

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi di bidang komputasi dan telekomunikasi telah memberi kemudahan kepada setiap orang untuk menyampaikan dan mendapatkan informasi. Kemudahan ini menyebabkan informasi menjadi semakin banyak dan beragam. Informasi dapat berupa dokumen, artikel, surat, cerita, laporan penelitian, data keuangan dan lain-lain.

Ketersediaan informasi yang semakin banyak menjadikan ringkasan sebagai kebutuhan yang sangat penting. Dengan adanya ringkasan, pembaca dapat dengan cepat dan mudah memahami makna sebuah teks tanpa harus membaca keseluruhan teks. Hal ini dapat menghemat waktu pembaca karena dapat menghindari pembacaan teks yang tidak relevan dengan informasi yang diharapkan oleh pembaca, terutama jika informasi tersedia di internet cukup banyak.

Berbagai metode telah diterapkan dan masih terus dikembangkan oleh para peneliti tentang peringkasan. Salah satunya adalah peringkasan dokumen yang kemudian akan menghasilkan informasi yang penting atau inti kalimat dari dokumen dengan menggunakan komputer. Tujuannya adalah mengambil sumber informasi dengan mengutip sebagian besar isi yang penting untuk mencari kalimat inti dari sebuah dokumen artikel dan menampilkan kepada pembaca dalam bentuk yang ringkas sesuai dengan kebutuhan pembaca. Dengan demikian teknologi ini dapat membantu pembaca untuk menyerap informasi yang ada dalam artikel melalui ringkasan tanpa harus membaca seluruh isi dokumen.

Pencarian otomatis kalimat inti ini dibutuhkan pembobotan pada masing-masing term sehingga metode *term frequency – invers document frequency (tf-idf)* digunakan sebagai penentu bobot masing-masing term tersebut. Sehingga penulis mengajukan judul tugas akhir “PENCARIAN OTOMATIS KALIMAT INTI DALAM DOKUMEN TUNGGAL BERBAHASA INDONESIA”.

1.2 Rumusan Masalah

Dari latar belakang yang sudah dijelaskan di atas diperoleh rumusan masalah sebagai berikut

- a. Bagaimana mengolah dokumen dengan teks mining.
- b. Bagaimana menghitung bobot yang dimiliki suatu dokumen dalam keseluruhan teks menggunakan metode TF-IDF.
- c. Bagaimana mengurutkan bobot yang diperoleh dari metode TF-IDF.

1.3 Batasan Masalah

- a. Input hanya dari satu dokumen atau satu artikel saja.
- b. Dokumen berupa artikel dari website/situs olahraga.
- c. Dokumen yang digunakan sebagai data set mengambil dari situs Goal.com.
- d. Database untuk kata-kata yang tidak relevan mengambil dari kamus bahasa indonesia.
- e. Input dan output dalam bahasa indonesia.
- f. Tidak menangani kesalahan penulisan kata.
- g. Singkatan di anggap satu kata.
- h. Hanya melayani list dokumen.

1.4 Tujuan Penelitian

- a. Mengolah dokumen dengan teks mining dan melakukan pembobotan dengan tf-idf.
- b. Melakukan pencarian otomatis kalimat inti dalam sebuah dokumen tunggal berbahasa indonesia.
- c. Mengambil sumber informasi dengan mengutip sebagian besar isi yang penting dan menampilkan kepada pembaca dalam bentuk yang ringkas sesuai dengan kebutuhan pembaca.

1.5 Manfaat Penelitian

- a. Dapat membantu pembaca untuk menyerap informasi yang ada dalam artikel melalui ringkasan tanpa harus membaca seluruh isi dokumen.

- b. Dapat menghemat waktu pembaca karena dapat menghindari pembacaan teks yang tidak relevan dengan informasi yang diharapkan oleh pembaca.