

Penerapan automatic sumarization memakai metode text extraction dengan menggunakan centroid value, positional value dan first sentence overlap untuk artikel ilmiah berbahasa Indonesia

Hendrik Ganda Wijaya, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=124532&lokasi=lokal>

Abstrak

Summary yang dimaksud di dalam tugas akhir ini adalah sebuah abstraksi dari dokumen yang digunakan untuk menentukan apakah dokumen tersebut merupakan dokumen yang ingin dibaca oleh user atau tidak. Dengan adanya summary dari dokumen tersebut maka user tidak perlu menghabiskan waktu untuk membaca seluruh dokumen untuk mencari informasi. User cukup membaca summary-nya saja untuk melihat apakah memang dokumen tersebut merupakan dokumen yang ingin dibaca oleh user. Pertanyaan yang akan dijawab melalui tugas akhir ini adalah bagaimanakah proporsi yang ideal untuk centroid value, positional value dan first sentence overlap agar dapat mencapai hasil yang optimal di dalam menghasilkan summary? Dan sejauh manakah kinerja dari program yang menerapkan metode text extraction di dalam menghasilkan summary? Di dalam penelitian ini penulis mengambil sebuah algoritma dari sebuah artikel mengenai penelitian yang dilakukan oleh Dragomir R Radev dan Weiguo Fan.

Algoritma tersebut adalah metode text extraction dengan menggunakan centroid value, positional value dan first sentence overlap. Di dalam tugas akhir ini, penulis mencoba untuk menerapkannya di dalam artikel ilmiah berbahasa Indonesia. Langkah-langkah yang dilakukan penelitian ini mencakup : menghilangkan stop word dari dokumen input, melakukan stemming pada dokumen input, mencari centroid value, positional value dan first sentence overlap dan total nilai dari masing-masing kalimat, kemudian menyusun summary dari kalimat-kalimat dengan nilai yang tertinggi.

Hasil penelitian ini menunjukkan bahwa proporsi yang paling ideal untuk centroid value, positional value dan first sentence overlap adalah 211 dan 212. Hasil penelitian ini juga menunjukkan bahwa kinerja dari program yang menerapkan metode text extraction ini cukup optimal (dengan nilai rata-rata 3,6 dari skala 1-5). Dari hasil penelitian ini, penulis menyimpulkan bahwa pengujian yang dilakukan terhadap summary bersifat subyektif. Penulis juga menyimpulkan bahwa terdapat hubungan yang erat antara tingkat readability input dengan tingkat readability output. Untuk penelitian selanjutnya, penulis menyarankan untuk menambahkan unsur format dan keyword ke dalam unsurpenghitungan. Selain itu penulis juga menyarankan untuk mencari penguji yang berpengalaman di dalam membuat summary. Agar hasil pengujian menjadi lebih terjamin.