

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi yang semakin pesat telah memaksa manusia untuk berusaha mengikutinya. Teknologi tersebut dapat digunakan oleh semua kalangan yang dapat memanfaatkannya untuk berbagai keperluan. Teknologi tersebut memudahkan mereka dalam memenuhi kebutuhan dengan lebih cepat, lebih efisien, dan tepat sehingga waktu yang akan dipergunakan dapat semakin dipangkas. Hal tersebut juga diiringi dengan semakin banyaknya informasi yang tersebar secara luas di dunia maya.

Jumlah halaman yang tersedia di Internet hampir dua kali lipat setiap tahunnya (Radev *et al*, 2000). Pada bulan Juli 2011, jumlah host yang diiklankan di DNS mencapai angka 849.869.781 (ISC, diakses tanggal 28 Agustus 2013). Oleh karena itu mesin pencarian memberikan banyak referensi media online dari sebuah kata kunci yang di berikan oleh pengguna. Dari berbagai media online tersebut seringkali berita yang disajikan memiliki topik yang sama, namun berbeda dalam penyajian beritanya. Hal ini dapat membingungkan pengguna ketika ingin mendapatkan informasi yang relevan terkait suatu topik tertentu karena banyaknya sumber berita yang beredar. Pengguna tentunya akan kesulitan demi mendapat informasi yang relevan jika harus membaca satu per satu dari sekian banyaknya informasi yang tersebar online. Berdasarkan masalah tersebut maka dibutuhkan

suatu aplikasi yang menyediakan ringkasan dari banyak dokumen yang tersebar di berbagai media online tersebut.

Tesis ini memberikan pendekatan ringkasan teks sebagai solusi untuk masalah ini. Pendekatan ini mengurangi waktu yang diperlukan untuk menemukan dokumen web yang memiliki informasi yang relevan. Ringkasan dapat didefinisikan sebagai teks yang dihasilkan dari satu atau lebih teks, yang menyampaikan informasi penting dalam teks asli dengan rasio kompresi yang cukup dan secara signifikan berisi teks lebih sedikit dari sumber (Radev *et al*, 2002).

Ringkasan dokumen dari banyak sumber menyediakan ringkasan dari informasi yang terdapat di berbagai media online. Ringkasan dokumen dari banyak sumber adalah proses penyaringan informasi penting dari satu set dokumen untuk menghasilkan informasi yang lebih relevan untuk pengguna. Hal ini dapat dilihat sebagai perpanjangan ringkasan dokumen tunggal yang hanya menyediakan ringkasan dari satu dokumen saja. Namun, hal-hal seperti redundansi, cakupan, keterkaitan, rasio kompresi, dll, lebih menonjol dalam ringkasan dokumen dari banyak sumber dan membuat aplikasi ini lebih kompleks. Dokumen sumber yang digunakan dalam ringkasan ini di dapat dari semua media online seperti koran online, e-book dan artikel-artikel yang ada di dunia maya.

Aplikasi peringkasan dokumen dari banyak sumber menggunakan pengelompokan kalimat dengan cara *sentence scoring* menggunakan metode TF-IDF. Aplikasi ini membuat ringkasan dari masing-masing dokumen yang ada dengan cara memberikan nilai pada tiap-tiap kalimat menggunakan metode TF-IDF. Metode TF-IDF merupakan sebuah metode pembobotan kata dengan cara

menghitung frekuensi kata tersebut di dalam sebuah dokumen dan di dalam kumpulan dokumen (Ramos, 2004). Metode ini dapat dipergunakan untuk memberikan bobot masing-masing kata dan kemudian menghitung skor dari kalimat di dalam sebuah dokumen. Kemudian hasil dari pemberian skor pada kalimat tersebut dipilih beberapa kalimat dengan skor tertinggi dan kalimat tersebut digabungkan untuk menjadi ringkasan akhir. Dengan adanya aplikasi ini diharapkan mampu memberikan ringkasan yang relevan dari informasi-informasi yang beredar di media online bagi pengguna.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas dapat dirumuskan masalah sebagai berikut :

1. Bagaimana mengelompokkan kalimat dari beberapa dokumen menggunakan *sentence scoring* dengan metode TF-IDF?
2. Bagaimana membangun perangkat lunak untuk membuat ringkasan yang bersumber dari satu atau lebih dokumen berbasis web?

