ICEST-2K24 [ International Conference on Engineering, Science and Technology ]
In Association with
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN - 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

# A Comprehensive Multimodal Analysis of Research Papers through Natural Language Processing (NLP) and Deep Learning

**Pankaj Sonawane, Hetvi Shah, Sahil Doshi, Aayush Parikh**
Department of Computer Engineering, SVKM's Dwarkadas J. Sanghvi College of Engineering Mumbai, India

## ABSTRACT
The project introduces an innovative automated system poised to transform research paper analysis across diverse academic fields, offering dynamic summaries and flow charts through cutting-edge technologies like deep learning and natural language processing (NLP). Its core aim is to enhance the accessibility and understanding of scholarly literature by swiftly extracting key insights from dense academic texts. Leveraging deep learning-based NLP, the system generates concise summaries, expediting the review process for academics and researchers. Additionally, employing computer vision and deep learning algorithms, it creates comprehensive flow charts that aid in visualizing complex concepts within research papers, promoting better comprehension. Users can interact with these flow charts to intuitively explore relationships and significant elements within the papers. Notably, the system meets usability standards, allowing users to input research articles and access the generated summaries and flow charts. By amalgamating mind mapping with dynamic summarization, this groundbreaking project addresses the challenges posed by the exponential growth of scholarly research, underscoring its commitment to facilitating efficient and thorough research consumption for academics, researchers, and institutions alike.
Keywords— BERT Bidirectional Encoder Representations from Transformers, BART Bidirectional and Auto-Regressive Transformers, ROUGE Recall-Oriented Understudy for Gisting Evaluation, BLEU Bilingual Evaluation Understudy

## I. INTRODUCTION

Our project tackles the challenge of navigating the vast expanse of scholarly literature by introducing an advanced automated system for research paper analysis. Utilizing state-of- the-art deep learning techniques, our system aims to enhance accessibility and comprehension of research papers through several key features. Employing advanced natural language processing methods, it dynamically generates summaries tailored to researchers' needs, significantly expediting the review process. Additionally, leveraging computer vision and deep learning algorithms, the system creates hierarchical mind maps that visually represent the paper's content, facilitating better comprehension of complex ideas. It also provides keyword extraction and relevant reference identification, further aiding in understanding the paper's essence. By seamlessly integrating flow chart generation with dynamic summarization, our project represents a paradigm shift in research paper analysis, offering researchers greater efficiency and customization in navigating academic literature, thus advancing the utilization of deep learning for intelligent and accessible research consumption.

**Copyright © 2024 The Author(s):** This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)**

## II. LITERATURE REVIEW

Recent advancements in natural language processing (NLP) have spurred significant developments in text summarization techniques. Several studies have contributed diverse methodologies to address the challenges inherent in summarizing large volumes of text. Tham Vo's work introduces a pioneering text graph-based summarization model known as TGA4ExSum. By integrating bidirectional attention auto- encoding and multi-headed attention mechanisms, Vo's model demonstrates notable efficacy in summarizing textual data, outperforming conventional baselines. The utilization of graph- based architectures underscores the importance of contextual and structural representations in enhancing summarization tasks across various NLP benchmarks. Tong Chen et al. present an innovative approach to abstractive summarization by incorporating knowledge graphs into the summarization process. Through the integration of large language models (LLMs) and transformer modules, Chen et al.'s method generates summaries of high informativeness and relevance. The incorporation of both textual and graphical inputs enriches the summarization process, promising improved quality in summarization outputs. Semantic representation forms the crux of Georgios Alexandridis et al.'s work, which introduces a deep learning-based approach to summarization. Their model, leveraging a deep encoder-decoder architecture, excels in capturing the nuanced semantic meanings of words, resulting in the production of comprehensive summaries. The emphasis on semantic understanding signifies a shift towards more contextually relevant summarization outputs. Citation graph- based summarization is explored by Chenxin An et al., who introduce the CGSUM model. Through the integration of information from source papers and references, CGSUM surpasses existing methods in summarization quality, underscoring the critical role of citation graph structures in summarization tasks. Mohamed Elhoseiny and Ahmed Elgammal propose an automated approach to generate MindMaps from textual data, utilizing Detailed Meaning Representation (DMR). While effective, their method faces challenges in parsing complex statements and assuming central ideas, necessitating further research in hierarchical abstract information generation and word vector utilization. Mengting Hu et al. present a graph-based summarization approach that combines the LexRank algorithm and DistilBERT. Their methodology, characterized by superior summarization quality and efficiency, showcases the potential of lightweight BERT alternatives in streamlining summarization tasks. By integrating diverse methodologies such as graph-based architectures, semantic understanding, and knowledge graph integration, these studies collectively contribute to advancing the state-of-the-art in text summarization, paving the way for more sophisticated and contextually relevant summarization techniques.

