

Abstrak

Comparative Text Mining (CTM) merupakan bagian dari *text mining* yang secara khusus mempunyai fungsi: menemukan tema umum dari semua koleksi, dan menemukan tema khusus dari suatu dokumen. Penggunaan CTM salah satunya yaitu untuk *summarizing reviews*. *Summarization* merupakan proses otomatis yang menghasilkan versi dokumen yang lebih pendek (50% atau kurang) namun tetap berguna bagi pengguna. Dengan *summarization*, pengguna diharapkan dapat menangkap isi dokumen tanpa harus melihat keseluruhan isi dokumen.

Metode *clustering* adalah metode yang memiliki kemampuan untuk menganalisis serta mengelompokkan secara otomatis dokumen-dokumen. Teknik *clustering* pada umumnya menggunakan kata dan dokumen biasanya dianggap sebagai kumpulan kata-kata tanpa adanya urutan atau yang disebut *bag of word*. *Suffix Tree Clustering* (STC) adalah algoritma pertama yang menggunakan frasa (*multi-word terms*) sehingga prosesnya lebih sederhana dibandingkan dengan algoritma yang lain. STC adalah algoritma *incremental*, kompleksitas waktu perhitungannya linear dan memenuhi kriteria untuk *document clustering web*.

Penerapan algoritma STC untuk CTM dilakukan untuk membuktikan dan melakukan kajian terhadap performansi dari algoritma STC tersebut. Dalam penelitian dilakukan pengamatan pengaruh parameter terhadap optimalisasi hasil yang dicapai, yaitu pengaruh parameter nilai *background theme* dan *bestRank range*.

Kata kunci: Comparative Text Mining (CTM), metode clustering, Suffix Tree Clustering (STC), dan tema.