

ABSTRAK

Penelitian ini dilakukan di bidang analisis sentimen untuk bahasa Indonesia. Membangun sistem yang otomatis memprediksi semantik orientasi dari leksikon menggunakan sumber yang tersedia dalam bahasa Indonesia. Pada penelitian ini digunakan korpus berbahasa Indonesia yang mempunyai 9 juta kata, diperoleh dari artikel-artikel media online yaitu Kompas dan Tempo yang kemudian dianotasikan dengan POS tagger. Selanjutnya dataset dibangun dengan mengambil semua kata sifat didalam korpus, membuang kata sifat yang tidak mempunyai orientasi atau berorientasi netral sehingga diperoleh 923 kata sifat yang unik.

Sistem ini termasuk beberapa tahapan yaitu proses teks preprocessing dan clustering. Teks preprosesing bertujuan untuk meningkatkan akurasi yang akan diperoleh. Dan metode clustering akan mengklusterkan setiap kata sifat ke orientasi yang sesuai dengannya yaitu positif atau negatif. Dengan penambahan teknik pada teks preprocessing dapat diraih akurasi 72%.

Kata kunci: analisis sentimen, bahasa Indonesia, orientasi semantic otomatis, kata sifat