

1. Pendahuluan

1.1 Latar belakang

Kemajuan yang pesat di bidang teknologi informasi khususnya Internet telah menimbulkan ledakan informasi. Setiap hari jutaan orang di dunia mencari, menciptakan, maupun menyebarkan informasi baru melalui Internet. Akibatnya, informasi tentang apapun tersedia melimpah di Internet. Dalam proses penelusuran informasi di Internet sering diperoleh informasi yang sangat banyak, tetapi sebagian besar diantaranya adalah informasi yang kurang begitu relevan.

Proses mencari dan mendapatkan informasi merupakan hal yang berkaitan erat dengan dunia informasi. Ketepatan hasil pencarian dan waktu pencarian menjadi beberapa faktor yang diperhatikan. Dokumen berbasis teks menjadi salah satu bentuk penyimpanan informasi. Biasanya pencarian dilakukan pada koleksi dokumen yang menjadi objek pencarian. Pencarian secara manual dapat dilakukan dengan membaca setiap dokumen pada koleksi dokumen untuk mendapatkan dokumen yang tepat dan sesuai kebutuhan. Namun, pencarian seperti itu membutuhkan waktu yang lama jika jumlah dokumen sangat banyak. Dengan demikian dibutuhkan sistem temu kembali informasi (*information retrieval system*) yang dapat mencari informasi yang sesuai dengan kebutuhan.

Sistem temu kembali informasi (*Information Retrieval*) merupakan sistem yang digunakan untuk menemukan kembali dokumen-dokumen yang sesuai dengan *keyword* (*query*) yang diberikan oleh *user* pada suatu koleksi dokumen sehingga diperoleh informasi yang relevan sesuai dengan kebutuhan *user*. Sedangkan *query* dalam *Information Retrieval* merupakan *keyword* (kata kunci) berupa satu atau beberapa term yang akan dicari. Secara umum sistem temu kembali terdiri dari beberapa komponen yaitu koleksi dokumen yang menjadi objek pencarian, *indexing*, informasi yang dibutuhkan *user* atau *query*, strategi pencarian (*searching* atau *matching*), hasil dokumen yang relevan dan penilaian relevansi (*relevance feedback*).

Ada dua faktor utama yang perlu diperhatikan dalam membangun suatu *information retrieval system*. Kedua faktor utama tersebut, yaitu efektifitas (*effectiveness*) dan efisiensi (*efficiency*). Faktor efektifitas (*effectiveness*) tidak hanya menyangkut akurasi hasil pencarian, tapi juga menyangkut akurasi pembobotan dokumen hasil pencarian. Sistem *information retrieval* yang baik adalah sistem *information retrieval* yang mampu mengambil semua dokumen relevan dan kemudian mengurutkan dokumen relevan tersebut pada urutan teratas. Agar diperoleh sistem dengan dokumen relevan berada pada urutan teratas dibutuhkan suatu metode pembobotan yang efektif dalam menentukan bobot suatu dokumen. Untuk alasan akurasi pembobotan inilah, digunakanlah Metode Savoy.

Metode savoy merupakan salah satu teknik pembobotan *term* dalam suatu dokumen. Pada metode Savoy ini, dalam menentukan bobot suatu *term* tidak hanya berdasarkan frekuensi kemunculan term di satu dokumen, tetapi juga memperhatikan frekuensi terbesar pada suatu term yang dimiliki oleh dokumen bersangkutan. Dalam menganalisa hasil penerapan dari metode savoy akan digunakan metode *term weighting* yang lain yaitu TF dan IDF sebagai pembanding dimana setiap tahap dalam proses *indexing* (meliputi *parsing*,

tokenization, stopwords removal, stemming) dan juga proses *searching* dilakukan dengan proses yang sama untuk setiap metode pembobotan, sehingga *precision, recall dan IAP* yang dihasilkan dipengaruhi oleh metode pembobotannya..

Metoda Savoy yang diterapkan pada *information retrieval* juga digunakan untuk menghasilkan nilai relevansi suatu dokumen terhadap *query* atau yang dikenal dengan *similarity score*. Tiap dokumen hasil pencarian mempunyai *similarity score* yang berbeda. *Similarity score* ini nantinya digunakan sebagai acuan dalam proses pengurutan dokumen

1.2 Perumusan masalah

Berdasarkan uraian diatas, maka permasalahan yang muncul dan yang menjadi objek penelitian pada Tugas Akhir ini ialah:

1. Bagaimana menerapkan *term weighting* dengan metode Savoy dalam proses *indexing* pada *information retrieval*.
2. Bagaimana pengaruh penerapan *term weighting* dengan metode Savoy pada pencarian informasi yang relevan dengan keinginan *user*.
3. Bagaimana menganalisis performansi dari penerapan *term weighting* dengan metode Savoy dalam proses *indexing* pada *information retrieval*. Performansi diukur berdasarkan *Precision, Recall, dan IAP*. Dalam menganalisis performansi akan digunakan teknik *term weighting TF* dan *IDF* sebagai pembandingan.

