

Abstrak

Berita elektronik merupakan media informasi yang paling populer dan interaktif saat ini. Begitu interaktifnya, hingga perkembangannya cukup pesat. Terbukti bertambah banyaknya situs perusahaan maupun situs personal, yang berarti semakin meningkatkan jumlah informasi dan data. Peningkatan yang pesat ini juga dipacu oleh penggunaan internet yang semakin berkembang dibandingkan era sebelumnya. Sebagai akibatnya, jumlah informasi meningkat secara eksponensial.

Banyaknya data yang ada, semestinya dapat memberikan manfaat yang banyak pula. *Clustering* merupakan salah satu metode untuk pengelompokan dokumen dengan menemukan keterkaitan antar dokumen. Saat ini, kebanyakan metode klusterisasi hanya mengandalkan perhitungan kesamaan berdasarkan kata dan tidak memperhatikan aspek lain, misalnya kesamaan frasa, misalnya *Vector Space Model*. Pada tugas akhir ini berusaha mengklusterkan dokumen dengan metode *Document Index Graph* yang menggunakan kombinasi dua kesamaan dokumen yaitu; kesamaan berbasis kata dan kesamaan berbasis frasa.

Metode ini diuji coba dengan menggunakan sampel berita berbahasa Indonesia dari media massa berbasis web. Pemilihan *similarity blend factor* dan *similarity threshold* yang tepat akan meningkatkan kualitas kluster. Hasil klusterisasi dievaluasi berdasarkan nilai *precision* dan *recall*.

Kata kunci: *clustering, Document, Index, Graph, similarity.*