

Abstrak

Pesatnya penggunaan Internet sebagai sarana penyebarluasan informasi mengakibatkan tersedianya informasi dalam jumlah yang banyak, beraneka ragam dan pada umumnya tidak terstruktur. Hal ini mendorong kebutuhan untuk mengelola informasi guna mendapatkan pengetahuan yang bermanfaat. Salah satu komponen penting dalam pengelolaan informasi adalah pengelompokan teks, dengan kategorisasi dan klastering, yang dapat membantu pengguna dalam mendapatkan informasi berdasarkan kategori tertentu secara cepat dan akurat.

Pada tugas akhir ini digunakan metode *concept based* (*statistical concept based mining model* dan *modified statistical concept based mining model*) dan metode pembobotan tradisional seperti TF IDF, untuk menghitung bobot suatu term. Output dari penghitungan ini akan menjadi input untuk kategorisasi dan klastering dokumen berita berbahasa Indonesia. Dari hasil pengelompokan dokumen, diukur performansi kategorisasi dan klastering dari ketiga metode tersebut.

Metode *concept based* memiliki performansi yang tinggi dan bervariasi pada berbagai kategori. Dari hasil penelitian pada tugas akhir ini juga diketahui bahwa metode *concept based mining model* memiliki akurasi yang lebih baik dari pada metode TF IDF. Hal ini menunjukkan *concept based mining model* dapat menjadi sebuah metode pembobotan term dalam suatu dokumen supaya performansi klastering dan kategorisasi bagus.

Kata kunci : *statistical concept based mining model*, *modified statistical concept based mining model*, TF IDF, kategorisasi, klastering.