

Abstrak

Dengan bertambah pesatnya informasi/dokumen yang beredar di internet sehingga memungkinkan untuk suatu dokumen dapat dikelompokkan ke dalam dua atau lebih kategori sekaligus. Oleh karena itu dibutuhkan suatu metode untuk mengelompokkan dokumen-dokumen tersebut ke dalam dua atau lebih kategori sekaligus.

Overlapping Cover Coefficient Clustering Method (OC3M) adalah suatu metode pengelompokan dokumen dengan model probabilitik, kesamaan *term*, dan *seed* dokumen sebagai inisialisasi awal dari pembentukan *cluster*. Pada metode ini diterapkan sifat *overlap*, yaitu kondisi dimana dokumen dapat menempati lebih dari satu *cluster*.

Pengujian yang dilakukan pada tugas akhir ini dalam mengelompokkan dokumen dengan algoritma OC3M yaitu menganalisis *cluster* yang dihasilkan berdasarkan nilai *Silhouette Coefficient*-nya serta menganalisis hal-hal yang mempengaruhi kualitas *cluster* yang terbentuk. Kualitas *cluster* yang terbentuk dipengaruhi oleh banyaknya dokumen yang digunakan, tipe dokumen, kemiripan dokumen dengan pusat *cluster*, dan juga dipengaruhi oleh *overlapping coefficient* yaitu parameter yang menentukan banyaknya suatu dokumen yang *similar* dapat dikelompokkan ke dalam *cluster* yang berbeda. Dari hasil percobaan, kualitas *cluster* yang terbentuk dengan menggunakan algoritma OC3M memiliki kualitas yang cukup baik, ini di tunjukkan dengan nilai *silhouette coefficient* yang bernilai positif.

Kata kunci : *OC3M, Overlap, Overlapping, Clustering, Cluster*