

Abstraksi

Clustering merupakan pengelompokkan data tanpa proses label kedalam kelompok – kelompok atau cluster-cluster, yang mempunyai nilai parameter kualitas cluster yang dibentuk (purity cluster), dan sebelum melakukan pembentukan cluster-cluster didalam clustering dilakukan proses pengelolaan data menjadi kumpulan teks yang disebut dengan text mining. Pada text mining ada dua jenis kata yaitu kelompok kata yang berimbuhan dan kelompok kata yang sudah hilang imbuhan. Untuk melakukan proses pembentukan kata dasar dari kata yang berimbuhan dibutuhkan suatu proses yang disebut stemming. Algoritma affix removal stemming merupakan salah satu algoritma stemming yang bisa diterapkan dalam data teks bahasa Indonesia karena mampu menstemming data teks dengan nilai akurasi mencapai 90%. Pada studi kasus clustering teks terjemahan ayat-ayat Al-Qur'an, dibuktikan bahwa pengaruh penerapan stemming pada text mining, dapat mengurangi jumlah data teks yang akan diproses pada clustering dan mampu memberikan nilai purity cluster yang lebih baik dari pada clustering yang tidak menerapkan stemming.

Kata Kunci: *Clustering, Cluster, Text Mining, Affix Removal Stemming, Purity Cluster, Al-Qur'an.*