

BAB I PENDAHULUAN

1.1 Latar Belakang

Mesin pencari *web* saat ini merupakan alat bantu dalam melakukan pencarian informasi di internet. Mulai dari kalangan akademisi ataupun profesional hingga orang biasa, mesin pencari *web* sangat dibutuhkan dalam membantu mencari informasi cepat dan aktual. Dengan menuliskan satu atau beberapa gabungan kata kunci dari informasi yang akan dicari, mesin pencari *web* akan memberikan informasi sesuai dengan kata kunci yang dimasukkan [1].

Mesin pencari *web* bekerja dengan cara menyimpan informasi tentang banyak halaman *web*, yang diambil langsung dari *World Wide Web* (WWW). Halaman-halaman ini diambil dengan menggunakan *web crawler* yang mengikuti setiap pranala/link yang dilihatnya. Isi setiap halaman kemudian dianalisis untuk menentukan cara indeks-nya (misalnya, kata-kata diambil dari judul, sub judul, atau *field* khusus yang disebut *meta tag*). Data tentang halaman *web* disimpan dalam sebuah *database* indeks untuk digunakan dalam pencarian selanjutnya.

Query yang dimasukkan oleh pengguna saat melakukan pencarian pada umumnya pendek, terdiri dari satu atau dua kata saja. Padahal, jumlah dokumen relevan yang diperoleh dipengaruhi oleh jumlah kata kunci dalam *query*. Hal ini mengakibatkan dokumen relevan yang terpanggil hanya sedikit karena sedikitnya informasi dalam *query* yang dimasukkan pengguna, sehingga pengguna harus mengulang pencarian dengan *query* yang berbeda sampai informasi yang didapatkan benar-benar relevan atau sesuai dengan yang diinginkan oleh pengguna.

Pencarian informasi merupakan kegiatan yang tidak bisa dipisahkan dari kehidupan manusia untuk memenuhi kebutuhan informasinya, terutama dalam mencari informasi pariwisata. Informasi menjadi faktor penting dalam perkembangan dunia pariwisata. Sektor pariwisata merupakan faktor penting bagi pendapatan daerah dan negara. Indonesia adalah negara yang memiliki tempat wisata yang beragam, antara lain wisata alam dan wisata budaya. Masing-masing dari wisata tersebut memiliki keindahan dan ciri khas tersendiri. Informasi mengenai wisata di Indonesia ini menuntut akan ketersediaan mesin pencari yang

dapat melakukan pencarian informasi pariwisata dan memberikan hasil pencarian yang relevan dan sesuai dengan keinginan pengguna, sehingga memudahkan pengguna dalam mencari informasi pariwisata.

Vector space model merupakan salah satu model yang digunakan untuk mengetahui kemiripan dokumen [2]. Dalam konteks yang lebih spesifik, *vector space model* menghasilkan nilai *recall* yang besar [3]. Pada pemodelan ini, pembuatan indeks dilakukan dengan cara mengubah berbagai bentuk kata menjadi kata dasar sehingga membantu mempersingkat proses saat mencocokkan antara dokumen dan *query* [4]. Pada *information retrieval*, *query reformulation* yang sering dipakai adalah dengan *query expansion*, yaitu dengan memanjangkan *query* yang dimasukkan oleh *user* dengan menambahkan beberapa *term* kedalamnya. *Query expansion* merupakan salah satu teknik yang dapat digunakan dalam membantu pengguna memberikan *query* yang baik [5]. Metode *query expansion* pada penelitian ini adalah dengan cara menemukan padanan kata atau sinonim dari *query* pada kamus Tesaurus Bahasa Indonesia, kemudian sinonim tersebut ditambahkan ke *query* untuk melengkapkan informasi yang ingin dicari oleh pengguna.

Pada penelitian sebelumnya Mahdarani Dwi Laxmi [6] menggunakan metode pembobotan *tf-idf*, *cosine similarity* dan *wordnet* berbahasa Inggris untuk perluasan kueri pada sistem temu kembali berbahasa Indonesia, dengan akurasi hasil pencarian sebesar 70%. Data yang digunakan didapatkan dari media sosial Twitter. Data berupa *tweets* yang ada dalam situs tersebut. Pada penelitian tersebut masih memiliki kekurangan yaitu penggunaan *WordNet* berbahasa Inggris kurang efektif saat dataset berbahasa Indonesia karena perlu melakukan proses *translate*, sehingga mempengaruhi proses ekspansi kueri dan akurasi dari sistem yang dibangun.

