

BAB I

PENDAHULUAN

Pada bab ini menjelaskan hal yang melatar belakangi pentingnya dilakukan penelitian, rumusan masalah, dan batasan permasalahan yang dilakukan dalam penelitian, tujuan dan relevansi atau manfaat penelitian terhadap perkembangan solusi dari permasalahan yang diangkat, serta sistematika penulisan dalam laporan tugas akhir ini.

1.1 Latar Belakang

Jumlah dokumen web berita di situs web universitas semakin banyak dan *up-to-date*. Kebutuhan pengunjung akan informasi tersebut semakin tinggi. Selama ini pengguna atau pembaca web universitas masih kesulitan mencari berita yang mereka butuhkan, karena berita yang ada masih menyebar dan belum terklasifikasi sesuai dengan topik dan isi beritanya, sedangkan di sisi lain pengguna membutuhkan penemuan informasi yang cepat dan akurat. Untuk itu diperlukan suatu aplikasi pencarian yang mampu mengatasi permasalahan tersebut.

Aplikasi web dalam penelitian ini akan memudahkan dalam mencari informasi berita berdasarkan kelas beritanya dilihat dari kesesuaian isi beritanya. Dokumen/ berita akan diklasifikasikan berdasarkan topik beritanya, misal dokumen yang berisi beasiswa masuk pada klasifikasi berita beasiswa, dan sebagainya, dimana kelas-kelas berita sudah ditentukan sebelumnya (*predefined*). Sedangkan proses pengukuran tingkat similaritas antar dokumen dilakukan dengan membandingkan suatu *query* dengan dokumen. Agar hasil pengukuran tingkat similaritas dokumen dengan *query* mendapatkan hasil yang optimal maka digunakan algoritma TF-IDF (*Term Frequency – Inversed Document Frequency*) dan VSM (*Vector-Space Model*) dari IR (*Information Retrieval*) model untuk mencari nilai *cosine* (menghitung nilai *cosinus* sudut antara dua *vector*) sebagai pengukur tingkat kemiripan antara dokumen dengan *query*. Selain itu agar pencarian dokumen mudah ditemukan, maka dokumen akan diklasifikasikan berdasarkan kelas-kelas yang sudah ditentukan sebelumnya (*predefined*) menggunakan metode k-NN (*k-nearest neighbour*). Untuk mempercepat hasil pencarian kata kunci maka pada

aplikasi ini akan digunakan teknik *caching* terhadap dokumen-dokumen yang pernah dikembalikan sebelumnya kepada pengguna. Teknik ini dimaksudkan untuk meningkatkan transfer dokumen dengan menyimpan dokumen yang pernah diakses pada *cache* tersebut, sehingga apabila ada dokumen yang ingin diakses adalah dokumen yang sama maka akses akan dapat dilakukan lebih cepat.

Aplikasi ini diharapkan mampu mengklasifikasikan berita berdasarkan dokumen yang sudah terangkum dengan baik berdasarkan kebutuhan informasi pengguna, serta mampu mengembalikan hasil pencarian yang paling sesuai dengan *query* pengguna, serta proses perhitungan kemiripan dokumen lebih cepat.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, maka secara spesifik permasalahan yang dapat dirumuskan adalah sebagai berikut:

1. Bagaimana mengklasifikasikan dokumen sesuai dengan kandungannya (topik berita)?
2. Bagaimana meningkatkan presisi dari hasil pemrosesan *query* dengan melibatkan klasifikasi berita?
3. Bagaimana agar daftar dokumen yang relevan dengan *query* pengguna tersampaikan lebih cepat?

1.3 Batasan Masalah

Batasan masalah pada Tugas akhir ini meliputi:

1. Klasifikasi dokumen menggunakan metode k-NN (*k-Nearest Neighbor*.)

2. Dokumen berita yang digunakan hanya terbatas pada 10 universitas peringkat teratas berdasarkan <http://www.webometrics.info/en/asia/indonesia>.
3. Pengumpulan data dilakukan secara manual dngan mengambil 220 berita dari 10 universitas peringkat teratas berdasar <http://www.webometrics.info/en/asia/indonesia>.
4. Metode yang digunakan untuk menghitung kemiripan *query* dengan dokumen adalah *cosine similarity*.

1.4 Tujuan Penelitian

Penelitian ini akan membangun aplikasi web, maka sistem diharapkan mampu:

1. Mengklasifikasikan dokumen sehingga mudah diakses oleh pengguna.
2. Menyediakan daftar dokumen teranking yang relevan dengan *query* pengguna.
3. Mengembalikan hasil pencarian dengan cepat dan relevan terhadap *query*.

1.5 Metode Penelitian

1. Studi literatur. Literature yang digunakan mencakup buku teks, jurnal ilmiah, artikel dan sumber lain baik dalam bentuk *hard-copy* maupun *online* yang terkait dengan *cosine similarity*, *k-NN*, *informasi retrieval*, perhitungan kemiripan, dan klasifikasi dokumen.
2. *Survey website* universitas yang nantinya akan digunakan sebagai acuan dalam implementasi penelitian ini.
3. Berdasarkan studi literatur dan survey yang telah dilakukan sebelumnya, maka dapat memulai menganalisis dan merancang aplikasi web.

4. Implementasi dan pengujian sistem, berdasarkan analisis dan rancangan yang telah dilakukan maka dapat dimulai *coding*, dan pengujian sistem.
5. Penarikan kesimpulan, hasilnya dianalisis untuk menarik kesimpulan mengenai fungsionalitas dan kinerja mesin pencari yang telah diimplementasikan.

1.6 Sistematika Penulisan

Sistematika pembahasan pada laporan Tugas Akhir ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Bab ini terdiri dari latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan dari tugas akhir ini.

BAB II : DASAR TEORI

Pada bab ini dijelaskan mengenai hal-hal yang melandasi atau yang digunakan sebagai pedoman untuk menyelesaikan permasalahan yang terdapat dalam Tugas Akhir ini.

BAB III : ANALISIS DAN PERANCANGAN SISTEM

Pada bab ini dijelaskan mengenai analisis permasalahan, solusi, cara kerja dan rancangan dari sistem aplikasi untuk menyelesaikan permasalahan.

BAB IV : IMPLEMENTASI DAN PENGUJIAN

Pada bab ini diuraikan tentang implementasi dan pengujian terhadap aplikasi prototipe dikembangkan meliputi lingkungan implementasi, fungsi-fungsi yang berjalan, user interface utama yang disediakan, uji cpba dan analisis terhadapnya.

BAB V : PENUTUP

Bab ini berisi kesimpulan dan saran yang diperoleh setelah dilakukannya analisis terhadap hasil ujicoba.