

Abstrak

Pencarian dokumen yang ada pada saat sekarang ini adalah dengan menampilkan hasil pencarian terurut berdasarkan peringkat kecocokan (document ranking). Hasil yang ditampilkan terkadang tidak sesuai (relevan) dengan yang diinginkan oleh pengguna.

Salah satu cara untuk mengelompokkan dokumen adalah dengan *clustering*. Pada Tugas Akhir ini akan dilakukan pengelompokkan dokumen berbahasa Indonesia yang jumlahnya statis dengan algoritma *Descriptive K-Means*. Perbedaannya dengan K-Means biasa adalah *Descriptive K-Means* menggunakan tahapan khusus untuk pencarian label dan pengalokasian dokumen.

Setelah implementasi, algoritma ini bisa membentuk cluster yang dokumen-dokumen di dalamnya sesuai dengan labelnya, hal ini dikarenakan query yang digunakan untuk pengalokasian dokumen relasi AND, artinya setiap kata yang terdapat pada label harus ada pada setiap dokumen.

Namun algoritma ini kurang efektif ketika kata yang dicari sedikit terdapat dalam dokumen koleksi, sehingga cluster yang terbentuk sering terkena *pruned* karena cluster tersebut jumlah dokumen yang terdapat di dalamnya tidak memenuhi jumlah minimal untuk terbentuknya suatu cluster.

Dalam pengujian ditemukan pencarian paling efektif ketika nilai $K=2$ berdasarkan nilai *precision* dan *recall*-nya. Tingkat keefektifan tersebut semakin berkurang seiring bertambahnya nilai K . Ukuran *cluster* yang dihasilkan tidak bisa diukur dari nilai *convergence* yang didapat karena sangat bergantung pada inputan kata dari pengguna dan jumlah dokumen dalam cluster.

Kata kunci: *cluster, clustering, Descriptive K-Means, preprocessing, precision, recall.*