

Information retrieval of text document with weighting tf-idf and lcs

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20448250&lokasi=lokal>

Abstrak

Information retrieval of text document requires a method that is able to restore a number of documents that have high relevance according to the user's request. One important step in the process is a text representation of the weighting process. The use of LCS in Tf-Idf weighting adjustments considers the appearance of the same order of words between the query and the text in the document. There is a very long document but irrelevant cause weight produced is not able to represent the value relevance of documents. This research proposes the use of LCS which gives weight to the word order by considering long documents related to the average length of documents in the corpus. This method is able to return a text document effectively. Additional features of word order by normalizing the ratio of the overall length of the document to the documents in the corpus generate values of precision and recall as well as the method of Tasi et al.

Sistem temu kembali dokumen teks membutuhkan metode yang mampu mengembalikan sejumlah dokumen yang memiliki relevansi tinggi sesuai dengan permintaan pengguna. Salah satu tahapan penting dalam proses representasi teks adalah proses pembobotan. Penggunaan LCS dalam penyesuaian bobot Tf-Idf mempertimbangkan kemunculan urutan kata yang sama antara query dan teks di dalam dokumen. Adanya dokumen yang sangat panjang namun tidak relevan menyebabkan bobot yang dihasilkan tidak mampu merepresentasikan nilai relevansi dokumen. Penelitian ini mengusulkan penggunaan metode LCS yang memberikan bobot urutan kata dengan mempertimbangkan panjang dokumen terkait dengan rata-rata panjang dokumen dalam korpus. Metode ini mampu melakukan pengembalian dokumen teks secara efektif. Penambahan fitur urutan kata dengan normalisasi rasio panjang dokumen terhadap keseluruhan dokumen dalam korpus menghasilkan nilai presisi dan recall yang sama baiknya dengan metode Tasi dkk.