Daftar Istilah

Clustering	: mengelompokkan data dengan karakteristik yang sama ke suatu ‘wilayah’ yang sama dan data dengan karakteristik yang berbeda ke ‘wilayah’ yang lain.
hierarchical clustering	: mengelompokkan data dengan membuat suatu hirarki berupa dendogram dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan.
Agglomerative Clustering	: proses clustering dari *N* klaster menjadi satu kesatuan klaster, dimana *N* adalah jumlah data, sedangkan divisive melakukan proses clustering yang sebaliknya yaitu dari satu klaster menjadi *N* klaster.
Multiple document summarization	: proses eringkasan untuk beberapa dokumen
Text mining	: suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponenkomponen dalam *data mining* yang salah satunya adalah kategorisasi.
Tokenization	: memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi *token* atau potongan kata tunggal atau *termmed word*.
Stemming	: proses konversi term ke bentuk umumnya :
Sentence scoring	: Pemberian nilai pada kalimat berdasarkan fitur-fitur yang dieroleh