

Abstrak

Evaluasi mesin peringkas teks otomatis adalah proses yang dilakukan untuk menentukan kualitas dari suatu mesin peringkas teks otomatis. Proses ini menghasilkan keluaran berupa nilai yang didapat dengan cara membandingkan antara ringkasan yang dihasilkan oleh mesin peringkas teks otomatis (disebut juga sebagai ringkasan kandidat) dengan ringkasan (ideal) lain yang dibuat oleh manusia (disebut juga sebagai ringkasan referensi) yang berasal dari teks sumber yang sama.

Pada Tugas Akhir penulis mengimplementasikan pengembangan dari metode *Longest Common Subsequences* (LCS) dan *Skip Bigram Co Occurrences Statistics* (SB) dalam teknik ROUGE untuk mengevaluasi mesin peringkas teks otomatis *multi-document*. Dimana pengembangan dari kedua metode tersebut bertujuan untuk menghasilkan suatu cara evaluasi yang memiliki tingkat keterhubungan yang baik dengan *human judgement*, karena informasi yang terambil menjadi semakin banyak.

Pengujian dilakukan dengan menggunakan *Pearson's Coefficient of Correlation* untuk mengukur hubungan linear antara nilai yang diberikan oleh aplikasi dan nilai rata-rata *human judgement* terhadap ringkasan yang sama. Hasil pengujian menunjukkan bahwa evaluasi dengan menggunakan algoritma LCS+SB mampu menghasilkan nilai evaluasi yang lebih tinggi dibandingkan algoritma LCS maupun SB, dengan hasil korelasi yang terkadang lebih tinggi dibandingkan kedua algoritma yang lain. Selain itu, korelasi dapat ditingkatkan dengan melakukan penambahan jumlah ringkasan referensi yang digunakan.

Kata kunci: *Summarization Evaluation, Longest Common Subsequences, Skip Bigram Co Occurrences Statistics, ROUGE, Pearson's Coefficient of Correlation*