## III. METHODOLOGY

*A.* Summarization:

BART: An inventive sequence-to-sequence paradigm called BART (Bidirectional and Auto-Regressive Transformers) was created for challenges involving natural language processing. BART generates text that is coherent and rich in context by combining bidirectional and auto-regressive training objectives, which sets it apart from typical autoregressive models. BART gains a flexible knowledge of language through pre-training on large-scale corpora utilising denoising autoencoder objectives, where input sentences are distorted and the model is trained to rebuild them. BART's flexible comprehension enables it to perform a wide range of downstream tasks, including language translation and text summarization. BART is a strong and adaptable model for a variety of natural language processing applications because of its bidirectional and auto-regressive fusion, which helps it capture complex connections withinsequences.

BERT: Contextualised word embeddings are transformed by the ground-breaking natural language processing algorithm BERT (Bidirectional Encoder Representations from Transformers). Because BERT uses a bidirectional transformer design, it may take into account a word's whole context within a sentence, in contrast to standard models that analyse text in a unidirectional fashion. BERT acquires deeply contextualised representations that capture complex syntactic and semantic links through pre- training on huge corpora and concealed language modelling aims. With the help of this contextual knowledge, BERT surpasses earlier state-of-the-art techniques in a variety of downstream tasks, including sentiment analysis, question answering, and named entity recognition. The success of BERT emphasises how important bidirectional context modelling is to improve language comprehension and performance in various natural language processing applications.

*B.* Flow chart generation:

The Python function {flowchartMaking} creates flowcharts automatically from text summaries using cutting-edge NLP techniques. It utilises a state-of-the-art architecture developed by Facebook AI called the BART (Bidirectional and Auto- Regressive Transformers) model. This model is particularly good at condensing lengthy paragraphs into concise summaries without sacrificing crucial details; it may be obtained via the Hugging Face {transformers` library. By using BART, the function ensures that the output summaries contain the most crucial information from the input text.

After obtaining the summary, the function uses NLTK's `sent_tokenize` function to divide it into individual sentences. Because it enables the creation of a flowchart later on that illustrates how the information in the summarised text makes sense, this stage is crucial for granular analysis. The function's primary skill is its use of graph theory to create visual representations from textual summaries. The directed graphs are created from the summary sentences, where each sentence is a node. Depending on the semantic content of the phrase, phrases containing conditional assertions, or "if" conditions, are represented as decision nodes (diamond-shaped), whilst other sentences are represented as action nodes (rectangle-shaped). This differentiation facilitates readers' comprehension of the flowchart by enabling them to discern between decision points and actionable processes.[7]

Furthermore, directed edges are used by the function to establish relationships between nodes, representing the thoughts in the summary in chronological order. This visual tool expedites the process of comprehending the condensed information while also organising the data logically. Once the process is complete, the created flowchart is displayed using matplotlib, and it is saved as a PNG image with the file name "flowchart1.png." This graphic output enables users to comprehend the summary text's organisation and flow intuitively, facilitating efficient analysis and decision-making. In conclusion, by automating the transformation of textual summaries into visual flowcharts using a combination of complex NLP algorithms and graph theory concepts, the `flowchartMaking` function enhances the accessibility and usefulness of summarised material.

*C.* Keyword Extraction:

The provided code makes excellent use of the RAKE (Rapid Automatic Keyword Extraction) technique to extract keywords from a given text by utilising the `rake_nltk' library. Depending on how frequently and pertinently they occur in the text, this algorithm utilises an advanced way to identify possible keywords and assigns them a score. The `generate_keywords` function encapsulates this process and offers a useful method for extracting essential concepts and themes from textual input. Readers can obtain a brief synopsis of the main points of the text.

Through its interaction with the `rake_nltk} library, the code streamlines the keyword extraction process, enabling users to acquire insightful information about the text content more quickly. When everything is said and done, this code provides a powerful tool for automating the extraction of pertinent keywords, improving the analysis and interpretation of textual material.

*D.* Reference Similarity Matching:

The `find_references` function can be used to retrieve the "References" section from a text. It starts by looking for the term "references" somewhere in the text, regardless of case sensitivity. The text that follows this sentence is recorded if it can be found, with the assumption that this material is the references section. This section is then divided into separate reference points by newline characters. Every point undergoes a process to remove leading and trailing whitespaces. Additionally, any empty strings that remain after splitting are eliminated.

