

Abstrak

Saat ini kebutuhan akan segala jenis informasi telah menjadi suatu hal yang sangat penting. Informasi dapat berasal dari berbagai sumber elektronik maupun non-elektronik, salah satunya adalah berita. Internet merupakan sarana yang paling ampuh dalam penyebaran informasi. Berita pun kini dapat dengan mudah didapatkan melalui internet, melalui situs-situs media yang sangat beragam. Sehingga, jumlah berita yang tersedia sangat besar dan menyulitkan bagi pembaca berita untuk mendapatkan secara tepat berita yang diinginkannya.

Tugas Akhir ini berusaha untuk mengelompokkan berita berbahasa Indonesia. *Clustering dokumen text* dapat menjadi suatu alternatif dalam mengelompokkan berita. *Document clustering* merupakan suatu proses mengelompokkan dokumen-dokumen, sehingga dokumen yang memiliki banyak kemiripan (*similar*) masuk dalam satu kluster, dan dokumen yang berbeda (*dissimilar*) ke dalam kluster yang berbeda pula. Sehingga berita dengan topik yang masih berkaitan ditempatkan dalam satu kelompok (*cluster*). Algoritma clustering yang digunakan adalah algoritma *K-means clustering*.

Analisis terhadap penggunaan *stemming* diujicobakan juga dalam tahap *preprocessing* untuk melihat pengaruhnya terhadap kualitas *cluster* yang dihasilkan dengan menggunakan metode *Purity*. Dari hasil ujicoba hasil *purity* yang terbaik didapatkan dari proses *clustering* yang melalui tahap *stemming*.

Kata kunci: *clustering*, berita berbahasa Indonesia, *K-means*, *stemming*, *purity*.