

Abstrak

Pengelompokan dokumen merupakan suatu permasalahan yang sangat penting saat ini. Hal ini disebabkan semakin banyak orang menyimpan dokumen berbentuk digital, tetapi belum dikelompokkan dengan baik. Ini menyebabkan timbulnya kesulitan dalam menemukan dokumen pada saat kita membutuhkan informasi didalam dokumen tersebut. Banyak jenis algoritma yang dapat digunakan untuk pengelompokan dokumen. Dalam tugas akhir ini, pengelompokan dokumen dilakukan dengan menggunakan *Active Fuzzy Constrained Clustering*.

Pada tugas akhir ini , dilakukan generalisasi pada proses *build vector* sebelum dokumen dikelompokkan dengan *Active Fuzzy Constrained Clustering*. Metode *build vector* yang akan digunakan adalah *Vector Space Model (VSM)* dan *Generalized Vector Space Model (GVSM)* dengan tambahan jenis informasi semantik. Dengan penerapan kedua metode ini, dapat dilihat perbedaan serta faktor yang dapat mempengaruhi dalam meningkatkan akurasi algoritma *Active Fuzzy Constrained Clustering* dalam pengelompokan dokumen.

Perhitungan *similarity* dengan menggunakan kamus kata *wordnet (GVSM)* sangat bergantung pada *wordnet*, jika kata kata pada dokumen pengujian dapat terdeteksi dengan baik pada *wordnet* maka akurasi hasil *clustering* juga akan lebih baik.

Kata Kunci: *Active Fuzzy Constrained Clustering, Vector Space Model (VSM), Generalized Vector Space Model (GVSM), wordnet.*