

ABSTRAK

Peningkatan jumlah dokumen berita menimbulkan permasalahan baru yang berasal dari besarnya volume data yang ada. Semakin besar volume data tersebut, maka proses pemerolehan informasi cenderung membutuhkan waktu *retrieval* yang lebih lama. Salah satu metode yang diusulkan untuk mempersingkat waktu *retrieval* adalah dengan cara mengelompokkan koleksi dokumen.

Dalam penelitian ini, *Fuzzy C-Means* dipilih sebagai algoritma pemodelan *cluster*. Representasi data menggunakan TF-IDF yang telah melalui operasi teks seperti *stopword* dan *stemming*. Hasil pengelompokkan koleksi dokumen kemudian diterapkan pada sistem pemerolehan informasi untuk mengetahui seberapa baik pengelompokkan koleksi dokumen dalam mempersingkat waktu *retrieval*, dan pengaruhnya terhadap nilai *precision*.

Data yang digunakan adalah dokumen berita berbahasa indonesia sebanyak 100 berita. Pengelompokkan menggunakan *Fuzzy C-Means* menghasilkan *purity* sebesar 98%. Hasil pengujian pada sistem pemerolehan informasi, ditemukan bahwa waktu *retrieval* lebih singkat hingga 70.8%, dan nilai *precision* sedikit lebih rendah yaitu 82.7%, dibandingkan dengan tanpa pengelompokkan yaitu 83.54%.

Kata Kunci: pemerolehan informasi, pengelompokkan dokumen, fuzzy c-means.

ABSTRACT

The increase in number of news inflicts a new kind of problem derived from the volume of data contained there. The bigger the data volume, the process to retrieve information would tend to take longer retrieval time. One of the methods suggested to shorten the retrieval time is to cluster the document collections.

In this particular research, Fuzzy C-Means was chosen as the modeling cluster algorithm. The data representation used TF-IDF which had already through text operation such as stopword and stemming. Then the results from the collection of documents were assigned to the information retrieval system to comprehend how well the clustering of document collections in lessening the retrieval time and the effect to the precision value.

The data which used were Indonesian news documents numbering 100 news. Clustering using Fuzzy C-Means gives 98% of purity. From the result test in the information retrieval system, it was found that the retrieval time lessened until 70.8% and precision value slightly lower which was 82.7%, compared to without clustering which was 83.54%.

Keyword: *information retrieval, documents clustering, fuzzy c-means.*