

Coverage, diversity, and coherence optimization for multi-document summarization

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20448078&lokasi=lokal>

Abstrak

A great summarization on multi-document with similar topics can help users to get useful information. A good summary must have an extensive coverage, minimum redundancy (high diversity), and smooth connection among sentences (high coherence). Therefore, multi-document summarization that considers the coverage, diversity, and coherence of summary is needed. In this paper we propose a novel method on multi-document summarization that optimizes the coverage, diversity, and coherence among the summary's sentences simultaneously. It integrates self-adaptive differential evolution (SaDE) algorithm to solve the optimization problem. Sentences ordering algorithm based on topical closeness approach is performed in SaDE iterations to improve coherences among the summary's sentences. Experiments have been performed on Text Analysis Conference (TAC) 2008 data sets. The experimental results showed that the proposed method generates summaries with average coherence and ROUGE scores 29-41.2 times and 46.97-64.71% better than any other method that only consider coverage and diversity, respectively.

Peringkasan yang baik terhadap dokumen-dokumen dengan topik yang seragam dapat membantu pembaca dalam memperoleh informasi secara cepat. Ringkasan yang baik merupakan ringkasan dengan cakupan pembahasan (coverage) yang luas dan dengan tingkat keberagaman (diversity) serta keterhubungan antarkalimat (coherence) yang tinggi. Oleh karena itu dibutuhkan metode peringkasan multi-dokumen yang mempertimbangkan tingkat coverage, diversity, dan coherence pada hasil ringkasan. Pada paper ini dikembangkan sebuah metode baru dalam peringkasan multi-dokumen dengan mengoptimasi tingkat coverage, diversity, dan coherence antarkalimat hasil ringkasan secara simultan. Optimasi hasil ringkasan dilakukan dengan menggunakan algoritma self-adaptive differential evolution (SaDE). Algoritma pengurutan kalimat yang menggunakan pendekatan topical closeness juga diintegrasikan ke dalam tiap iterasi algoritma SaDE untuk meningkatkan koherensi antarkalimat hasil ringkasan. Uji coba dilakukan pada 15 topik dataset Text Analysis Conference (TAC) 2008. Hasil uji coba menunjukkan bahwa metode yang diusulkan dapat menghasilkan ringkasan dengan rata-rata koherensi 29-41,2 kali lebih tinggi serta skor ROUGE 46,97-64,71% lebih besar dibandingkan dengan metode yang hanya mempertimbangkan coverage dan diversity hasil ringkasan.