Research Paper Summarizer Using AI

G. Santhoshi[1], M Jyothi[2], Kovvuri Ramya Sri[3], G. Hasika[4], G. Varsha[5], R. Snigdha[6]
[1,2]Assistant Professor, Department of Information Technology, G. Narayanamma Institute of Technology and Science for Women, Hyderabad, India.
[3]Assistant Professor, Department of Computer Science Technology, G. Narayanamma Institute of Technology and Science for Women, Hyderabad, India.
[4,5,6]Student Department of Information Technology, G. Narayanamma Institute of Technology and Science for Women, Hyderabad, India.
Emails: gunasanthoshi@gmail.com[1], jyothiamireddy@gnits.ac.in[2], 33ramyasri@gmail.com[3], 22251a1243@gmail.com[4], varshh321@gmail.com[5], rsnigdha037@gmail.com[6]

Abstract

An AI-powered tool that can analyze and summarize research papers, making it easier for students to understand complex academic articles. The amount of digital information is growing rapidly, making it hard to handle and understand all the text available in different areas. It's really important to quickly and accurately summarize large articles or research papers of text to find information, combine knowledge, and make decisions. This research paper explains how we developed and tested a system that can turn long documents into short, clear summaries. Develop algorithms for extracting key phrases and terms that capture the core concepts and topics of the research paper. Develop features for Highlighting Keywords, read aloud option, Plagiarism Check, Extracting Images, and focus areas. This tool plays an important role and help researchers and high academic professors to get updated with the current technologies in their respective fields. The Research Paper Summarizer Project utilizes advanced Natural Language Processing (NLP) to analyze and summarize research papers effectively.
Keywords: Natural Language Processing (NLP), Highlighting Keywords, Read Aloud, Plagiarism, Images, Research.

1. Introduction

In present digital world scenario, there are so many research papers that it can be challenging for students, researchers, and professors to keep up with the latest developments. Understanding and getting updated with such large amount of information takes lot of time, which is not always feasible. To address this issue, we developed an AI-powered tool specifically designed to analyze and summarize research papers efficiently. This tool not only generates clear, efficient summaries but also highlights key phrases and terms, reads the text aloud, checks for plagiarism, extracts relevant images, and focuses on core concepts. We use advanced Natural Language Processing (NLP) techniques, our tool simplifies the complex academic literature to simple and accessible summaries. This project aims to support academic and researchers by providing an effective solution to manage and comprehend the large amount of complex information in various fields. [10]

2. Literature Survey

[1] It creates a summary by first organizing the document in layers and then choosing sentences step by step, considering what has already been included. It treats the task of picking sentences for the summary like a decision-making problem, where the document provides the information, and selecting each sentence is like taking an action. [2] This review on text summarization was conducted using a Systematic Literature Review (SLR) approach. SLR is a method to find, evaluate, and interpret all relevant research on a specific topic or set of research questions. [3] The software uses the external tool WordNet to improve the generated summary. WordNet is a database that groups words by their meanings. The Natural Language Toolkit (NLTK)

for Python is used to connect to WordNet through the program. The quality of the summarization is evaluated using ROUGE. [4] Sentence scoring features are grouped into seven categories. One category, frequency-keyword heuristics, uses the most common words in the document to identify its main themes. Sentences that include these frequent words are scored based on how often these words appear. Another category, indicator phrases, focuses on words that usually appear in important or informative parts of the text. [5] Extractive unsupervised summarization creates a summary from a document without using any pre-labeled data or classifications. There are three main methods to do this: graph-based, latent variable, and term frequency. These methods are easy to implement and provide good results. They often produce better outcomes compared to other advanced techniques. [16]