The `find_most_similar_points` function then discovers, from a set of reference points, the reference points that are most similar to a target text. The degree of similarity is calculated by counting the number of common words (independent of case) between each reference point and the target text. These similarity scores, which are stored in a dictionary, map each reference point to its matching score. The vocabulary is then arranged in descending order using these scores. Points with zero similarity are filtered out to ensure that only relevant matches are considered. If there are fewer similar points than the specified count, the function adds dissimilar points to the list. Finally, the function offers a meticulously curated list of the most similar reference points, along with an analysis of their significance to the text.

The code extracts reference points from a text that has a "References" section and a target text using `find_references`. The five reference points that are most similar to the target text are then identified using {find_most_similar_points}. This is used for content comparison or information retrieval, where users can find relevant references within a corpus.

## IV TESTING & TEST CASES

*A.* Test Cases:

Input 1:

```
Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics , pages 5082–5092
Florence, Italy, July 28 - August 2, 2019. Initially, a theoretical model for
semantic-based text generalization is intro-
duced and used in conjunction with a deep
encoder-decoder architecture in order to pro-
duce a summary in generalized form. Sub-
sequently, a methodology is proposed which
transforms the aforementioned generalized
summary into human-readable form, retaining
at the same time important informational as-
pects of the original text and addressing the
problem of out-of-vocabulary or rare words. The overall approach is evaluated on two pop-
ular datasets with encouraging results. In the latter case, new sen-tences are generated which concatenate the ove
all meaning of the initial text, rephrasing its con-
tent.
```

Fig. 1 Output Summary 1

Output Flowchart 1:

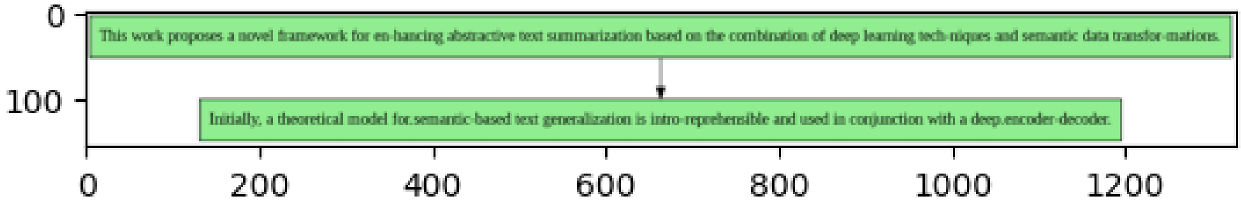
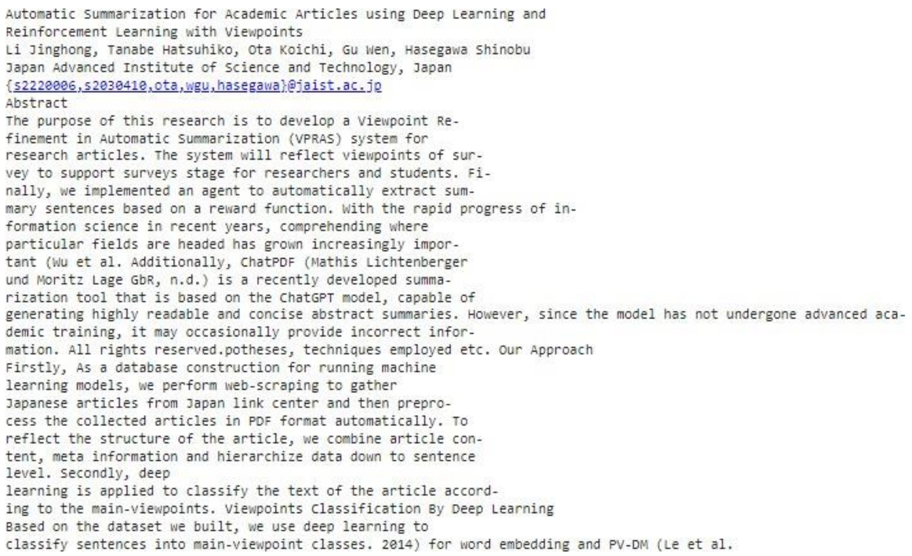



Fig. 2 Output Flow Chart 1

Input 2:



