

## Abstrak

Peringkasan teks (*text summarization*) adalah proses penyaringan informasi yang paling penting dari sebuah atau beberapa dokumen sehingga dihasilkan versi yang lebih singkat yang digunakan oleh *user (users)* untuk *task* tertentu (*tasks*) dengan memanfaatkan aplikasi yang dijalankan pada komputer.

Pada Tugas Akhir ini diimplementasikan *centroid-based summarization*. *Centroid* dari klaster merupakan kumpulan dari *term* melalui proses perhitungan yang dianggap penting terhadap sebuah klaster dari dokumen-dokumen. Pada Tugas Akhir ini, penulis mencoba melakukan modifikasi terhadap *centroid* yang digunakan yaitu berupa *concept*. *Concept* adalah *term* yang mempunyai peranan semantik di sebuah kalimat dalam sebuah dokumen. Nilai *concept* akan dihitung dengan menggunakan *statistical concept based mining model*. Aplikasi ini dilakukan terhadap dokumen-dokumen berbahasa Indonesia dan Inggris dengan isi berita yang mirip. *Feature-feature* ekstraksi kalimat yang digunakan adalah *centroid value* yaitu penjumlahan nilai-nilai *centroid* dari kata-kata yang terdapat dari kalimat tersebut, letak/posisi kalimat dalam paragraf, *similarity with title* yaitu nilai kemiripan suatu kalimat yang dibandingkan dengan judul dari dokumen dengan menghitung kata-kata yang *overlap*, dan *redundancy penalty*. Parameter-parameter di atas akan dihitung untuk menentukan ranking dari kalimat. Hasil dari aplikasi ini adalah ringkasan ekstraktif yang terdiri dari kalimat-kalimat dengan ranking tertinggi.

Pengujian dilakukan dengan menggunakan ROUGE *evaluation toolkit* dengan membandingkan hasil aplikasi ini dengan *summarizer* lainnya yaitu MEAD. Hasil pengujian menunjukkan bahwa metode *centroid-based* dengan *feature* ekstraksi yang diaplikasikan pada tugas akhir ini cenderung lebih rendah dibandingkan dengan *score* yang dihasilkan oleh MEAD.

**Kata kunci** : peringkasan teks, *centroid-based*, *concept*, *feature* ekstraksi