

Abstrak

Pesatnya pertumbuhan dan keanekaragaman informasi di dunia maya contohnya dokumen, *user* semakin sulit mencari informasi yang diinginkan. Dalam pencarian menggunakan *search engine* sering diperoleh informasi yang sangat banyak namun sebagian besar informasi tersebut terkadang tidak sesuai dengan keinginan *user*. Untuk itu diperlukan pengelompokan dokumen dengan menerapkan salah satu teknik yang terkait dengan Text Mining dan Information Retrieval yaitu *Cluster-based retrieval (CBR)*. Penerapan *CBR* dalam *search engine* diharapkan dapat mempermudah pengguna *search engine* dalam menemukan dokumen yang mereka cari.

Metode *CBR* yang digunakan adalah *Cover Coefficient Clustering Method (C3M)*, metode pengelompokan objek yang menerapkan prinsip *partitional clustering* yaitu memetakan dokumen tepat dalam satu *cluster (non-overlapping)* dengan menggunakan perhitungan *cover coefficient* dalam menentukan jumlah *cluster* dan *seedpower*. *C3M* menggunakan model probabilistik dan kesamaan term (*term similarities*) serta menggunakan *seed* dokumen sebagai inisialisasi/pola awal *cluster*. Probabilitas disini menggambarkan derajat asosiasi (keterhubungan) antar dokumen-dokumen atau *term-term*.

Pengujian dilakukan untuk menganalisis hasil pengelompokan *C3M* berdasarkan parameter kualitas *cluster* yang dihasilkan dengan menggunakan nilai *silhouette coefficient* serta menganalisis hal-hal yang mempengaruhi kualitas *cluster*. Kualitas *cluster* yang dihasilkan dipengaruhi oleh relevansi dokumen hitlist yang ter-*retrieve*, jumlah *cluster* yang terbentuk, jumlah *term query* serta jumlah dokumen *hitlist*. Penerapan *CBR* menggunakan *C3M Algorithm* menghasilkan kualitas *cluster* yang cukup baik, ditunjukkan dengan nilai *silhouette coefficient* yang bernilai positif.

Kata kunci: *cluster based retrieval, cover coefficient clustering algorithm, partitional clustering, silhouette coefficient.*