

ABSTRAK

SELFIRA MADDOA, Implementasi Metode TF-IDF dan *Cosine Similarity* Pada Sistem Pencarian Artikel yang Relevan (dibimbing oleh Ir. Ida, S.Kom.,M.T, dan Darniati S.Kom.,M.T).

Perkembangan teknologi informasi menyebabkan peningkatan volume data teks digital, khususnya artikel ilmiah, yang menuntut adanya sistem pencarian informasi yang mampu menyajikan hasil secara relevan dan kontekstual. Pencarian berbasis pencocokan kata kunci secara literal dinilai belum optimal dalam menangani variasi bahasa dan konteks kueri. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi metode *Term Frequency–Inverse Document Frequency* (TF-IDF) dan *Cosine Similarity* pada sistem pencarian artikel ilmiah berbahasa Indonesia. Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen, di mana data berupa judul dan abstrak artikel diperoleh dari repositori digital terbuka. Tahapan pra-pemrosesan teks meliputi *case folding*, *tokenisasi*, *stopword removal*, dan *stemming* untuk meningkatkan kualitas representasi data. Hasil penelitian menunjukkan bahwa sistem mampu menghasilkan nilai *precision* hingga 0,9333 dan *F1-score* sebesar 1,000, yang mengindikasikan bahwa metode TF-IDF dan *Cosine Similarity* efektif dalam meningkatkan relevansi hasil pencarian. Dengan demikian, sistem yang dikembangkan mampu memberikan hasil pencarian yang lebih akurat dan kontekstual dibandingkan metode pencarian berbasis kata kunci literal, serta layak diterapkan pada repositori artikel ilmiah berskala kecil hingga menengah.

KATA KUNCI

TF-IDF, *Cosine Similarity*, Sistem Pencarian Informasi, Artikel Ilmiah, *Text Mining*

ABSTRACT

SELFIRA MADDOA, *Implementation of TF-IDF Method and Cosine Similarity in Relevant Article Search System (supervised by Ir. Ida, S.Kom.,M.T, and Darniati S.Kom.,M.T).*

The rapid growth of information technology has led to a significant increase in the volume of digital text data, particularly scientific articles, which requires effective information retrieval systems capable of providing relevant and contextual results. Conventional keyword-based search methods are often insufficient in handling linguistic variations and complex query contexts. Therefore, this study aims to implement and evaluate the Term Frequency–Inverse Document Frequency (TF-IDF) and Cosine Similarity methods in an Indonesian scientific article search system. This research adopts a quantitative approach with an experimental method, using article titles and abstracts obtained from open-access digital repositories as the dataset. Text preprocessing stages include case folding, tokenization, stopword removal, and stemming to improve data consistency and representation quality. The results indicate that the proposed system achieves a precision value of up to 0.9333 and an F1-score of 1,000, demonstrating that the combination of TF-IDF and Cosine Similarity effectively enhances the relevance of search results. Thus, the developed system provides more accurate and contextual article retrieval compared to literal keyword matching and is suitable for implementation in small to medium-scale academic repositories.

Keywords:

TF-IDF, Cosine Similarity, Information Retrieval System, Scientific Articles, Text Mining