Automatic Summarization for Academic Articles using Deep Learning and
Reinforcement Learning with Viewpoints
Li Jinghong, Tanabe Hatsuhiko, Ota Koichi, Gu Wen, Hasegawa Shinobu
Japan Advanced Institute of Science and Technology, Japan
{s2220006,s2030410,ota,wgu,hasegawa}@jaist.ac.jp
Abstract
The purpose of this research is to develop a Viewpoint Re-
finement in Automatic Summarization (VPRAS) system for
research articles. The system will reflect viewpoints of sur-
vey to support surveys stage for researchers and students. Fi-
nally, we implemented an agent to automatically extract sum-
mary sentences based on a reward function. With the rapid progress of in-
formation science in recent years, comprehending where
particular fields are headed has grown increasingly impor-
tant (Wu et al. Additionally, ChatPDF (Mathis Lichtenberger
und Moritz Lage GbR, n.d.) is a recently developed summa-
rization tool that is based on the ChatGPT model, capable of
generating highly readable and concise abstract summaries. However, since the model has not undergone advanced aca-
demic training, it may occasionally provide incorrect infor-
mation. All rights reserved.potheses, techniques employed etc. Our Approach
Firstly, As a database construction for running machine
learning models, we perform web-scraping to gather
Japanese articles from Japan link center and then prepro-
cess the collected articles in PDF format automatically. To
reflect the structure of the article, we combine article con-
tent, meta information and hierarchize data down to sentence
level. Secondly, deep
learning is applied to classify the text of the article accord-
ing to the main-viewpoints. Viewpoints Classification By Deep Learning
Based on the dataset we built, we use deep learning to
classify sentences into main-viewpoint classes. 2014) for word embedding and PV-DM (Le et al.

Fig. 3 Output Summary 2

Output Flowchart 1:

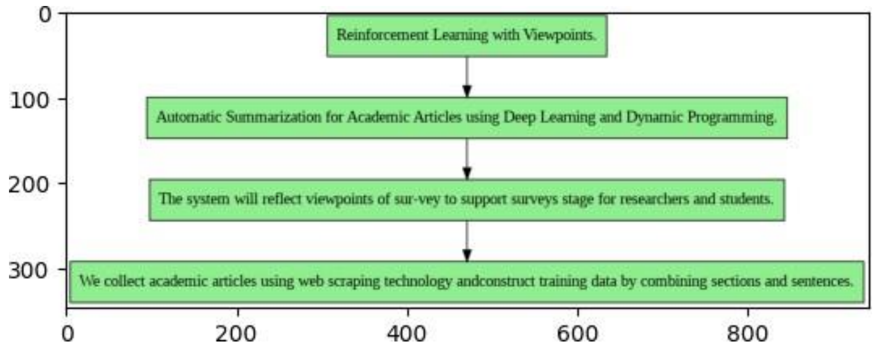



Fig. 4 Output Flow Chart 2

### *B.* Testing

ROUGE, which stands for "Recall-Oriented Understudy for Gisting Evaluation," is a crucial instrument for assessing the effectiveness of machines learning and natural language processing (NLP) automatic summarization algorithms. Its main goal is to determine the degree of similarity between summaries that are automatically produced by systems and those that are created by humans, or reference summaries. ROUGE gives recall top priority and highlights how the system can condense all relevant information from the source documents into the summary. ROUGE guarantees that important details are not missed by summarization techniques by emphasizing recall. Each of the metrics that make up ROUGE, such as ROUGE-N, ROUGE-L, and ROUGE-W, is intended to measure a different component of summary quality. ROUGE-L determines the longest common subsequence (LCS), ROUGE-N quantifies n- gram overlap, and ROUGE-W gives n-grams weighted counts. ROUGE scores are determined by comparing the generated summary with one or more reference summaries, calculating the overlap of n-grams or LCS, and then calculating metrics for precision, recall, and F1-score.

ROUGE scores: {'rouge-1': {'r': 0.2684268426842684, 'p': 1.0, 'f': 0.4232437087187465}, 'rouge-2': {'r': 0.17338534893801474, 'p': 0.9615384615384616, 'f': 0.29379360740031907}, 'rouge-l': {'r': 0.2684268426842684, 'p': 1.0, 'f': 0.4232437087187465}}

## V. RESULTS & DISCUSSIONS

TABLE I. EVALUATION OF HYBRID SUMMARY

| | **Recall** | **Precision** | **F-1 score** |
|---|---|---|---|
| Rouge-1 | 0.268 | 1.000 | 0.423 |
| Rouge-2 | 0.173 | 0.961 | 0.294 |
| Rouge-L | 0.268 | 1.000 | 0.423 |

TABLE II. EVALUATION OF CHATGPT SUMMARY

| | **Recall** | **Precision** | **F-1 score** |
|---|---|---|---|
| Rouge-1 | 0.056 | 0.756 | 0.104 |
| Rouge-2 | 0.017 | 0.365 | 0.032 |
| Rouge-L | 0.052 | 0.704 | 0.097 |

The comparison between the summarization results of our algorithm and ChatGPT highlights the algorithm's superior performance in capturing relevant information. In Table 1, the algorithm consistently achieves higher recall and precision scores across all Rouge measures compared to ChatGPT in Table 2. Particularly noteworthy is the algorithm's perfect precision (1.000) in Rouge-1 and Rouge-L, indicating its exceptional ability to accurately extract key details from the source material.

