

Abstrak

Seiring dengan pesatnya pertambahan jumlah dan keanekaragaman dokumen yang dapat berdampak besar pada saat pencarian suatu dokumen. Pencarian dokumen yang ada pada saat sekarang ini adalah dengan menampilkan hasil pencarian terurut berdasarkan peringkat kecocokan (document ranking). Hasil yang ditampilkan terkadang tidak sesuai (relevan) dengan yang diinginkan oleh pengguna.

Salah satu cara untuk mempermudah pengguna dalam mencari dokumen adalah dengan menggunakan pengkategorian dokumen. Dengan adanya nama kategori dari setiap pengelompokkan dapat memudahkan pengguna karena nama kategori dapat mencerminkan isi dari suatu kumpulan dokumen. Salahsatu cara untuk mengelompokkan dokumen adalah dengan *clustering*. Pada Tugas Akhir ini akan dilakukan pengelompokkan dokumen berbahasa Indonesia dengan algoritma *K Means*. Sebelum melakukan pengelompokkan diperlukan proses *pre processing* yaitu *case folding*, *parsing*, pembuangan *stopword* dan *stemming*. Proses ini diperlukan untuk mengurangi jumlah kata yang diproses. Nama kategori diberikan pada masing-masing kluster. Penamaan kategori dihitung dari frekuensi kemunculan kata terbanyak dari setiap kluster.

Analisis yang dilakukan pada tugas akhir ini adalah membandingkan dua metode pembobotan dan korelasi jarak antar objek. Selain itu, analisis juga dilakukan untuk mengetahui kesesuaian nama kategori dengan dokumen dalam satu kluster dan menganalisis hasil kluster yang dihasilkan dengan melihat nilai *convergence*, *precision* dan *recall*.

Kata Kunci: nama kategori, *k means*, *clustering*, dokumen, *pre processing*.