

Pengindeksan berdasarkan maximal frequencies dalam sistem temu kembali informasi teks berbasis Indonesia

Dwi Astuti Aprijani, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=125374&lokasi=lokal>

Abstrak

Penelitian ini mengkaji penggunaan Maximal Frequent Sequences (MFS) sebagai indeks dalam koleksi dokumen untuk Sistem Temu-kembali Informasi teks berbahasa Indonesia. Indeks, sebagai representasi dokumen, harus menggambarkan informasi keseluruhan yang terkandung dalam dokumen. MFS sebagai salah satu cara untuk mengekspresikan multi kata (frase) dapat digunakan sebagai indeks dalam koleksi dokumen untuk merepresentasikan isi dokumen. Kekuatan utama MFS dapat membentuk indeks yang sangat solid karena menoleransi adanya kata-kata pemisah di antara suatu pasangan kata dan jumlah istilah yang digunakan sebagai indeks sedikit. Hasil uji coba terhadap 1162 dokumen ilmiah dan 2999 dokumen berita menunjukkan bahwa jumlah istilah pada pengindeksan berdasarkan MFS jauh lebih sedikit dibandingkan jumlah istilah pada pengindeksan berdasarkan frekuensi kata (term frequency), yaitu 31% untuk koleksi dokumen ilmiah non-stemming, 48% untuk koleksi dokumen ilmiah stemming, 30% untuk koleksi dokumen berita non-stemming, dan 59% untuk koleksi dokumen berita stemming. Pada koleksi dokumen ilmiah, MFS dengan nilai threshold yang lebih kecil relatif mampu memperbaiki nilai presisi, baik presisi rata-rata sesudah memproses 20 dokumen maupun presisi rata-rata sesudah memproses dokumen dengan jumlah yang sesuai dengan jumlah dokumen relevan. Pada koleksi dokumen berita, MFS dengan nilai threshold yang kecil relatif memperbaiki nilai presisi rata-rata sesudah memproses dokumen dengan jumlah yang sesuai dengan jumlah dokumen relevan, namun threshold yang lebih besar justru memperbaiki presisi rata-rata sesudah memproses 10 dokumen. Efek stemming meningkatkan presisi rata-rata sesudah memproses 10 dokumen dan presisi rata-rata sesudah memproses dokumen dengan jumlah yang sesuai dengan jumlah dokumen relevan pada sistem temu-kembali informasi ini.