

BAB I

PENDAHULUAN

1.1 Latar Belakang

Belakangan ini terdapat beragam sistem informasi berbasis teks, yaitu informasi yang disimpan dalam dokumen-dokumen berupa file text. Karena banyaknya dokumen yang dapat disimpan, maka pengguna sistem informasi mengalami kesulitan untuk mendapatkan informasi yang diinginkan dan pengguna tidak dapat melihat dokumen satu demi satu untuk mendapatkan informasi yang tepat. Sehingga diperlukan suatu cara agar pengguna dapat mengakses informasi secara cepat dan tepat.

Hal ini diimbangi dengan meningkatnya penulisan ilmiah oleh mahasiswa di Jurusan Matematika Universitas Diponegoro. Sebuah penulisan ilmiah dapat ditemukan dengan mudah oleh manusia, tetapi jika dilakukan secara terkomputerisasi akan membawa permasalahan tersendiri. Begitu pula dengan tingkat relevansi atau kecocokan suatu dokumen dengan dokumen lainnya. Manusia dapat dengan mudah menentukan tingkat relevansi suatu dokumen dengan dokumen lainnya atau tidak. Salah satu cara adalah dengan menggunakan *query* (kata kunci). Dengan mengetikkan kata kunci sistem informasi akan menampilkan dokumen-dokumen tersebut, biasanya terurut menurut tingkat relevansinya. Hal ini disebut Sistem Temu Kembali Informasi (*information retrieval*).

Sistem Temu Kembali Informasi merupakan sistem yang berfungsi untuk menemukan informasi yang relevan dengan kebutuhan pemakai. Salah satu hal

yang perlu diingat adalah bahwa informasi yang diproses terkandung dalam sebuah dokumen yang bersifat tekstual. Dalam konteks ini, temu kembali informasi berkaitan dengan representasi, penyimpanan, dan akses terhadap dokumen representasi dokumen. Dokumen yang ditemukan tidak dapat dipastikan apakah relevan dengan kebutuhan informasi pengguna yang dinyatakan dalam *query* (Tarto, 2008).

Salah satu model sistem temu kembali informasi yang paling sederhana namun produktif adalah model ruang vektor. Vektor model ini mempresentasikan term yang terdapat pada dokumen dan query. Elemen vektor tersebut adalah bobot *term* yang menjadi penilaian dan perankingan dokumen. Dalam model ruang vektor ini hal yang perlu diperhatikan adalah pembobotan *term* (*term weighting*) (Arifin, 2002).

Metode pembobotan yang umumnya diunggulkan dalam beberapa penelitian menggunakan model ruang vektor yaitu *Term Frequency Inverse Document Frequency* TF-IDF (Arifin, 2002). Dalam perhitungan bobot *term*, sekalipun *term frequency* banyak digunakan, namun hal itu hanya mendukung proporsi jumlah dokumen yang dapat ditemukan kembali oleh proses pencarian pada sistem temu kembali informasi. Sedangkan proporsi jumlah dokumen yang ditemukan dan dianggap relevan untuk kebutuhan pengguna akan lebih meningkat bila vektor bobot tersebut menggunakan term yang jarang muncul pada koleksi dokumen. *Term* demikian diharapkan mampu mengelompokkan sejumlah dokumen yang memuatnya, sehingga berbeda dengan seluruh anggota koleksi dokumen lain yang tidak memilikinya. Kriteria ini dapat diakomodasi dengan menghitung invers frekuensi dokumen. Dengan digabungkannya kedua metode ini yaitu frekuensi

kemunculan *term* dan invers frekuensi yang mengandung kata tersebut, diharapkan mampu meningkatkan proporsi jumlah dokumen yang dapat ditemukan kembali dan yang dianggap relevan secara sekaligus. Sehingga kriteria *term* yang paling tepat adalah *term* yang sering muncul dalam dokumen secara individu, namun jarang dijumpai pada dokumen lainnya.

Dari penjelasan di atas maka dalam penulisan Tugas Akhir ini akan dibuat suatu perangkat lunak yang dapat mencari dokumen-dokumen penulisan ilmiah dari mahasiswa jurusan Matematika Universitas Diponegoro yang relevan sesuai tingkat pembobotannya.

1.2 Perumusan Masalah

Berdasarkan penjelasan pada latar belakang yang sudah dijelaskan dapat dirumuskan masalah dalam Tugas Akhir ini adalah bagaimana mengimplementasikan metode pembobotan TF-IDF pada perangkat lunak untuk mendapatkan hasil dokumen relevan yang teranking sesuai tingkat pembobotannya berdasarkan query masukan oleh pengguna.

1.3 Ruang Lingkup Masalah

Ruang lingkup penulisan Tugas Akhir ini dibatasi pada:

1. Aplikasi ini mencari dokumen pendek berupa abstrak tugas akhir berbahasa Indonesia dari mahasiswa matematika Universitas Diponegoro yang berformat teks (.txt).
2. Query inputan yang digunakan adalah berupa kata, bukan simbol atau notasi tertentu.
3. Aplikasi ini dibangun menggunakan perangkat lunak Borland Delphi 6.0.

1.4 Tujuan dan Manfaat

Tujuan umum dari penulisan Tugas Akhir ini adalah membangun perangkat lunak sistem temu kembali informasi untuk pencarian dokumen-dokumen teks yang berupa abstrak Tugas Akhir agar pengguna mendapatkan dokumen yang relevan berdasarkan *query* yang dimasukkan, serta telah teranking sesuai tingkat pembobotannya.

Manfaat dari penulisan tugas akhir ini adalah memudahkan pengguna mendapatkan dokumen relevan berdasarkan *query* yang dimasukkan, serta telah teranking berdasarkan tingkat pembobotannya.

1.5 Sistematika Penulisan

BAB 1 : PENDAHULUAN

Bab 1 berisikan tentang latar belakang pengambilan Tugas Akhir yang berjudul “Sistem Temu Kembali untuk Dokumen Matematika dengan Menggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF)”, perumusan masalah, ruang lingkup pembatasan masalah, tujuan serta manfaat dari penulisan, sistematika penulisan Tugas Akhir.

BAB 2 : LANDASAN TEORI

Bab 2 membahas tentang pengertian Sistem Informasi Temu Kembali secara umum, penjelasan tentang model ruang vektor Sistem Temu Kembali Informasi, penjelasan metode pembobotan TF-IDF, *Data Flow Diagram* (DFD), serta performansi sistem.

BAB 3 : ANALISIS DAN PERANCANGAN SISTEM

Bab 3 berisikan analisis sistem yang berupa penjelasan tentang model fungsional perangkat lunak yang meliputi *Data Context Diagram* (DCD), *Data Flow Diagram* (DFD). Perancangan sistem berupa penjelasan tentang perancangan fungsi, dan perancangan antarmuka perangkat lunak.

BAB 4 : IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab 4 berisikan tentang penjelasan implementasi perancangan perangkat lunak yang sudah dijelaskan pada bab sebelumnya serta pengujian terhadap perangkat lunak.

BAB 5 : KESIMPULAN DAN SARAN

Bab 5 berisikan tentang kesimpulan yang dapat diambil dalam penulisan Tugas Akhir serta saran-saran membangun yang dapat dikembangkan untuk pengembangan sistem di masa datang.