Algoritme pelacakan kata dengan toleransi kesalahan

Lim Yohanes Stefanus, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=90071&lokasi=lokal>

Abstrak

Makalah ini membahas salah satu masalah yang sangat umum dalam pemrosesan teks dengan bantuan komputer, yaitu bagaimana melacak dengan cepat suatu kata atau ungkapan tertentu dalam suatu berkas teks yang besar. Dalam banyak situasi, kata yang akan dilacak itu mungkin tidak diketahui dengan persis.

Misalnya, ejaan kata itu sendiri kurang betul, atau teks tersebut mengandung kesalahan-kesalahan penulisan. Jadi bagaimana seharusnya kriteria untuk menentukan kesuksesan pelacakan? Jika sistem disuruh mencari kata "objek" dalam teks yang mengandung kata "obyek", apakah sistem harus menyatakan bahwa pelacakan itu berhasil menemukan kata yang sesuai?

Untuk membantu manusia yang pada umumnya bersifat mudah membuat kesalahan, sistem pelacakan teks sebaiknya bekerja berdasarkan kriteria kesuksesan pelacakan yang longgar. Tentu saja kelonggaran ini harus dibatasi. Kita tidak menginginkan sistem yang melemparkan ratusan jawaban untuk satu permintaan melacak kata yang kita yakin ejaannya benar. Sebaiknya sistem itu bekerja sebagai berikut: mula-mula ia akan berusaha menemukan kata yang persis; kalau kata yang persis tidak ada, ia akan berusaha menemukan kata yang ejaannya paling mendekati menurut ukuran kedekatan tertentu; dan seterusnya. Yang menjadi persoalan adalah bagaimana bentuk ukuran kedekatan tersebut.