

Perbandingan kinerja algoritma agglomerative hierarchical pada document clustering

Catur Adi Nugroho, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=124011&lokasi=lokal>

Abstrak

Laporan Tugas Akhir ini berisi mengenai penelitian yang dilakukan oleh penulis dalam membandingkan kinerja beberapa algoritma, yang tergolong ke dalam agglomerative hierarchical, dalam hal melakukan clustering dokumen untuk mendapatkan solusi hierarchical cluster. Algoritma yang diperbandingkan adalah algoritma single link, complete link, dan average. Proses perbandingan dilakukan berdasarkan kualitas cluster yang dihasilkan pada sejumlah dataset. Hasil penelitian menunjukkan bahwa algoritma average merupakan algoritma yang terbaik dalam menghasilkan solusi hierarchical cluster, diikuti oleh algoritma single link, dan algoritma complete link.

Penelitian ini juga melakukan penerapan teknik dalam feature selection untuk melihat seberapa besar efisiensi yang bisa diperoleh tanpa harus mengurangi kualitas solusi cluster yang dihasilkan. Teknik feature selection yang dipergunakan meliputi pembatasan nilai Document Frequency dan Information Gain. Efisiensi yang dilakukan oleh kedua teknik ini adalah melakukan pemilihan kata-kata yang penting saja yang diikutsertakan dalam proses clustering. Penelitian ini mencoba melihat seberapa besar efisiensi yang dapat diperoleh masing-masing teknik dan kemudian membandingkannya satu sama lain. Hasil penelitian menunjukkan bahwa kedua teknik baik pembatasan nilai Document Frequency dan Information Gain mampu melakukan efisiensi pada titik-titik reduksi yang sudah ditetapkan yaitu sebesar 10%-90% dari jumlah kata unik yang ada tanpa kualitas yang berkurang. Selain itu, hasil penelitian menunjukkan bahwa kedua teknik ini sama efektifnya dalam mereduksi dimensi dari dataset yang dipergunakan.