

Abstrak

Seiring dengan perkembangan Teknologi, maka penambahan keanekaragaman dokumen juga akan berkembang dengan cepat. Hal ini memberi dampak pada performansi *search engine* sekarang. Banyaknya dokumen yang ter-retrieve yang ditampilkan ke *user* banyak mengandung dokumen-dokumen yang tidak *relevant*. Hal ini membuat *user* kesulitan untuk mencari dokumen-dokumen yang diperlukan.

Clustering adalah salah satu solusi untuk memecahkan masalah tersebut. *Clustering* sendiri adalah proses pengelompokan dokumen-dokumen yang memiliki tingkat kemiripan yang tinggi menjadi satu *cluster*. Dalam Tugas Akhir ini dilakukan pengelompokan dokumen hasil pencarian dengan menggunakan Algoritma *Buckshot*. *Buckshot* merupakan salah satu metode *clustering* yang menggabungkan antara Algoritma *HAC* (*Hierarchical Agglomerative Clustering*) dengan *K-Means*. *Buckshot* memakai algoritma *HAC* dalam pemilihan *initial centroid*nya sehingga kualitas *cluster* yang dihasilkan menjadi lebih baik dibandingkan dengan *K-Means*.

Analisa yang dilakukan pada tugas akhir ini adalah membandingkan kualitas *clustering* menggunakan Algoritma *Buckshot* dengan kualitas *clustering* menggunakan Algoritma *K-Means*. Adapun kualitas *clustering* bisa dilihat dari nilai *Cohesion*, *Separation* dan *Cohesion/Separation*. Menurut pengujian, kualitas *cluster* yang dihasilkan Algoritma *Buckshot* lebih baik daripada *K-Means* asalkan jumlah *cluster* yang dipilih user tidak sama dengan jumlah *sample* dokumen acak yang digunakan oleh *system*. Nilai *Precision* dan *Recall* juga berpengaruh terhadap kualitas *cluster* yang dihasilkan.

Kata Kunci : *Clustering, Search engine, Algoritma Buckshot, Algoritma HAC, Algoritma K-Means, Cohesion, Separation, Precision, Recall*