

Abstrak

Berita saat ini tidak hanya disebarakan melalui media elektronik dan media cetak, namun juga disebarakan melalui dunia internet. Sehingga jumlah berita yang tersedia sangatlah besar dan butuh waktu lama untuk mengelompokkannya secara manual. *Clustering* dapat digunakan sebagai solusi dari permasalahan tersebut. Salah satu metode yang dapat digunakan untuk *clustering* dokumen berita ini adalah metode *Hierarchical Agglomerative Clustering* (HAC). Pada tugas akhir ini metode HAC yang digunakan ialah *Single Link*, *Complete Link*, dan *Average Link*.

Metode HAC mempunyai kemampuan menggabungkan data dengan membuat hirarki, dimana data yang memiliki kemiripan akan ditempatkan di hirarki yang berdekatan dan yang tidak memiliki kemiripan ditempatkan pada hirarki yang berjauhan. Setiap dokumen akan dianggap sebagai sebuah *cluster*, kemudian digabungkan sesuai dengan metode HAC yang digunakan, berulang hingga jumlah *cluster* sesuai dengan yang diinginkan.

Hirarki yang terbentuk akan dihitung kualitasnya menggunakan *cophenetic correlation coefficient*, sementara kualitas *cluster* yang terbentuk akan dihitung menggunakan *silhouette coefficient* dan *purity*.

Kata kunci : *Clustering*, HAC, *cophenetic correlation coefficient*, *silhouette coefficient*, *purity*