Pertanika Proceedings 1 (5): 109 - 113 (2025)

PERTANIKA JOURNALS

PERTANIKA PROCEEDINGS

Journal homepage: http://www.pertanika.upm.edu.my/

Streamlining Academic Insights Gen AI in Research Paper Summarization

Priyadharshini P*, Sahanna B, Sangeetha G and Sivaranjani M

Department of Computer Science and Engineering, National Engineering College, Kovilpatti 628503, India

ABSTRACT

In today's fast-paced academic environment, researchers face the challenge of processing vast amounts of information to produce comprehensive research papers. This project introduces a web-based application designed to automatically gather insights from multiple PDFs or text files and organize them into clear, well-structured academic documents. By applying natural language processing (NLP) and machine learning (ML) techniques, the system extracts key sections, namely abstracts, introductions, methodologies, results, and conclusions from the inputted research papers, generates the content, and generates a single well-formatted document. The platform streamlines the process of literature reviews and report writing by automating tasks like text extraction, summarization, and content organization, saving both time and effort. Users can easily adjust the documents to focus on certain topics or follow academic writing standards, such as APA or MLA formats. Visualization tools are included in the system to address complex results and depict them in a simple manner. The tool automates repetitive work, enabling researchers to perform tasks efficiently and consistently. It allows for an increased focus on analysis and reasoning. Simplifying the process of writing research papers while maintaining rigorous academic standards is the key to the solution.

Keywords: Academic writing, automated research synthesis, data summarization, document generation, machine learning, natural language processing, PDF analysis

ARTICLE INFO

Article history:
Received: 29 August 2025
Published: 31 October 2025

DOI: https://doi.org/10.47836/pp.1.5.020

E-mail addresses:
Priyadharshinip.97@gmail.com (Priyadharshini P)
2112084@nec.edu.in (Sahanna B)
2112086@nec.edu.in (Sangeetha G)
2112090@nec.edu.in (Sivaranjani M)
* Corresponding author

INTRODUCTION

For Multi-Document Summarization (MDS) to work, a high-level sophisticated algorithm prevents a given document's contents from being repeated with unacknowledged paraphrasing. One is patented, which solves this problem with a two-step method, first done by clustering, preprocessing, and cleansing the articles, and later applying the

Text Rank Algorithm to the most important sentences (Sarwadnya & Sonawane, 2018). Later, with the addition of MMR to the system, the results improved. MMR proved to cut down the redundancy of selecting relevant sentences that were distinct from each other. The method was evaluated by ROUGE-1 and ROUGE-2, metrics for measuring summarization, assessing the level of coherence and informativeness (Gupta & Patel, 2020). The F-scores of 0.5103 and 0.4257 were achieved, proving the proposed method did succeed in minimizing redundancy and enhancing the quality of the summary (Gunawan & Harahap, 2019). With Indian languages, extractive summarization is an untouched area of research. To address this gap, systems targeting the Marathi language had been developed. This system operates by choosing informative and meaningful sentences to facilitate quicker content comprehension. Testing the system using the ROUGE evaluation method demonstrated the accuracy and relevance of the generated summaries (Lin & Guan, 2024).

As stated above, recent developments in automatic summarization utilize ELMo embeddings. These are deep contextual representations of words known as ELMO, which deep learning algorithms use to generate and comprehend the context of words in interpretation text (Awais & Nawab, 2024). ELMo constructs text better, increasing the coherence of the summaries produced.

In earlier uses of ELMo, it was primarily integrated in both abstractive summarization and deemed useful (Chen & Zhuge, 2018). Its application in extractive summarization is new, and that added implementation increased the system's ability to understand the text content and gave clear, practical summaries of vast quantities of written materials (Rezaei et al., 2019). These techniques are very helpful for handling large volumes of information and can be used in many areas such as education, news, and social media. This shows how powerful modern tools like ELMo can be in improving extractive summarization (Hájek & Horák, 2024).

PROBLEM STATEMENT

Writing a research paper takes a lot of time and effort. It involves collecting information from different places, summarizing the main points, and formatting the paper to follow some rules. Doing the tasks manually can lead to low-quality work, repeated information, and wasted time. It is also hard to extract content from different sources like PDFs, websites or plain text, because each one needs a different method to process.

Most tools available today do not let users customize the format and often do not summarize well. With the growing amount of research information, there is a strong need for a tool that can handle different types of inputs, pick out the important content, summarize it clearly, and create a well-organized research paper.

To solve this problem, we suggest building an automated system. This system will use smart tools to extract content, the Google Gemini API to summarize it, and formatting tools

to arrange the content nicely. It will help users create high-quality, ready-to-use research papers faster and with less effort.

RESULTS

In today's world full of information, researchers often find it hard to bring together insights from many academic papers. This web-based application helps solve that problem by automating the process of combining and organizing information. Using advanced NLP and machine learning, it can scan multiple PDFs or text files, pick out key points, and create a clear, well-structured research paper as shown in Figure 1. Researchers now have easy access to rich academic publications, which further adds to the growing internet boom. However, synthesizing the insights from numerous papers is still a daunting task. This issue is resolved with an online tool that automates the gathering and structuring of information, as shown in Figure 2. The tool reads hundreds of text files or PDFs, partitions