Pada penelitian ini dibuatlah sistem temu kembali informasi pariwisata dengan model *vector space model* dan *automatic query expansion* menggunakan kamus Tesaurus Bahasa Indonesia. Diharapkan penelitian ini dapat menghasilkan suatu perangkat lunak yang dapat digunakan untuk mencari dan menampilkan informasi yang lebih relevan berdasarkan kata kunci yang dimasukkan pengguna,

dengan didasarkan pada kesamaan makna kata kunci tersebut. Dengan pemodelan ruang vektor atau *vector space model* yang digunakan untuk menyimpan dokumen dalam bentuk matriks *term-document*, diharapkan dapat meningkatkan akurasi sistem dalam menemukan kembali dokumen yang relevan.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Query yang dimasukkan oleh pengguna saat melakukan pencarian pada umumnya pendek, terdiri dari satu atau dua kata saja. Hal ini mengakibatkan dokumen relevan yang terpanggil hanya sedikit ataupun tidak terpanggil sama sekali, karena sedikitnya informasi dalam *query* yang dimasukkan pengguna.

1.2.2 Metode Usulan

Pada penelitian diusulkan pemodelan *vector space model* yang dipadukan dengan *automatic query expansion* menggunakan kamus Tesaurus Bahasa Indonesia. *Query expansion* digunakan untuk memperluas makna dari *query* agar dapat memanggil lebih banyak dokumen relevan. Sedangkan pemodelan *vector space model* digunakan untuk mempermudah dalam menemukan dokumen relevan.

1.2.3 Pertanyaan Penelitian

Berdasarkan latar belakang yang telah dijelaskan diatas, pertanyaan penelitian dalam pembuatan tugas akhir ini adalah:

1. Apakah penggunaan *query expansion* dapat meningkatkan dokumen relevan yang terpanggil?
2. Berapa bobot *f-measure* dari sistem temu kembali informasi pariwisata yang dibangun dengan pemodelan *vector space model* dan *automatic query expansion* menggunakan kamus Tesaurus Bahasa Indonesia?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Penerapan pemodelan *vector space model* dan *automatic query expansion* menggunakan kamus Tesaurus Bahasa Indonesia bertujuan agar mesin pencari yang dibangun dapat memberikan lebih banyak dokumen relevan dan sesuai dengan keinginan pengguna.

1.3.2 Manfaat

Sistem temu kembali informasi pariwisata yang dibangun dapat mencari tingkat relevansi dokumen dengan kata kunci yang telah diperluas maknanya, sehingga dapat melengkapi informasi yang dibutuhkan oleh pengguna dan dapat memberikan lebih banyak dokumen yang relevan.

1.4 Batasan Masalah

Adapun batasan masalah dari tugas akhir ini adalah:

1. Dokumen yang digunakan dalam penelitian ini merupakan halaman *web* dari <https://www.twisata.com/>, <https://www.nativeindonesia.com/>, <https://pesona.travel/>, <https://visitingjogja.com/>, <https://eksotisjogja.com/>, <https://anekatempatwisata.com/>, <https://www.water-sport-bali.com> dan <https://mytrip123.com/>.
2. Dokumen yang diproses adalah dokumen teks berbahasa Indonesia.
3. Data yang diproses merupakan isi dari *meta tag title* dan *meta tag description* halaman *web*.
4. Aplikasi ini tidak memperhatikan kesalahan penulisan pada dokumen *web*.
5. Situs yang digunakan merupakan situs dengan tema pariwisata atau berisi informasi tentang pariwisata.
6. Tidak ada batasan tahun pada dokumen *web* yang di *crawling*.
7. Kamus tesaurus yang digunakan merupakan kamus Tesaurus Bahasa Indonesia terbitan Pusat Bahasa Departemen Pendidikan Nasional.
8. Dokumen *web* yang digunakan pada penelitian berjumlah 532 dokumen dan telah di klasifikasikan secara manual oleh peneliti.

1.5 Sistematika Penulisan

Sistematika penulisan proposal ini terdiri dari beberapa bab, yaitu:

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini membahas tentang landasan teori dan penelitian sebelumnya.

BAB III METODE USULAN

Pada bab ini membahas tentang metode yang diusulkan (algoritma), arsitektur (Input, Proses, Output), tahapan penelitian, serta rencana pengujian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menguraikan tentang hasil, berupa kode program aplikasi beserta pengujian aplikasi tersebut.

BAB V PENUTUP

Bab ini berisikan kesimpulan dari hasil aplikasi yang dibuat dan saran-saran terhadap kekurangan dari aplikasi tersebut.