



Detecting Plagiarism in Research Papers Using AHP and Google Dorking

Alok Mathur (20BCE2959)| Prasoon Soni (20BCE2960)| Prof. Rajkumar S| Scope

Introduction

- Researchers face a significant challenge when submitting research proposals to funding agencies due to the possibility of rejection due to plagiarism or duplicity. The process can be time-consuming and resource-intensive, taking weeks to months or even years to get approved, with researchers often unaware of the reasons for rejection. Basic plagiarism checks may not always catch paraphrasing and other malpractices, resulting in stalled submissions and lost funding opportunities. This paper proposes a system to address these issues and enhance the proposal submission process for researchers.

Motivation

- AHP and Google Dorking detect research paper plagiarism. AHP calculates similarity, and Google Dorking retrieves information. Combining both detects internal and external plagiarism, providing a detailed analysis of similarities and differences. This ensures original research, maintaining scientific community and academic research credibility.

SCOPE of the Project

- The project's scope is to develop a multi-level plagiarism detection system that uses AHP and Google Dorking. The system will identify internal and external plagiarism in research papers, addressing the growing concern of plagiarism in academic institutions and scientific communities. By preserving the credibility of the scientific community and academic research, the system benefits authors and academic institutions. The project aims to enhance the quality of research papers and the overall integrity of academic research by providing a comprehensive and effective solution to the problem of research paper plagiarism.

Methodology

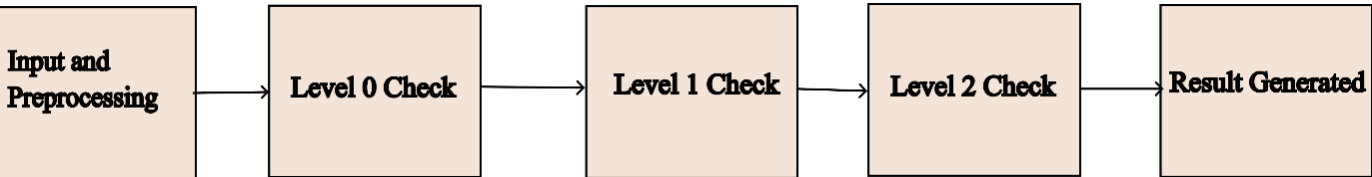


Figure 1 Flow Diagram of Proposed work

- The approach to plagiarism detection that incorporates a multi-level system for detecting plagiarism is illustrated in Figure 1.
- Three levels of plagiarism detection Level-0, Level-1, and Level-2.
- Level-0 uses Google Dorking technique to search for potential plagiarized papers on the internet.
- Level-1 checks plagiarism in the introduction, abstract, and references of the paper using advanced text comparison algorithms.
- Level-2 performs a full paper plagiarism check using advanced algorithms to scan the entire paper for plagiarism using AHP.
- AHP technique offers a comprehensive and effective solution to the problem of plagiarism detection.
- Multilevel plagiarism detection provides a detailed and nuanced analysis of potential plagiarism.

