

Abstrak

Berkembangnya teknologi di dunia maya membuat jumlah informasi berupa artikel berita semakin banyak. Untuk itu, diperlukan suatu pengelompokan terhadap artikel yang memudahkan pembaca mencari informasi dengan menerapkan salah satu fungsionalitas dari data mining, yaitu klasterisasi. Teknik klasterisasi yang ada saat ini masih belum secara tepat menangani data berdimensi tinggi dan database yang berukuran besar sehingga deskripsi dari klaster tersebut masih sulit untuk dipahami. Oleh karena itu dibutuhkan metode pengklasteran dimana hasil pengklasteran tersebut memiliki bentuk deskripsi klaster yang mudah dipahami.

Metode yang dapat diterapkan ini mampu mengurangi dimensionalitas data yang tinggi dan besarnya ukuran database. Ada beberapa metode yang dapat digunakan yaitu berdasarkan frequent term-based text clustering yang terdiri dari hierarchical frequent term-based clustering (HFTC) dan *frequent term-based clustering* (FTC).

Hasil dari klasterisasi berdasarkan *frequent term based text clustering* adalah berupa klaster yang memiliki deskripsi klaster yang mudah dipahami. Berdasarkan hasil percobaan dapat disimpulkan bahwa pada HFTC, F-measure nilainya semakin besar dengan *minimum support* yang semakin kecil. Hal ini menunjukkan kualitas klaster yang terbentuk pun semakin bagus. Nilai *Entropy* yang dihasilkan pada FTC bervariasi dan tidak memiliki pola pada tiap *minimum support* yang diinputkan. Serta Waktu yang dibutuhkan dalam pembentukan klaster akan semakin sedikit seiring dengan makin besarnya nilai *minimum support*.

Kata kunci: *klasterisasi, frequent term-based text clustering, HFTC, FTC, f-measure, Entropy*