

Peringkasan multi-dokumen untuk bahasa Indonesia menggunakan teknik centroid-based summarization dan teknik k-means-based summarization

Wisnu Linggakusuma Wardhana, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123671&lokasi=lokal>

Abstrak

Tersedianya sumber informasi yang tidak terbatas pada saat ini, menjadikan perolehan informasi melibatkan banyak sumber informasi. Hal-hal tersebut memicu penelitian mengenai metode peringkasan dokumen yang semula ditujukan untuk membuat sebuah ringkasan dari sebuah dokumen menjadi metode peringkasan yang ditujukan untuk menghasilkan ringkasan dari banyak dokumen. Peringkasan multi-dokumen merupakan suatu metode yang ditujukan untuk menyampaikan informasi-informasi utama dari banyak dokumen dalam ringkasan. Penelitian yang dilakukan oleh penulis ini merupakan penelitian dengan topik peringkasan multi-dokumen untuk dokumen berbahasa Indonesia. Pada penelitian ini, penulis menggunakan dua buah teknik peringkasan multi-dokumen yaitu centroidbased summarization dan k-means-based summarization. Pada penelitian ini penulis mencoba untuk mengaplikasikan kedua teknik tersebut untuk membuat ringkasan dari dokumen berbahasa Indonesia.

Untuk mengukur kualitas ringkasan yang dihasilkan oleh kedua teknik tersebut, penulis membuat ringkasan referensi untuk masing-masing sumber dokumen yang dibuat secara manual sebagai perbandingan. Hasilnya adalah pada teknik centroidbased summarization, kualitas ringkasan yang dihasilkan akan semakin bagus jika kelompok dokumen yang digunakan sebagai masukan berisi dokumen-dokumen yang relevan terhadap topik. Sedangkan pada teknik k-means-based summarization, kualitas ringkasan yang dihasilkan akan semakin bagus jika kelompok dokumen yang digunakan sebagai masukan merupakan kelompok dokumen yang besar (lebih banyak mengandung dokumen/kalimat). Evaluasi terhadap kualitas ringkasan juga dilakukan dengan menggunakan juri/penilai manusia. Hasilnya adalah pada teknik centroid-based summarization, para juri menilai ringkasan yang dihasilkan sudah bagus. Pada teknik k-means-based summarization dengan 10% compression rate, para juri menilai bahwa ringkasan.