okay so thank you so much for this session. Bye.	0.954128	0.904615	0.954128	0.954128

4.2.Text to summary rouge score table - Again, we use rouge score to determine thesummarygenerated byourfine-tuned BARTmodel,presentedinthetablebelow.

	Generated_summary	Ground_truth	rouge1	rouge2	rougeL	rougeLsum
0	Artificial neural networks (ANN) mimic the human brain's information processing. They have led to breakthroughs in fields such as computer vision, natural language understanding, and robotics. From self-driving cars to medical diagnostics, ANN are paving the way for exciting advancements in AI-driven technology.	Artificial Neural Networks (ANN) have revolutionized AI applications by mimicking human brain processes. ANN's ability to learn, adapt, and make accurate predictions through training has led to breakthroughs in computer vision, natural language understanding, and robotics. They are transforming industries such as self-driving cars and medical diagnostics, paving the way for exciting advancements in AI technology.	0.750000	0.470588	0.615385	0.615385
1	Artificial neural networks have revolutionized machine learning and artificial intelligence. ANNs excel at pattern recognition, classification, regression and other complex tasks. The ability of ANNs to learn from data and make accurate predictions makes them a powerful tool with the potential for significant impact across industries and scientific research.	Artificial neural networks (ANNs) have revolutionized machine learning by excelling in pattern recognition and prediction tasks through backpropagation. Deep neural networks (DNNs) with multiple layers have achieved impressive success in various domains. ANNs have the potential for significant impact across industries and scientific research.	0.516129	0.329670	0.473118	0.473118

4.3.Comparison of summaries - Here, we present the results of the comparison of thesummariesgeneratedbythefine-tunedBARTmodel,thatisdoneusingBERT.Themodeltakes in input two paragraphs or sentences and tells us the extent to which they are relatedtoeachother.Itratestheinputsusing3parameters,i.e.entailment,neutralandcontradiction.Here'sanexampleofthetwomeetingswerecorded,convertedthemtotext,generated their summaries, and finally compared them. Both the meetings were aboutdiscussing the topic 'ANN' and hence we find similarity between the two generatedsummaries.



5. Conclusion

Inthispaper,wehavedevelopedameetingsummarizationsystemthatisabletoautomatically generate summaries of meetings. The system collects meeting data from varioussources,preprocessesthedata,andappliesasummarizationmodeltocreateconcisesummariesof the most important points discussed during the meeting. The summaries are then compared toidentifycommonthemesorpatterns,and the qualityofthesummarizationmodelisevaluated.

Through the development of this system, we have shown that meeting summarization can be aneffectivewaytohelpstakeholdersstayinformedaboutthecontentofmeetings,withouthavingtospend thetimeandresourcesnecessary toattendeverymeetinginperson.By generatingsummaries automatically,the system is able to improve the efficiency and effectiveness ofcommunicationamongstakeholders.

However, it is important to note that the effectiveness of the meeting summarization systemdepends on the quality of the summarization model. As such, it is important to continue to refineand optimize the model to ensure that it is able to accurately capture the most important pointsdiscussedduringthemeeting.

Overall, the meeting summarization system developed in this paper represents an important steptowardsimprovingcommunicationandcollaborationamongstakeholders.Byprovidingconcise summariesofmeetings,thesystemmakesiteasierforstakeholderstostayinformedandmakemoreinformedd ecisionsbasedonthecontentofthemeetings.

References

[1] M. Bewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, VesStoyanov,Luke Zettlemoyer, “BART: Denoising Sequence-to-Sequence Pre-training for NaturalLanguage Generation, Translation, and Comprehension”, Proceedings of the 58th AnnualMeeting ofthe Association forComputationalLinguistics,pages7871–7880.

[2] J. Devlin, M. Wei, C. Kenton, L. Kristina Toutanova, “BERT: Pre-training of DeepBidirectional Transformers for Language Understanding”, Proceedings of NAACL-HLT2019,pages4171–4186Minneapolis,Minnesota,June2-June7,2019.c2019AssociationforComputationalLinguistics.

[3] K. Fleming.. “BERT: A Review of Applications in Natural Language Processing andUnderstanding”.10.48550/arXiv.2103.11943(2021).

[4] A. Gadford, J. Wook Kim, T. Xu, G. Brockman, Christine McLeavey, IlyaSutskever,“RobustSpeechRecognitionviaLarge-ScaleWeakSupervision”.

[5] L. Jones, A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, A. N. Gomez, Ł. Kaiser, I.Polosukhin."Attentionisallyouneed."*Advancesinneuralinformationprocessingsystems*30(2017).

[6] R.Kils,I.Gurevych."Sentence-bert:Sentenceembeddingsusingsiamesebert-networks." *arXivpreprintarXiv:1908.10084*(2019).

[7] Y. Lang, Y. Sun, J. Dai, X. Hu, Q. Guo, X. Qiu, X. Huang. "BART-Reader: PredictingRelations Between Entities via Reading Their Document-Level Context Information." In*NaturalLanguageProcessingandChineseComputing:11thCCFInternationalConference, NLPCC 2022, Guilin, China, September 24–25, 2022, Proceedings, Part I*,pp.159-171.Cham:SpringerInternationalPublishing,(2022).

[8] E. Maufiq, L. Zeratul, I. M. Yusoh, B. M. Aboobaider. "Enhancing the Takhrij Al-HadithbasedonContextualSimilarityusingBERTEmbeddings."*InternationalJournalofAdvancedComputer Science andApplications*12.11(2021).

[9] S. Michel, L. Chi-kiu, "Fully unsupervised crosslingual semantic textual similarity metricbased on BERT for identifying parallel data." *Proceedings of the 23rd Conference onComputationalNaturalLanguage Learning(CoNLL).*(2019).

[10] D. Mila.“Semantic Similarity- A Review of Approaches and Metrics”. InternationalJournalofAppliedEngineeringResearch.(2019).

[11] A.Mundara,C.Mankar,S.Nagrale,P.Malviya,A.Sangle,M.Navrange,“BARTModelfor Text Summarization : An Analytical Survey and Review”, International Journal ofAdvancedResearchinScience,CommunicationandTechnology(IJARSCT),Volume2, Issue5,May2022

[12] W.Niajie.,“LiteraturereviewonvulnerabilitydetectionusingNLP technology”(2021)..

[13] G. Raima, R. Stefan, S. Yücel. “Context-based Extraction of Concepts from UnstructuredTextualDocuments”.InformationSciences.588.10.1016/j.ins.2021.12.056(2021)..

[14] V. Ronit , M. Saraswathi , S. S. Pranav, “Implementation of Video and Audio to TextConverter”, International Journal of Research Publication and Reviews, Vol 4, no 5, pp1204-1208May2023

[15] E. Sambria, B. White, "Jumping NLP Curves: A Review of Natural Language ProcessingResearch [Review Article]," in IEEE Computational Intelligence Magazine, vol. 9, no. 2,pp.48-57,May2014,doi:10.1109/MCI.2014.2307227.

[16] L. C. Yenk, “ROUGE: A Package for Automatic Evaluation of Summaries”, Appl. Math.Comput,(2022).

[17] R.A.Yuhan,"Robustspeechrecognitionvialarge-scaleweaksupervision."arXivpreprintarXiv:2212.04356(2022).