3. Proposed System

In our proposed system we developed an AI summarization tool where the users can upload a research paper and get the summary of the paper [6]. This system is based on the research paper that is uploaded by the users. [13] This system will generate a summary by framing a meaningful sentence that are extracted from the paper to generate the extract summary for the research paper. This system will provide the images related to main content of the research paper that are extracted from the paper along with summary to visualize the images the are present in the research paper. bIt also includes the plagiarism checker it will give the how much percentage of text is included in the plagiarism. This system has a read aloud feature where the users can use it to read the summary that is generated [7-9]. It also underlines the keywords in the research paper to highlight the words in the summary. Keywords and read aloud module enhance the user interaction with the paper. This module identifies the important words in the paper and highlight them. It will help the users to locate essential information. [17] This system can create simplified and coherent summaries making complex papers more accessible and understandable to users. The system also allows the users to perform various controlling actions like to

view the images in the paper, to view the summary of the paper, and can also check for the plagiarism. [11] The flow of our system is first the user uploads the research paper and then text processing will be done on the paper that is uploaded. Then it will generate the summary of the paper and it will check for the post-processing and removes the stop words and again goes to the text processing stage. [12] This process will continue until it generates the meaningful summary of the paper, Shown in Figure 1.
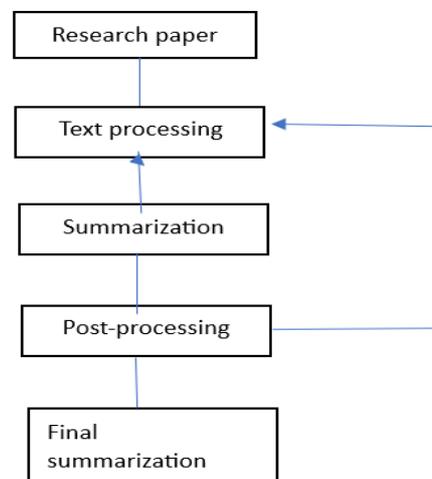

Figure 1 Architecture of The System

4. Implementation

4.1. Natural Language Toolkit Module

Our system uses the nltk module. It includes text processing libraries for tasks such as tokenization, stemming, lemmatization, part-of-speech tagging, and named entity recognition. Tokenization is used to split the text form the paper into words or the simple sentences. [14] Stemming is used to reduce the words to their root. Lemmatization will make sure that there is no grammatical mistakes in summary. Stop words is used to remove the unnecessary words from the paper like is, to, in etc. which will not effect to the meaning of the sentences. Named entity recognition will identify the proper nouns in the text to add it into the summary. [15] Also we incorporated a PIL for the image and imageTK libraries, Image library is used for opening,

manipulating, and saving the various images. ImageTk will be used to display the generated images on the TKinter GUI interface. PYPDF2 module it is used for extracting the text from the files it can access a metadata [18]. The image extracting is responsible for the extracting visual elements from the research paper. [20] This system reads the reads the text from the pdf and break down the sentences into meaningful smaller sentences. Count the frequency of each word to highlight the words in the summary. Joining all the meaning full sentences to generate a summary, shown in Figure 2.
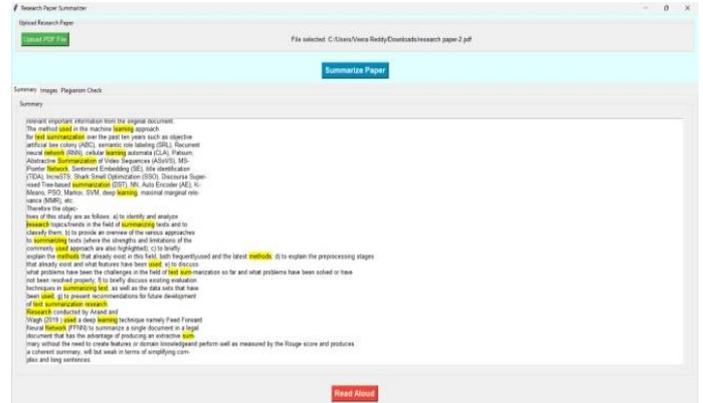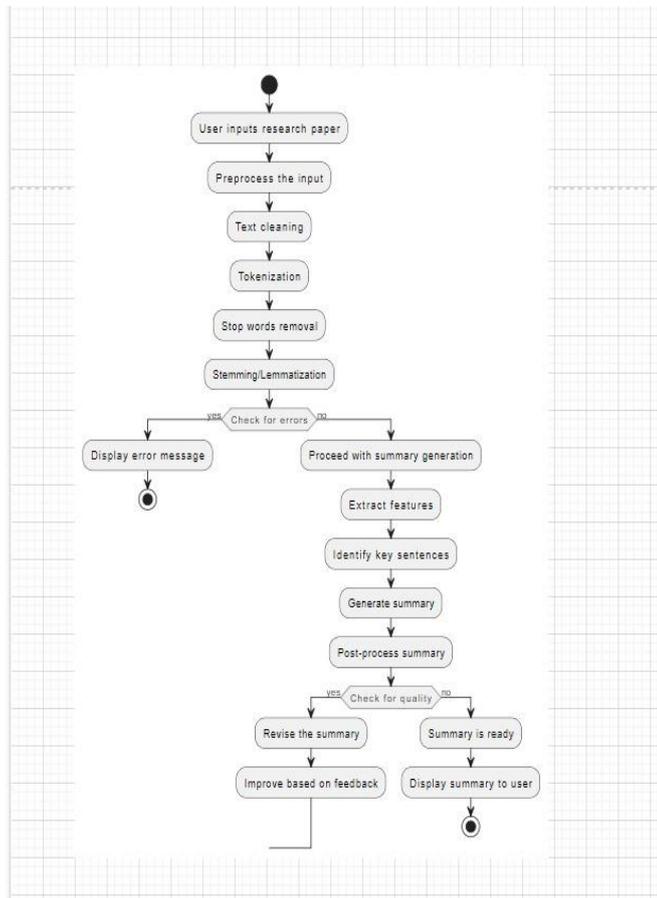

