

Ringkasan multi dokumen berbahasa indonesia secara otomatis menggunakan metode latent semantic analysis dan centroid-based summarization = Automatic Indonesian multi document summarization using latent semantic analysis method and centroid-based summarization

Tinaliah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20350297&lokasi=lokal>

Abstrak

Tesis ini menjelaskan mengenai penggabungan antara metode Latent Semantic Analysis dan Centroid-based Summarization dalam proses peringkasan multidokumen berbahasa Indonesia. Evaluasi peringkasan multidokumen berbahasa Indonesia akan dibandingkan dengan ringkasan yang dihasilkan oleh sistem terhadap hasil ringkasan referensi yang dibuat oleh ahli Bahasa Indonesia dengan mengukur akurasi kalimat yang sama muncul pada ringkasan yang dihasilkan oleh sistem terhadap ringkasan referensi. Besarnya korpus data training yang kita punya dapat meningkatkan nilai cosine similarity tiap kalimat yang dihasilkan pada metode LSA. Secara keseluruhan akurasi pada penggabungan antara metode latent semantic analysis dan metode centroid-based summarization menghasilkan akurasi yang lebih baik sebesar 26.62% dibandingkan dengan metode centroidbased summarization sebesar 23.81%, dengan selisih rata-rata akurasi penggabungan antara metode latent semantic analysis dan metode centroid-based summarization dengan rata-rata akurasi metode centroid-based summarization adalah sebesar 2,82%.

<hr>

This study describe the combination of Latent Semantic Analysis method and Centroid-based Summarization in multi-document summarization with Indonesia language. The evaluation result of the system is taken from comparing the summary made by system and the summary made by human. Larger of corpus training data will increase cosine similarity of each sentence in LSA. Overall, combination method of latent semantic analysis and centroid-based summarization is more accurate 26.62% compared to the centroid-based summarization method 23.81 %, with diference of the accuracy average between combination method of latent semantic analysis and centroid-based summarization and accuracy average of method of centroid-based summarization is 2.82%.