

BAB I PENDAHULUAN

1.1 Latar Belakang

Informasi menjadi kebutuhan pokok masyarakat saat ini, dengan informasi masyarakat dapat mengetahui segala peristiwa yang terjadi di dunia. Informasi kini bisa didapatkan melalui internet yang diakses kapan saja secara mudah. Berita online merupakan salah satu media yang menyajikan informasi terkini dengan berbagai topik berita. Saat ini banyaknya situs berita yang terdaftar pada *www (world wide web)* memungkinkan hasil pencarian berita menjadi sangat luas. Situs-situs tersebut belum tentu memberikan sumber berita yang jelas, sehingga pengguna harus memilah berita dari beberapa situs yang terpercaya [1]. Untuk itu, diperlukan suatu sistem pencarian berita dari situs berita nasional yang sudah terpercaya. Sistem pencarian tersebut menghasilkan beberapa berita yang diurutkan berdasarkan tingkat relevansi terhadap query. Relevan atau tidaknya berita akan diketahui secara jelas setelah pengguna membaca satu persatu isi berita yang disajikan sehingga membutuhkan waktu yang banyak. Sistem yang dibutuhkan adalah Sistem Temu-Balik Informasi (*Information retrieval / IR*) [3].

Information Retrieval atau Temu Balik Informasi merupakan bagian dari computer science tentang pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. *Information retrieval* adalah suatu system yang mampu melakukan penyimpanan, pencarian, dan pemeliharaan informasi. Secara teknis, tujuan *Information Retrieval* adalah mencocokkan (matching) term atau istilah yang dibangun (query) dengan term atau indeks yang ada dalam dokumen, sehingga dengan kecocokan tersebut maka dokumen-dokumen yang relevan akan terambil (retrived) dari database. Terambilnya dokumen berdasarkan permintaan pengguna dengan harapan bahwa content atau isi dari dokumen yang terambil tersebut relevan dengan kebutuhan informasi pencari informasi [2].

Stemming merupakan proses pengubahan kata menjadi kata dasar. Tujuan dari proses stemming adalah menghilangkan imbuhan-imbuhan seperti prefiks, infiks, sufiks, dan konfiks pada setiap kata[3]. Terdapat banyak stemmer yang sudah bisa digunakan untuk bahasa Indonesia. Salah satu contohnya adalah algoritma stemming *Nazief & Adriani*. Algoritma stemming

Nazief dan Adriani sebagai algoritma stemming untuk teks berbahasa Indonesia yang memiliki kemampuan persentase keakuratan lebih baik dari algoritma lainnya [10].

Untuk Menghitung kemiripan dokumen salah satunya dengan menggunakan metode *Dice Similarity*. *Dice Similarity* adalah metode untuk melihat tingkat kedekatan atau kesamaan (smilarity) term dengan cara pembobotan term. Dokumen dipandang sebagai sebuah vektor yang memiliki magnitude (jarak) dan direction (arah) [14]. Pada tahun 2016 Fatkhul Amin, Purwatiningsih, Edy Winarno melakukan penelitian yang berjudul Rancang Bangun Sistem Temu Kembali Informasi (*Information Retrieval System*) Dokumen Berbahasa Jawa menggunakan Metode DICE Similarity dengan tujuan mempermudah dalam melakukan pencarian dokumen teks berbahasa Jawa penelitian tersebut menghasilkan rata-rata recall = 0,04 dan rata-rata precision = 0,83[14].

Selain metode *dice similarity* terdapat beberapa metode lain yang dapat menghitung kemiripan antara *query* dengan dokumen seperti *Cosine similarity*. Penelitian yang telah dilakukan tentang *cosine similarity*, yaitu penelitian yang dilakukan oleh Viko Basmalah Wicaksono, dkk [8], yang berjudul Analisis Perbandingan Metode Vector Space Model dan Weighted Tree Similarity dengan *Cosine Similarity* pada kasus Pencarian Informasi Pedoman Pengobatan Dasar di Puskesmas memberikan nilai rata-rata *precission* 44,82983% dan *recall* 99,08165%

Dari uraian latar belakang tersebut maka penulis akan melakukan penelitian yang berjudul **“Penerapan Algoritma Stemming Nazief & Adriani untuk Membandingkan Metode Cosine Similarity dan Dice Similarity pada Pencarian Dokumen Berita”**. Pada penelitian ini peneliti akan melakukan perbandingan metode *Cosine Similarity* dan *Dice Similarity* dengan menggunakan Metode *Stemming Nazief dan Adriani* pada proses stemming. Diharapkan dengan adanya penelitian ini dapat mengetahui akurasi perbandingan dan waktu proses perbandingan antara metode *Cosine Similarity* dan *Dice Similarity*.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Tingkat kemiripan dari suatu dokumen dapat dideteksi menggunakan metode *Cosine Similarity* dan *Dice Similarity*. Metode *Cosine* dan *Dice Similarity* merupakan metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua vector dengan menggunakan kata kunci. *Nazief & Adriani* merupakan Algoritma stemming pada Bahasa Indonesia dengan tingkat persentase keakuratan yang lebih baik. Dengan permasalahan tersebut maka dilakukan penelitian untuk mencari yang lebih baik antara metode *Cosine* dan *Dice Similarity* jika menggunakan *Stemming Nazief & Adriani* pada proses Stemming.

1.2.2 Metode Usulan

Metode yang di usulkan yaitu penerapan metode *Cosine similarity* dan *Dice Similarity* untuk menghitung kemiripan dokumen, serta menerapkan metode *stemming Nazief & Andriani* untuk meningkatkan proses preprocessing.

1.2.3 Pertanyaan Penelitian

1. Bagaimana perbandingan akurasi dan waktu pencarian teks menggunakan *Cosine similarity* dan *Dice Similarity* dengan menerapkan metode *Stemming Nazief & Adriani*?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Dari pertanyaan penelitian di atas maka tujuan dari penelitian ini yaitu mengetahui perbandingan akurasi dan waktu pencarian teks menggunakan *Cosine* dan *Dice Similarity* dengan menerapkan metode Stemmer *Nazief & Adriani*.

1.3.2 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah memudahkan pemilihan metode yang digunakan untuk mengukur kemiripan antara *query* dengan dokumen dan sebagai referensi pada perbandingan metode *cosine* dan *dice* dalam system temu kembali informasi.

1.4 Batasan Masalah

1. Dokumen berita diambil dari situs <https://pesona.travel.com>

1.5 Sistematika Proposal

Adapun sistematika yang digunakan dalam penulisan skripsi adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini akan menjelaskan latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab akan menjelaskan mengenai konsep-konsep dan teori-teori yang mendukung dalam penyusunan penelitian ini. Serta mendasari teori-teori yang akan digunakan dalam menyelesaikan permasalahan.

BAB III METODE USULAN

Pada bab akan menjelaskan mengenai rancangan sistem yang akan dibuat, termasuk adanya tahapan analisa permasalahan, perancangan sistem, pengujian sistem, analisa evaluasi pengujian dan penarikan kesimpulan.

BAB IV IMPLEMENTASI SISTEM

Pada bab ini akan menjelaskan implementasi pembuatan aplikasi berdasarkan hasil rancangan yang telah di buat sebelumnya.

BAB IV KESIMPULAN

Bab ini berisi tentang kesimpulan dan saran dari penelitian