

ANALYSIS OF THE EFFECTIVENESS OF THE COSINE SIMILARITY ALGORITHM AND THE BOYER- MOORE ALGORITHM FOR DIGITAL DOCUMENT SEARCH IN THE HEALTH OFFICE

Student Name : Wawan Ade Saputra
Student ID : 6304221420
Supervisor : Lidya Wati, M.Kom

ABSTRACT

This research analyzes the effectiveness of the Cosine Similarity algorithm and the Boyer-Moore algorithm for text document search within the SIPANDOK system. Cosine Similarity measures semantic similarity between a query and documents using TF-IDF weighting, whereas Boyer-Moore performs direct string matching. The system was developed using the Rapid Application Development (RAD) method and tested using 56 documents and 55 queries from Bengkalis Regency Health Office. Evaluation was conducted using metrics including Precision, Recall, F1-Score, Accuracy, and execution time. The research results indicate that Boyer-Moore is more effective in finding relevant documents (recall-oriented), while Cosine Similarity offers faster average execution time. Testing on the same collection of 55 queries against the same set of documents showed the fastest recorded execution times were 0.2201 seconds for Cosine Similarity and 0.0304 seconds for Boyer-Moore. Each algorithm demonstrates its own strengths within the context of text document Information Retrieval.

Keywords: Cosine Similarity, Boyer-Moore, Information Retrieval, TF-IDF, Precision, Recall, F1-Score, Accuracy.