

INTISARI

IMPLEMENTASI PERINGKAS TEKS OTOMATIS PADA JURNAL BAHASA INDONESIA DENGAN METODE SENTENCE SCORING DAN RANKING

Samani

10/300484/PA/13276

Peringkasan teks merupakan proses untuk mengompres dokumen original kedalam versi yang lebih pendek dengan mengekstrak informasi paling penting dari teks dokumen tersebut. Penelitian ini mengimplementasikan peringkasan teks pada jurnal bahasa Indonesia untuk menghasilkan ringkasan dengan metode *sentence scoring* dan *ranking*. Penelitian ini bertujuan untuk mengetahui akurasi dari ringkasan yang dihasilkan sistem yang menggunakan metode *sentence scoring* dan *ranking*.

Metode *sentence scoring* dan *ranking* termasuk dalam kategori ekstraksi, yaitu dalam menghasilkan ringkasan tidak dilakukan pembuatan kalimat baru, melainkan hanya dilakukan ekstraksi dari kalimat yang memiliki bobot fitur paling tinggi. *Sentence scoring* yang dilakukan yaitu menjumlahkan nilai dari Fitur *thematic*, fitur *position*, fitur *cueword* dan fitur *background*. Dalam penelitian ini dianalisa perbandingan hasil akurasi dari sistem yang menggunakan pembobotan tf dan yang menggunakan pembobotan tfidf sebagai fitur *thematic*, serta yang menggunakan pembobotan paragraf dan pembobotan bab sebagai fitur *position*. Sistem ini diimplementasikan menggunakan bahasa pemrograman PHP.

Hasil penelitian menunjukkan bahwa menggunakan pembobotan tfidf sebagai fitur *thematic* memiliki akurasi yang lebih baik dibanding menggunakan pembobotan tf. Selain itu pembobotan bab sebagai fitur *position* memiliki akurasi yang lebih baik dibandingkan menggunakan pembobotan paragraf. Dari keempat percobaan, diperoleh kesimpulan bahwa kombinasi dari pembobotan tfidf, pembobotan bab, *cueword*, dan *background* memiliki nilai akurasi rata-rata yaitu sebesar 42.19% yang merupakan nilai akurasi yang tertinggi yang diperoleh dalam penelitian ini menggunakan metode *sentence scoring* dan *ranking*.

Kata kunci: peringkasan teks, pembobotan kalimat, jurnal, fitur kalimat, tfidf

ABSTRACT

IMPLEMENTATION OF AUTOMATIC TEXT SUMMARIZATION ON INDONESIAN LANGUAGE JOURNAL USING SENTENCE SCORING AND RANKING METHOD

Samani

10/300484/PA/13276

Text summarization is to compress an original document into a shortened version by extracting the most important information out of the document. This research implements text summarization on Indonesian journal to generate an automatic summary using sentence scoring and ranking method. The purpose of this research is to find out the accuracy of the summary generated by system that using sentence scoring and ranking method.

Sentence scoring and ranking method is included in the category of extraction. This method constructs the summary by taking sentences that has the highest score out of the original document. Sentence scoring summing the value of thematic feature, position feature, cueword feature and background feature. The accuracy ratio of system that using tf scoring as thematic feature and system that using tfidf as thematic value is analyzed in this research, as well as the accuracy of system using section scoring as position feature and system that using paragraph scoring as position value. This sistem is implemented using PHP programming language.

The result shows that using tfidf scoring as thematic feature has better accuracy than using tf scoring. Using section scoring as position feature has better accuracy than using paragraph scoring. The combination of tfidf scoring, section scoring, cueword and background has an avarage of accuracy value 42.19% which is the best results that can be obtained in this research using sentence scoring and ranking method.

Keyword: text summarization, sentence scoring, journal, feature based, tfidf