

ABSTRAK

PERBANDINGAN AKURASI DAN PEMBOBOTAN FITUR TEKS PADA PERINGKASAN DOKUMEN BAHASA INDONESIA MENGGUNAKAN ALGORITMA GENETIKA

Oleh:

Gozalli Achmad Irawan

Peringkasan teks otomatis menjadi salah satu kajian yang aktif dilakukan semenjak berkembangnya teknologi jaringan internet. Pesatnya perkembangan jumlah dokumen teks, serta kebutuhan akan mesin pencari yang akurat pada internet mendorong dikembangkannya sebuah metode untuk melakukan peringkasan dokumen teks secara otomatis. Beberapa penelitian mengenai peringkasan teks sudah banyak dilakukan, namun penelitian tersebut masih jarang dilakukan pada dokumen Bahasa Indonesia.

Pada penelitian ini dilakukan peringkasan teks Bahasa Indonesia dengan teknik ekstraksi menggunakan algoritma genetika. Ekstraksi kalimat dilakukan dengan menggunakan lima fitur kalimat, yaitu *positive keyword*, kalimat yang menyerupai judul, kemiripan antar kalimat, kalimat semantik, serta dengan menambahkan satu fitur kalimat frasa. Bobot optimal tiap fitur dihasilkan oleh algoritma genetika. Dari hasil penelitian didapatkan akurasi terbaik pada hasil ringkasan sebesar 91% pada tingkat kompresi sebesar 30%. Peningkatan akurasi hasil ringkasan dengan penambahan fitur frasa terjadi pada tingkat kompresi 10% sebanyak 3% dan pada tingkat kompresi 20% sebanyak 1,5%. Hasil penelitian ini menunjukkan bahwa penambahan fitur frasa dapat meningkatkan akurasi hasil peringkasan dokumen.

Kata kunci: algoritma genetika, ekstraksi frasa, ekstraksi kalimat, peringkasan teks otomatis

ABSTRACT

COMPARISON OF ACCURACY AND TEXT FEATURES WEIGHTING IN SUMMARIZATION ON INDONESIAN DOCUMENT USING GENETIC ALGORITHMS

By:

Gozalli Achmad Irawan

Automatic text summarization becomes an active research conducted since the development of internet network technology. The rapid development of the number of text documents, and the need for an accurate search engine on the internet encouraged the development of automatic text summarization methods. Some research on text summarization has been done, but the research is still rare in Indonesian documents.

In this research applied Indonesian text summarization with extraction technique using genetic algorithms. Sentence extraction was done using the five features of the sentence, which is positive keywords, sentences like the title, the similarity between sentences, semantic sentence, and by adding a sentence phrase feature. Optimal weights of each feature was generated by the genetic algorithm. The result showed the best accuracy on the summary of 91% at compression rate of 30%. Improved accuracy of the summary with the addition of the phrase feature occurs at compression rate of 10% by 3% and compression rate of 20% by 1,5%. These results indicate that the addition of the phrase feature can improve the accuracy of document summarization result.

Keyword: genetic algorithm, phrase extraction, sentence extraction, automatic text summarization,