

Variable latent semantic indexing sebagai teknik text retrieval yang menggunakan singular value decomposition

Suharto Anggono, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=126245&lokasi=lokal>

Abstrak

Sistem text retrieval (perolehan kembali teks) menanggapi query terhadap koleksi dokumen teks. Dua teknik yang telah diperkenalkan yang dapat digunakan pada text retrieval dan melibatkan penggunaan singular value decomposition (SVD, dekomposisi nilai singular) adalah variable Latent Semantic Indexing (VLSI) serta Latent Semantic Indexing (LSI), pendahulunya. Dokumen dapat dideskripsikan oleh kumpulan term, misalnya berupa kata. Dari dokumendokumen yang ada, dapat dibentuk matriks term-dokumen, A , yang berisi bobot term berdasarkan kemunculannya dalam dokumen. Pada VLSI, sebagaimana LSI, digunakan aproksimasi rank-rendah terhadap A untuk dicocokkan dengan vektor query. Namun, berbeda dengan LSI, aproksimasi pada VLSI tergantung pada distribusi probabilitas vektor query. Distribusi itu dicirikan oleh matriks cooccurrence, CQ . Aproksimasi untuk A itu bisa didapatkan melalui SVD terhadap $CQ^{1/2}A$. Telah dilakukan penelitian terhadap VLSI dengan mempelajari literatur, mengerjakan secara manual prosedur VLSI, dan melakukan percobaan penggunaan VLSI. Percobaan dilakukan dengan koleksi 100 dokumen yang pernah digunakan pada mata kuliah Pemrosesan Teks, koleksi 9 dokumen berupa judul technical memo, koleksi MED, dan koleksi CACM. Dari analisis, ditemukan bahwa yang dikalikan dengan A untuk kemudian di-SVD tidak harus $CQ^{1/2}$. Dari percobaan dengan koleksi MED dan koleksi CACM, ditemukan bahwa perolehan dokumen dengan VLSI bisa bagus dalam hal precision pada sedikit dokumen berperingkat teratas.