

Nazilatul Mahbubah, 2016 *Peringkasan Multi Dokumen Bahasa Indonesia Menggunakan Pendekatan Ekstraksi Fitur Dengan Metode Centroid-Based*. Skripsi ini dibawah bimbingan Badrus Zaman, S.Kom., M.Cs dan Indah Wediningsih, S.Si, M.Kom. Program Studi S1-Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Airlangga.

ABSTRAK

Saat ini, *Information Retrieval* digunakan dalam pengelolaan dokumen di internet dan pengambilan informasi terhadap dokumen-dokumen tersebut. Namun *output* dari *Information Retrieval* tidak sepenuhnya sesuai keinginan pengguna. Jika *output* dokumen sangat banyak, hal ini akan memberatkan pengguna untuk memeriksa dokumen satu per satu. Oleh karena itu diperlukan sistem peringkasan teks otomatis (*automatic text summarization*) yang digunakan untuk meringkas isi dokumen, sehingga pengguna cukup membaca ringkasannya saja.

Untuk menghasilkan sebuah ringkasan dokumen, sistem menghitung bobot tiap kalimat yang merepresentasikan nilai kepentingan kalimat tersebut. Metode yang digunakan untuk menghitung bobot kalimat dalam penelitian ini adalah Metode *Centroid-based*. Tahap dalam system ini untuk menghasilkan sebuah ringkasan adalah melakukan *preprocessing* terhadap dokumen-dokumen, menghitung bobot kalimat menggunakan metode *centroid-based*, dan pemilihan ringkasan menggunakan kombinasi tiga parameter. Dokumen sampel yang digunakan berupa 100 artikel berita *online* yang didapatkan dari situs detik.com.

Proses evaluasi dilakukan dengan membandingkan ringkasan hasil sistem dengan ringkasan yang dibuat oleh tiga *evaluator*. Evaluator adalah orang-orang yang bekerja pada bidang *broadcasting*. Dari proses tersebut akan menghasilkan nilai *F-Measure* yang merepresentasikan tingkat keberhasilan sistem dalam menghasilkan ringkasan. Evaluasi yang dilakukan adalah dengan membandingkan hasil ringkasan yang dihasilkan dengan hasil ringkasan yang dihasilkan oleh fitur *Auto Summarize* Ms Word 2007. Dari penelitian ini diperoleh hasil evaluasi sistem dengan nilai *F-Measure* sebesar 0,47 yang lebih tinggi dari fitur *Auto Summarize* Ms Word 2007 dengan nilai 0,30.

Kata Kunci : *Information Retrieval, Automatic Text Summarization, Centroid-based*