In contrast, ChatGPT presents lower recall and precision scores, suggesting it struggles to precisely capture the essence of the text. While ChatGPT's F-1 scores are relatively higher for Rouge-1 and Rouge-L, indicating a balanced performance in terms of precision and recall, its inability to match the algorithm's precision levels implies a potential loss of critical information.

Overall, the algorithm's superior precision implies a higher level of accuracy and reliability in its summaries, ensuring that important details are not overlooked. While ChatGPT may demonstrate a more balanced performance across Rouge measures, its lower precision raises concerns about the comprehensiveness and accuracy of its summaries compared to the algorithm. Therefore, the results suggest that the algorithm outperforms ChatGPT in producing more accurate and informative summaries.

## VI. CONCLUSIONS & FUTURE SCOPE

In conclusion, our study delved into the efficacy of research paper analysis, including nlp, and Deep Learning models, like T5, nltk, spacy, network, bert-extractive-summarizer, rouge for analyzing text data, particularly focusing on summarizing the research paper in easier ways by creating a flow chart to speed up the analysis.

Our findings underscored the importance of tailored techniques and extraction methods in enhancing research paper analysis accuracy. While the bert-extractive-summarizer model exhibited promising performance, Bart method (Abstractive summarization) also showcased notable accuracy. However, the Deep Learning model, being powerful, demonstrated higher accuracies compared to traditional machine learning approaches.

The functional requirements document offers a thorough blueprint for developing an advanced automated system that would revolutionise the analysis of research papers. By merging state-of-the-art deep learning and natural language processing (NLP) methods, the system aims to provide dynamic characteristics.

Looking ahead, expanding the scope to include more ways to increase the analysis accuracy by using intricate and newer frameworks that can foster inclusivity and applicability of multiple documents i.e, multiple related research papers to make the summary domain focused. Customizing the analysis and incorporating multiple domain documents can significantly enhance the accuracy effectively.

Future research endeavors could focus on optimizing preprocessing techniques and incorporating more specific features to further boost accuracy and robustness in multi modal analysis. Additionally, exploring hybrid models integrating machine learning and deep learning techniques could offer novel insights into analysis tasks, paving the way for more nuanced analysis.

## REFERENCES

[1] C. An, M. Zhong, Y. Chen, D. Wang, X. Qiu, and X. Huang, “Enhancing Scientific Papers Summarization with Citation Graph,” Proceedings of the ... AAAI Conference on Artificial Intelligence.

[2] P. Kouris, Georgios Alexandridis, and A. Stafylopatis, “Abstractive Text Summarization Based on Deep Learning and Semantic Content Generalization,” Association for Computational Linguistics.

[3] T. Chen, X. Wang, T. Yue, X. Bai, C. X. Le, and W. Wang, “Enhancing Abstractive Summarization with Extracted Knowledge Graphs and Multi-Source Transformers,” Applied sciences .

[4] M. Hu, H. Guo, S. Zhao, H. Gao, and Z. Su, “Efficient Mind-Map Generation via Sequence- to-Graph and Reinforced Graph Refinement,”

[5] M. F. Nurrokhim, L. S. Riza, and Rasim, “Generating mind map from an article using machine learning,” Journal of Physics: Conference Series .

[6] T. Vo, “An approach of syntactical text graph representation learning for extractive summarization,” International Journal of Intelligent Robotics and Applications.

[7] M. Elhoseiny and A. Elgammal, “Text to multi-level MindMaps,” Multimedia Tools and Applications.

[8] Brian prasad, “A flow-chart-based methodology for process improvement”, 2020.

[9] Phu Pham, Loan T. T. Nguyen, Witold Pedrycz & Bay Vo, “Deep Learning, Graph-based Text Representation and Classification: A Survey, Perspectives, and Challenges”, 2022.

[10] Gregory César Valderrama Vilca & Marco Antonio Sobrevilla Cabezudo, “A Study of Abstractive Summarization Using Semantic Representations and Discourse Level Information”, 2017.

[11] Marko Grobelnik, “Assigning Keywords to Documents Using Machine Learning”, 2021.

[12] Shengbin Liao, “TopicLPRank: a keyphrase extraction method based on improved TopicRank”, 2023.