

Abstrak

Dokumen dengan kategori yang sama dalam jumlah yang besar sulit untuk dibedakan kesamaannya antara dokumen yang satu dengan dokumen yang lainnya. Salah satu cara yang dapat digunakan untuk mengatasi masalah ini adalah dengan *document clustering*. Untuk jumlah *cluster*-nya, *user* bahkan tidak mengetahui berapa jumlah yang tepat untuk melakukan *clustering* dokumen-dokumen tersebut. Untuk itu diperlukan metode *clustering* yang dapat menghasilkan jumlah *cluster* secara otomatis. Satu dari banyak metode yang dapat digunakan untuk menghasilkan jumlah *cluster* secara otomatis adalah *Hill Climbing*.

Hill Climbing akan melakukan identifikasi terhadap pergerakan varian dari tiap tahap pembentukan *cluster* dan menganalisis polanya agar dapat menemukan nilai *global optimum* sehingga jumlah *cluster* akan terbentuk secara otomatis. Sedangkan untuk metode *clustering* yang digunakan adalah salah satu metode dalam *Hierarchical Agglomerative*, yaitu *Centroid Linkage Hierarchical Method* (CLHM). Setiap dokumen akan dianggap sebagai sebuah *cluster*, kemudian digabungkan dengan algoritma CLHM yang berulang hingga jumlah *cluster* sesuai dengan yang diinginkan. Hasil dari *cluster* yang terbentuk akan dihitung kualitasnya dengan parameter evaluasi dan nilai *purity*.

Kata kunci : *Clustering*, CLHM, *Hill Climbing*, *varians*, *purity*