Figure 2 Flow Chart for The Summarization

5. Result

5.1. Generated Summary

Firstly, the file should be uploaded (PDF). After uploading a path will be displayed in the interface. Later, after selecting the option named 'Summarize Paper', the summary will be generated and displayed as shown in Figure 3.

Figure 3 Result of Summary Generated

5.2. Images

Based on the paper uploaded the images present in the paper could be charts, flowcharts or any graphs, such images will be extracted and displayed in the option name 'Image'. [19], shown in Figure 4 & Figure 5.
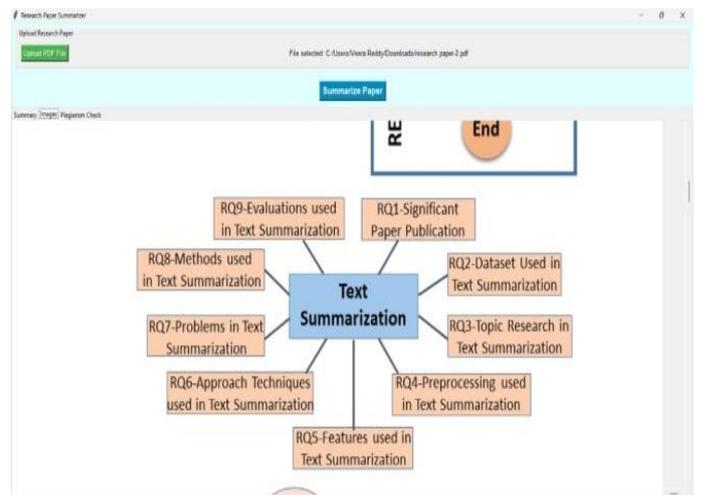

Figure 4 Image Generated from The Paper

Figure 5 Image Generated from Paper

5.3. Plagiarism

It gives the plagiarism percent detected, Figure 6 Clicking on the option 'Plagiarism'.
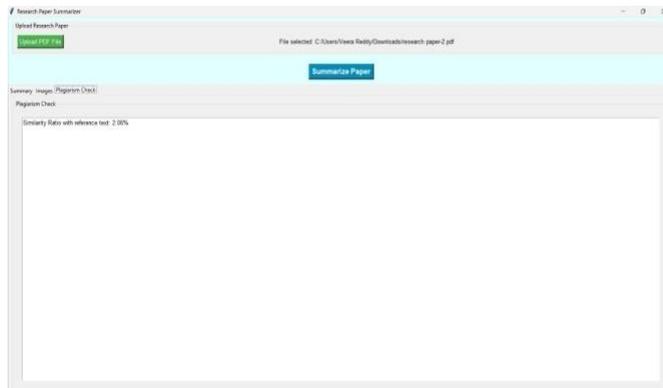

Figure 6 Plagiarism Percent

Conclusion

This paper presents an AI tool which summarizes the research papers and give effective summary. It utilizes Natural Language Processing (NLP) techniques to give meaningful and clear summaries. In conclusion, our AI tool makes it much easier for students and researchers to understand and keep up with the vast amount of research papers by quickly summarizing them and highlighting key information. This helps save time and improves learning and decision-making in various academic fields.