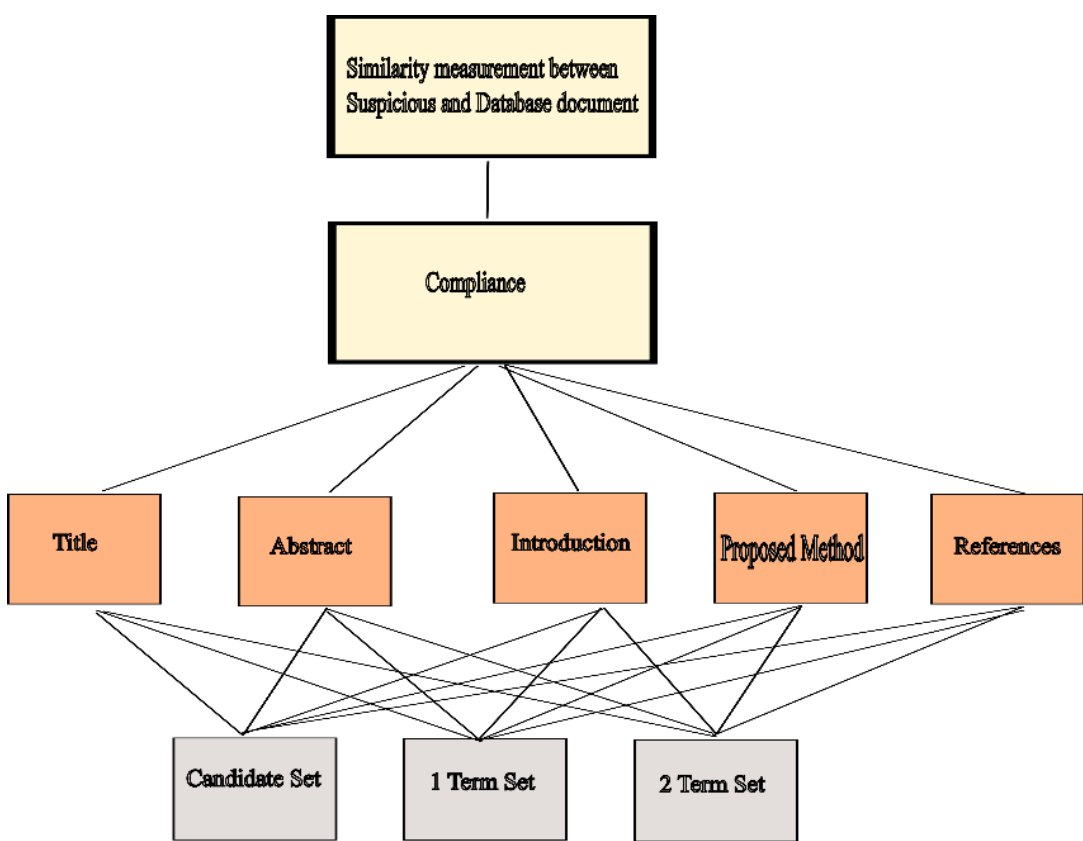


Figure 2 AHP Modelling done in Level-2

- Level 2 plagiarism detection method that involves pre-processing data to extract frequent words, bigrams, and trigrams, assigning weights to the terms, and using the AHP model to calculate similarity scores between the paper and others. The AHP method is used to determine the relevance of multiple windows, which are compared to reference documents to calculate similarity scores. Criteria, including the location and type of matched terms, are compared in a pairwise manner using a numerical scale, and weights are calculated based on these comparisons. The goal is to determine the similarity between a document and a set of reference documents.

Results

- The multi-level approach identifies plagiarism at different levels of granularity, providing a comprehensive assessment of plagiarism. Additionally, the approach significantly reduces the amount of time taken to detect plagiarism.
- We have taken 10 papers from PubMed and analysis of these paper are shown.



Figure 3 Level 0 Check on paper that gives links of article similarity

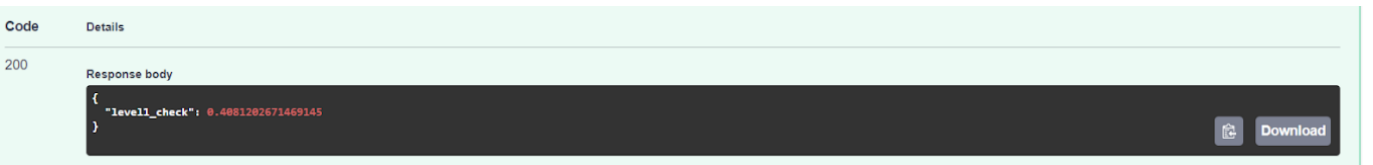


Figure 4 Level plagiarism check result on the same paper

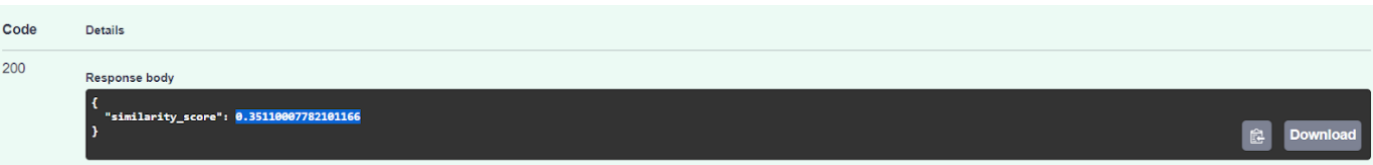


Figure 5 Level 2 Plagiarism check result

Plagiarism Level	Mean Similarity Score	Median Similarity Score	Minimum Similarity Score	Maximum Similarity Score
Level 0	0.45	0.47	0.01	0.83
Level 1	0.32	0.30	0.01	0.69
Level 2	0.20	0.19	0.01	0.38

Table 1 Analysis of 10 research papers taken from PubMed

Conclusion

In conclusion, research proposals play a critical role in advancing academic and scientific research. However, the process of submitting a proposal can be a daunting and time-consuming task for researchers. One of the significant challenges faced by researchers is the possibility of their proposals being rejected due to duplicity or plagiarism, leading to wasted time and resources. The research paper proposes a system to address the issue of duplicity and plagiarism in research proposals and improve the submission process for researchers. The proposed approach uses the Analytic Hierarchy Process technique to provide a multi-level plagiarism detection system that is both accurate and efficient. This system can ensure that submitted papers are original and have not been copied from any other source, and it can help teachers identify and reformulate plagiarized papers quickly, allowing them to focus their efforts on the remaining students. The system is designed to be accurate and efficient, and it aims to comprehensively identify instances of plagiarism, allowing educators and professionals to take appropriate action against offenders. The system comprises three levels of detection: level-0, level-1, and level-2. In each level, different techniques are used to check for plagiarism in various parts of the paper, including the introduction, abstract, references, and the entire paper. Level-0 is the Google Dorking search. Level-1 is plagiarism detection using the references, keywords and literature review. Level-2 is complete plagiarism check of paper using AHP. Overall, the proposed system can significantly streamline the research proposal submission process, saving researchers time and resources and advancing the field of research.

References

- Mansoor, M. N., & Al Tamimi, M. S. H. (n.d.). Plagiarism detection system in scientific publication using lstm networks. Iotpe.com. Retrieved February 2, 2023, from <http://www.iotpe.com/IJTPE/IJTPE-2022/IJTPE-Issue53-Vol14-No4-Dec2022/3-IJTPE-Issue53-Vol14-No4-Dec2022-pp17-24.pdf>
- Arabi, H., & Akbari, M. (2022). Improving plagiarism detection in text document using hybrid weighted similarity. Expert Systems with Applications, 207(118034), 118034. <https://doi.org/10.1016/j.eswa.2022.118034>