Batasan masalah dari Tugas Akhir ini adalah:

1. Koleksi dokumen dan *query* yang digunakan adalah dokumen dalam teks bahasa Inggris dan berformat .txt.
2. Koleksi dokumen menggunakan koleksi dokumen yang diambil dari <ftp://ftp.cs.cornell.edu/SMART> dan pada koleksi dokumen tersebut sudah terdapat kumpulan *query* beserta *relevance judgement* untuk setiap *query* tersebut. Koleksi dokumen yang digunakan yakni: ADI, CRAN, MED, dan CISI.

1.3 Tujuan

Tujuan pembuatan Tugas Akhir ini adalah :

1. Menerapkan *term weighting* dengan metode Savoy dalam proses *indexing* pada *information retrieval*.
2. Merancang dan membangun perangkat lunak untuk menerapkan *term weighting* dengan metode Savoy dalam proses *indexing* pada *information retrieval*.
3. Menganalisis performansi *precision, recall, dan IAP* terhadap hasil dari penerapan *term weighting* dengan metode Savoy dalam proses *indexing* pada *information retrieval*.

1.4 Metodologi penyelesaian masalah

Metodologi penyelesaian masalah yang digunakan dalam menyelesaikan Tugas Akhir ini, yaitu :

a) Studi Literatur

Langkah ini bertujuan untuk memahami dasar teori mengenai *information retrieval, indexing, dan metode Savoy* serta hal lain yang mendukung

penyelesaian tugas akhir ini. Sumber dasar teori dapat berupa buku, *paper*, maupun halaman web.

- b) Pengumpulan dan Pemahaman Data
Yaitu mengumpulkan dokumen teks yang diperlukan, mempelajari data.
- c) Analisis Kebutuhan dan Perancangan perangkat lunak
Melakukan analisis kebutuhan perangkat lunak yang akan dibangun, agar didapatkan gambaran umum seperti apa perangkat lunak yang ingin dibangun, kemudian merancang perangkat lunak yang sesuai untuk memenuhi kebutuhan. Rancangan perangkat lunak dapat menjadi panduan saat implementasi perangkat lunak.
- d) Implementasi
Pada tahapan ini dilakukan pembangunan perangkat lunak yang telah dirancang dengan menggunakan teknik pemrograman tertentu. Pada tahapan ini dibangun sistem yang dapat menangani proses *indexing* terhadap data dokumen yang menjadi objek pencarian, *term weighting* dengan metode Savoy.
- e) Pengujian dan Analisis hasil Implementasi
Di tahap ini, akan dilakukan pengujian terhadap hasil implementasi guna menemukan dan menghilangkan *error/bug* yang mungkin masih ada. Selain itu, juga dilakukan pengukur performansi berdasarkan nilai *precision*, *recall*, dan *Interpolated Average Precision* (IAP) dari sistem *information retrieval* yang dibangun. Analisis yang dilakukan pada seluruh dataset dan juga pada jumlah dokumen yang berbeda.
- f) Pembuatan Laporan Tugas Akhir
Pada tahap akhir, akan dilakukan pembuatan dokumentasi yang berupa laporan Tugas Akhir.

1.5 Sistematika Penulisan

Sistematika Penulisan Tugas Akhir ini terdiri dari lima bab dengan disertai lampiran terkait pelaksanaan tugas akhir yaitu:

- | | |
|---------|--|
| BAB I | Pendahuluan
Bab ini membahas kerangka penelitian dalam tugas akhir, meliputi latar belakang, perumusan masalah, batasan masalah, tujuan perancangan dan metodologi yang digunakan dalam perancangan sistem. |
| BAB II | Dasar Teori
Bab ini menjelaskan seluruh teori yang menjadi landasan konseptual dan mendukung penyelesaian tugas akhir ini. |
| BAB III | Analisis dan Perancangan Sistem
Bab ini membahas mengenai pengumpulan data analisis dan perancangan perangkat lunak yang terdiri dari perancangan struktur data, perancangan modul. |
| BAB IV | Implementasi, Pengujian dan Analisis
Bab ini membahas implementasi detail sistem dan pengujian terhadap sistem. |
| BAB V | Kesimpulan dan Saran
Berisi tentang kesimpulan dan saran sebagai hasil dari analisis dan implementasi Tugas Akhir. |