ABSTRAK

Suatu kumpulan dokumen yang banyak akan sulit untuk diketahui karakteristik informasinya. Diperlukannya suatu sistem yang dapat membagi dokumen-dokumen tersebut menjadi kumpulan-kumpulan yang lebih kecil berdasarkan karakteristiknya. *Clustering* merupakan pengelompokan suatu objek berdasarkan informasi yang berada pada objek tersebut. Caranya yaitu dengan mengelompokan karakteristik informasi yang sama ke suatu kelas yang sama dan suatu objek dengan karakteristik informasi yang berbeda ke kelas yang lain. Oleh karena itu pada penelitian kali ini akan dilakukan klasterisasi dengan menggunakan *Fuzzy C-Means* (*FCM*). *FCM* merupakan suatu teknik peng-*cluster*-an yang mana keberadaan tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan. Dalam pengelompokan dokumen dengan menggunakan metode ini dibutuhkan proses *lowercasing*, penghilangan tanda baca dan angka, tokenisasi, *stopwords*, *stemming* dan pembobotan *term*. Pada *stemming*, algoritma yang digunakan adalah algoritma *stemming* Arifin-Setiono. Untuk pembobotan *term* metode yang digunakan adalah pembobotan tf-idf. *FCM* menggunakan model pengelompokan *fuzzy* sehingga data dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1.

Kata kunci : *Clustering*, dokumen, algoritma *stemming* Arifin-Setiono, pembobotan tf-idf, *Fuzzy C-Means*.