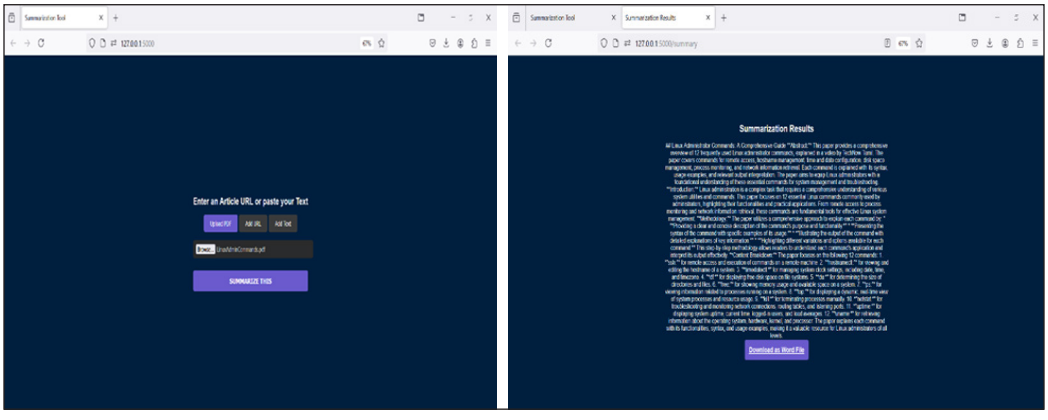

Figure 1. PDF uploaded page and summarized text

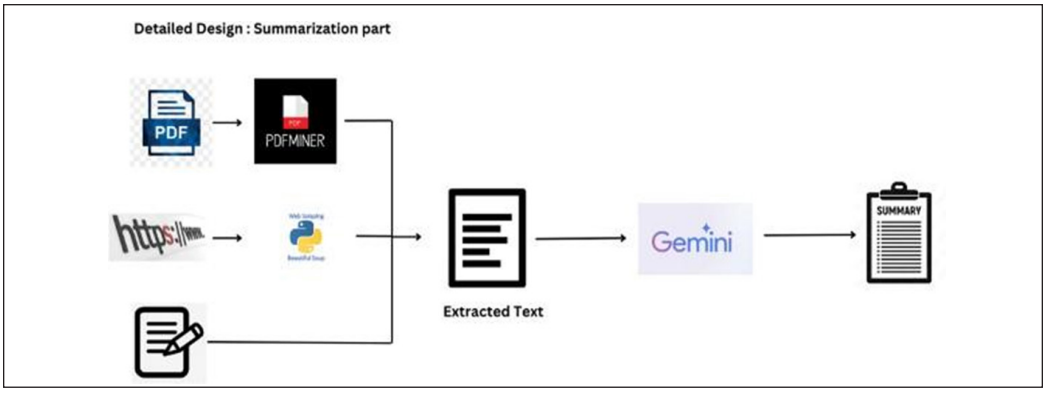

Figure 2. Step-by-step summarization

them into essential fragments, and then arranges and writes a complete research paper from them using advanced NLP and machine learning technology. The crafting of the logic and structure, reasoning and aims, as well as coherence of the primary document is aligned to academic standards. The tool enhances productivity, reduces errors, and saves valuable time, completing tedious tasks like formatting and sectioning documents, which improves researchers' efficiency. This means researchers can redirect their energies toward analysis and reflection.

CONCLUSIONS

This project helps researchers improve information overload by automating the generation of research papers from a collection of PDFs or content files. The application is crafted using Natural Language Processing (NLP) and Machine Learning (ML), which can automate paper generation by extracting marked abstraction, results, and conclusions and producing a single integrated paper formatted correctly, thereby increasing productivity and saving time for most researchers.

NLP sets the limits within which the topic is discussed, while ML completes the task through document interrelation. It improves text comprehension of a large volume of scholarly materials. Moreover, the paper argues that they can concentrate on novel analysis and concepts to a greater extent, as the tool automates tedious, time-consuming, and meticulous processes like extracting, formatting, organizing data, and many more. When optimized, the workflow becomes simpler and makes research easier, improving efficiency in academic and occupational settings.

ACKNOWLEDGEMENT

The authors sincerely thank to National Engineering College for providing the necessary facilities and resources to carry out this work. Authors also extend their appreciation to family and peers for their encouragement during this research project.

REFERENCES

Awais, M., & Nawab, R. M. A. (2024). Abstractive text summarization for the Urdu language: Data and methods. *IEEE Access*, *12*, 61198-61210. https://doi.org/10.1109/ACCESS.2024.3378300

Chen, J., & Zhuge, H. (2018). Extractive text-image summarization using multi-modal RNN. In *14th International Conference on Semantics, Knowledge and Grids (SKG)* (pp. 245-248). IEEE. https://doi.org/10.1109/SKG.2018.00033

Gunawan, D., & Harahap, S. H. (2019). Multi-document summarization by using textrank and maximal marginal relevance for text in Bahasa Indonesia. In *International Conference on ICT for Smart Society* (pp. 1-5). *IEEE*. https://doi.org/10.1109/ICISS48059.2019.8969785

Gupta, H., & Patel, M. (2020). Study of extractive text summarizer using the elmo embedding. In *Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 829-834). IEEE. https://doi.org/10.1109/I-SMAC49090.2020.9243610

Hájek, A., & Horák, A. (2024). Czegpt-2–training new model for Czech generative text processing evaluated with the summarization task. *IEEE Access*, *12*, 34570-34581. https://doi.org/10.1109/ACCESS.2024.3371689

Li, X., & Guan, S. (2024). Dynamic SLAM with Multiple Information Constraint. In *43rd Chinese Control Conference (CCC)* (pp. 7575-7581). IEEE. https://doi.org/10.23919/CCC63176.2024.10661864

Rezaei, A., Dami, S., & Daneshjoo, P. (2019). Multi-document extractive text summarization via deep learning approach. In *5th Conference on Knowledge Based Engineering and Innovation (KBEI)* (pp. 680-685). IEEE. https://doi.org/10.1109/KBEI.2019.8735084

Sarwadnya, V. V., & Sonawane, S. S. (2018). Marathi extractive text summarizer using graph based model. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICCUBEA.2018.8697741