References

[1]. C. Virmani, A. Pillai and D. Juneja, Int. J. of Comp. Intell. Res. 13 621–630 (2017).

[2]. H. P. Luhn, IBM Journal of Research and Development 2, 159–165 (1958)

[3]. H. P. Edmundson, Journal of the Association for Computing Machinery 16, 264–285 (1969)

[4]. Julian Kupiec, Jan Pedersen and Francine Chen "A Trainable Document Summarizer 2014

[5]. Ishitva Awasthi, Kuntal Gupte, Prabjot Singh Bhogal, Sahejpreet Singh Anand, Prof.Piyush Kumar Soni "Natural Language Processig based Text Summarization" 2021

[6]. Alexander R: Fabbri, Richard Socher, Dragomir Radev "SummEval: Re-Evaluating Summarization" 2019

[7]. Dr.Annapurna P Patil, Shivan Dalmia, Syed Abu Ayab Ansari, Tanay Aul, Varun Bhatnagar "Automatic Text Summarization" 2014.

[8]. Jen-Yuan Yeh, Hao-Ren Ke, Wei-pangYang "Text Summarization using a trainable summarizer and latent semantic analysis" 2004.

[9]. Xiaoyue Liu, Jonathan J., webster & Chunyu kit "An Extractive Text Summarization Based on significant Words" 2009.

[10]. Galindo, J., & Tamayo, P. (2000). Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. Computational economics, 15, 107-143.

[11]. Ch Ramesh and K Ramya Sri, "Evaluation of Machine Learning Models for Credit Scoring[J]", Test Engineering and management, vol 82, Page number 2798-2805.ISSN0193-4120, https://zenodo.org/record/7920846

[12]. Kovvuri Ramya Sri," Applied Regression Techniques through Cases Studies Using R", Lambert Academic Publication, July 10, 2023, ISBN-10 : 6206738671, ISBN-13 : 978-6206738671

[13]. Kovvuri Ramya Sri, "Convolutional Neural Network- An Advanced approach of Neural Network", July 6, 2023, ISBN-10 : 6206686531, ISBN-13 : 978-6206686538

[14]. Guna Santhoshi and Srinivasu Badugu "Development of Different Word Vectors and Testing Using Text Classification Algorithms for Telugu", 19 April 2022, pp 33–45, Springer, Singapore, ISBN 978-981-16-9669-5 https://doi.org/10.1007/978-981-16-9669-5_3

[15]. Santhoshi, G., et.al., published a Book with "Word Embedding and its Applications for Telugu Language" VSRD academic publishing a division of visual soft india pvt.ltd., Volume 1, Year 2023. ISBN-13: 978-93-91462-86-4 Pages: vi+96

[16]. Santhoshi, G., et.al., published a paper "Simulated Game Planning" in the

International Journal of Computer and Internet Security. ISSN 0974-2247 Volume 15, Number 1 (2023), pp. 21-26

[17]. M.Jyothi, Sesha Bhargavi Velagaleti, A. Vijaya Krishna, D.Sree Lakshmi "Improving Performance of Clinical and Operational Workflows in Health Tech Domain using Artificial Intelligence" International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; Volume 11 Issue VI Jun 2023.

[18]. Kovvuri Ramya Sri, K Madhavi, "Machine Learning Algorithm for Toxic Comments Analysis" International Journal for Research in Applied Science and Engineering Technology, 2023, 2321-9653 https://www.ijrase t.com/ https://doi.org/10. 22214/ijraset.202 3.54217 UGC 146

[19]. Guna Santhoshi, M. Deepthi, "Real Time Barcode Scanning without Eye Contact on Smartphones" International Journal for Research in Applied Science and Engineering Technology 2023 2321-9653 https://www.ijrase t.com/ https://doi.org/10. 22214/ijraset.202 3.54219 UG.

[20]. M.Jyothi "THE ALL-IN-ONE IOT ROBOT COMPANION" ISBN-10: 6206739457, ISBN-13 : 978-6206739456, sep-2023.