

## **OPTIMISING MULTI-DOCUMENT SUMMARISATION FOR EFFICIENT DIGITAL LIBRARY DOCUMENT RETRIEVAL**

**Anomija V.<sup>1\*</sup> and Akmal Jahan M. A. C.<sup>1</sup>**

<sup>1</sup>*Department of Computer Science, Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sammanthurai, Sri Lanka.*

*\*ano3071996@gmail.com*

Finding pertinent material fast is becoming more and more challenging for researchers and students as a result of the research publications' rapid development into digital libraries. The problem is exacerbated by subscription-based platforms, which frequently offer restricted access to entire publications, forcing consumers to make their decisions only on abstracts. To address these challenges, this study explores multi-document summarization techniques aimed at improving document retrieval and relevance assessment. We evaluate Maximal Marginal Relevance (MMR), Centroid, and PageRank algorithms, assessing their performance on datasets such as Multi-XScience and a manually curated set using ROUGE metrics. The findings demonstrate that MMR has the best recall-to-precision ratio, making it an excellent choice for summarizing a variety of scientific document sets. Centroid is appropriate in situations when speed is of the essence since, although being marginally less accurate, it drastically cuts down on processing time. Even if it is less useful in this situation, PageRank nonetheless offers insightful information for ranking-based strategies. To facilitate more effective navigation of large datasets and enable users to make well-informed decisions, this study highlights the significance of incorporating these summarizing techniques into digital library systems. The findings optimize the trade-off between computing speed and summary quality, which advances current attempts to improve scientific knowledge retrieval.

**Keywords:** *Centroid, Multi-Document Summarization, PageRank, ROUGE.*