

DOCUMENT ADDITIONAL EVALUATION IN INFORMATION RETRIEVAL SYSTEM

Julio Adisantoso, Yeni Herdiyeni, Ika Kartika

ABSTRAK

Saat ini pengguna cenderung menyukai pencarian berdasarkan makna. Hal ini disebabkan oleh adanya masalah sinonim dan polisemi dalam pemilihan penggunaan kata. Salah satu teknik yang mencoba mengatasi masalah tersebut Latent Semantic Indexing (LSI). Dalam pengaplikasiannya, LSI dapat menggunakan Singular Value Decomposition untuk mengestimasi struktur penggunaan kata dalam dokumen. Cara yang paling tepat untuk menambahkan dokumen atau istilah adalah melalui penghitungan ulang SVD (recomputing SVD). Namun hal ini menjadi kendala karena dibutuhkannya memory yang cukup besar dan waktu yang semakin lama untuk menghitung ulang matriks istilah-dokumen yang semakin besar. Cara lain yang dapat digunakan untuk mengatasi kendala tersebut adalah dengan menggunakan teknik folding-in. Tujuan penelitian ini adalah untuk melihat pengaruh penambahan dokumen terhadap hubungan tersembunyi antara semua istilah yang secara kontekstual berdekatan artinya dengan menggunakan teknik folding-in.

Tercakup didalamnya proses pembentukan matriks istilah-dokumen dengan menggunakan parsing, penghilangan stop list, serta stemming. Pembobotan istilah untuk dokumen menggunakan skema pembobotan l_{xn} , sedangkan pembobotan istilah untuk kueri menggunakan skema pembobotan c_{fx} . Pengujian menggunakan 150 dokumen untuk membangkitkan matriks istilah-dokumen asal dan 110 dokumen untuk evaluasi penambahan dokumen. Pengukuran kinerja temu kembali dilakukan dengan menggunakan average precision untuk mengetahui rank optimal terhadap sepuluh kueri. Ternyata pada penelitian ini dengan pemilihan rank yang kecil akan memberikan hasil dengan tingkat akurasi yang cukup tinggi dengan tertanganinya masalah sinonim. Penambahan dokumen dengan folding-in memberikan hasil yang memuaskan. Melalui penambahan 110 dokumen ke dalam koleksi sebagian besar nilai recall bisa dipertahankan apabila menggunakan rank yang kecil, meskipun terdapat distorsi seiring dengan penambahan dokumen.