

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi ini tidak dipublikasikan, namun tersedia di perpustakaan dalam lingkungan Universitas Airlangga, diperkenankan untuk dipakai sebagai referensi kepustakaan, tetapi pengutipan harus seizin penyusun dan harus menyebutkan sumbernya sesuai kebiasaan ilmiah.

Dokumen skripsi ini merupakan hak milik Universitas Airlangga.



Triyah Fatmawati, 2016. Penerapan Metode *Common Phrase Index* pada *Query* Frase untuk Penemukembalian Informasi. Skripsi ini dibawah bimbingan Badrus Zaman, S.Kom., M.Cs. dan Indah Werdiningsih, S.Si., M.Kom. Program Studi S1 Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Airlangga.

ABSTRAK

Seiring berkembangnya teknologi, proses pencarian informasi pada teks berita dapat dilakukan dengan mudah, karena teks berita tidak hanya disebarakan dalam bentuk cetak, seperti surat kabar, melainkan juga dalam bentuk elektronik yang dapat diakses menggunakan *search engine*. Dalam proses pencarian dokumen relevan pada *search engine*, frase sering digunakan sebagai *query*. Jumlah kata yang menyusun *query* frase sangat memengaruhi tingkat relevansi dokumen yang dihasilkan. Oleh karena itu dilakukan penelitian ini untuk menganalisis penerapan metode *common phrase index* pada sistem temu kembali informasi.

Penelitian akan dilakukan pada teks berita berbahasa Inggris sebanyak 100 dokumen dan diimplementasikan pada sebuah *prototype* untuk mengetahui tingkat relevansi dokumen yang dihasilkan. Tahap awal adalah penentuan *common word*, yang dilakukan berdasarkan jumlah keseluruhan frekuensi. Kemudian pada tahap implementasi, sistem dibangun dengan tahapan *pre-processing*, *indexing*, perhitungan *term weighting*, dan perhitungan *cosine similarity*. Sistem akan menampilkan dokumen hasil pencarian secara berurutan berdasarkan nilai *cosine similarity*, dari nilai terbesar hingga terkecil. Untuk menguji fungsionalitas sistem, dilakukan pengujian menggunakan metode *black box* dengan teknik *equivalence class partitioning*. Selanjutnya dilakukan uji coba menggunakan 100 dokumen dan 20 *query*. Hasil uji coba sistem tersebut digunakan untuk tahap evaluasi. Proses evaluasi terhadap sistem dilakukan dengan membandingkan dokumen relevan dengan hasil pencarian oleh sistem. Penentuan dokumen relevan dilakukan dengan menghitung nilai *kappa statistic* dengan *judge* sebanyak 3 orang.

Dari perhitungan tersebut dihasilkan nilai *kappa statistic* sebesar 0,71 sehingga dokumen relevan tersebut layak untuk evaluasi sistem. Kemudian dilakukan perhitungan *precision*, *recall*, dan *F-measure* pada 2 skenario. Skenario pertama menghasilkan *precision* sebesar 0,43, *recall* sebesar 0,50, dan *F-measure* sebesar 0,46. Skenario kedua menghasilkan *precision* sebesar 0,43, *recall* sebesar 0,49, dan *F-measure* sebesar 0,46. Dari hasil perhitungan tersebut dapat dikatakan bahwa tingkat keberhasilan sistem dalam menghasilkan dokumen relevan adalah rendah.

Kata Kunci — Frase, *Common Phrase Index*, Dokumen Relevan.