

## ABSTRAK

Dalam kurun waktu terakhir, pertambahan jumlah dokumen karya ilmiah berbahasa Indonesia meningkat sangat pesat. Tanpa ada pengubahan dalam sistem pemerolehan informasi, volume data yang meningkat dapat mengakibatkan turunnya performa sistem pemerolehan informasi, terutama dalam hal waktu retrieval.

Salah satu metode yang diusulkan untuk mempersingkat waktu retrieval adalah pengelompokan koleksi. Dalam tugas akhir ini, G-Means dipilih sebagai algoritma pemodelan cluster. Keuntungan implementasi G-Means adalah kemampuan algoritma ini untuk memilih jumlah cluster yang paling optimal.

Hasil pengelompokan koleksi kemudian diuji dalam lingkungan sistem pemerolehan informasi untuk melihat seberapa baik pengelompokan koleksi dalam mempersingkat waktu retrieval, dan seberapa besar pengaruhnya terhadap precision.

Data yang digunakan adalah karya ilmiah berbahasa Indonesia sebanyak 100 karya. Dari hasil pengujian, ditemukan bahwa waktu retrieval lebih singkat hingga 16,3%, dengan rerata waktu retrieval sebesar 12,88 detik dan precision sebesar 47%.

**Kata Kunci:**pemerolehan informasi, clustering, g-means, dokumen karya ilmiah

## ABSTRACT

*In recent years, Indonesian-written scientific papers grow significantly in term of number. Without any improvement in information retrieval systems, increasing data volume could lead to poor system performance, especially in its retrieval time.*

*One proposed method to improve retrieval time is collection clustering. G-Means was chosen for cluster modeling algorithm, as it can determine number of generated clusters automatically. Clustering collection results are tested in information retrieval system to find how significant clustering can reduce retrieval time, and whether it has impact to system's average precision.*

*We use 100 Indonesian scientific papers as collection. Based from the results, retrieval time gain 16.3% faster, with average retrieval time is about 12,88 seconds and average precision is about 47%.*

**Keyword:** information retrieval, clustering, g-means, scientific paper