

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 dilakukan pengelompokkan dokumen berbahasa Indonesia dari dokumen koleksi dengan algoritma LINGO. LINGO merupakan algoritma *clustering* yang lebih mengedepankan kualitas penamaan label pada klaster.

Setelah implementasi, algoritma ini bisa membentuk klaster dengan dokumen-dokumen di dalamnya sesuai dengan labelnya, hal ini dikarenakan setiap dokumen dialokasikan ke masing-masing klaster berdasarkan tingkat kemiripannya dengan label yang terbentuk.

Dalam menentukan label untuk penamaan klaster, algoritma ini memeriksa kemunculan *term* atau *complete phrase* dalam dokumen. Maka dari itu, algoritma ini sangat efektif jika dokumen-dokumen yang diproses banyak mengulang topik inti, sebaliknya akan kurang efektif jika topik inti dari dokumen diinterpretasikan dengan berbagai istilah yang beragam.

Dalam pengujian label klaster dengan metode *precision* dan *recall* pada metode pembobotan *Term Frequency* (TF) dan *Term Frequency – Inverse Document Frequency* (TF-IDF) didapatkan hasil yang bagus untuk keduanya.

Kata kunci: klaster, *clustering*, *LINGO*, *complete phrase*, *precision*, *recall*.