1.3 Tujuan

Tujuan dari pembangunan perangkat lunak peringkas dokumen dari banyak sumber berbasis web menggunakan *sentence scoring* dengan metode TF-IDF ini adalah :

1. Mengelompokkan kalimat dari beberapa dokumen menggunakan *sentence scoring* dengan metode TF-IDF.

2. Membangun perangkat lunak untuk membuat ringkasan yang bersumber dari satu atau lebih dokumen berbasis web.

1.4 Batasan Masalah

Batasan-batasan masalah yang terdapat dalam pembangunan perangkat lunak peringkas dokumen dari banyak sumber berbasis web menggunakan *sentence scoring* dengan metode TF-IDF ini adalah :

1. Ringkasan akhir yang di hasilkan berupa ringkasan ekstraktif.
2. Dokumen yang diringkas adalah artikel dari media online.
3. Dokumen yang diringkas adalah dokumen topik yang sama atau mirip.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini dalam bentuk teoritis adalah :

1. Penelitian ini dapat menambah pemahaman teori penggunaan metode TF-IDF pada pembangunan perangkat lunak peringkas dokumen.
2. Penelitian ini dapat menambah pemahaman teori dan menjadi bahan pertimbangan sejauh mana metode TF-IDF dapat digunakan dalam sebuah sistem peringkas dokumen yang dapat digunakan untuk *sentence scoring* dalam pemilihan kalimat.

1.6 Metodologi

Metode yang akan digunakan dalam pembangunan perangkat lunak peringkasan dokumen dari banyak sumber berbasis web menggunakan *sentence scoring* dengan metode TF-IDF adalah:

1. Metode Observasi

Metode observasi dilakukan dengan melakukan pengamatan dan pencatatan terkait dengan proses yang berhubungan dengan sistem yang akan dibangun.

2. Metode Kepustakaan

Dalam tahap ini, dilakukan pengumpulan data dari berbagai sumber yang mendukung seperti buku-buku referensi, skripsi, jurnal, serta data-data di internet yang terkait dengan pembangunan sistem peringkasan dokumen dari banyak sumber ini.

3. Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak mencakup proses spesifikasi kebutuhan perangkat lunak, seperti antarmuka dengan pengguna (*user interface*) maupun kinerja (*performance*) perangkat lunak pada berbagai fungsi yang dirancang untuk dapat dilaksanakan oleh sistem. Dalam metode ini ada 4 sub metode, yaitu:

- a. Analisis

Dalam tahap ini ditentukan spesifikasi kebutuhan sistem yang akan dibuat. Hasilnya berupa Spesifikasi Kebutuhan Perangkat Lunak (SKPL).

b. Perancangan

Dalam tahap ini dilakukan perancangan terhadap sistem yang akan dibuat berdasarkan spesifikasi yang telah ada. Hasilnya berupa Deskripsi Perancangan Perangkat Lunak (DPPL).

c. Pengkodean

Tahap pengkodean merupakan tahap implementasi rancangan sistem ke dalam program. Hasilnya berupa kode sumber yang siap dieksekusi.

d. Pengujian

Tahap pengujian merupakan tahap dimana sistem yang telah dibuat dinilai apakah secara fungsional sesuai dengan spesifikasi yang telah dibuat.

4. Metode Dokumentasi

Metode dokumentasi merupakan metode yang digunakan untuk mencatat data yang telah terkumpul untuk pembangunan perangkat lunak ke dalam bentuk dokumen.

1.7 Sistematika Penulisan Tugas Akhir

Adapun sistematika penulisan laporan tesis adalah sebagai berikut:

BAB 1 PENDAHULUAN

Pada bab ini akan dijelaskan mengenai latar belakang masalah, rumusan masalah, tujuan, batasan masalah, metode penelitian, dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada bab ini akan diuraikan secara singkat hasil-hasil penelitian yang pernah dilakukan yang masih memiliki relasi atau hubungan dengan permasalahan yang diangkat, serta mampu untuk mendukung dalam penyusunan Tesis ini dan teori-teori, pendapat, prinsip, dan sumber-sumber lain yang dapat dipertanggungjawabkan secara ilmiah.

BAB 3 METODOLOGI PENELITIAN

Pada bab ini akan dijelaskan mengenai metodologi penelitian yang akan digunakan dalam pembangunan perangkat lunak ini.

BAB 4 HASIL PENELITIAN DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai hasil analisis dan perancangan perangkat lunak yang akan dibuat, gambaran mengenai cara mengimplementasikan dan menggunakan sistem, serta hasil pengujian yang dilakukan terhadap perangkat lunak tersebut.

BAB 5 KESIMPULAN DAN SARAN

Pada bab ini akan dijelaskan mengenai kesimpulan yang didapat dari pembahasan laporan secara keseluruhan, serta kritik dan saran.

Pada bab selanjutnya akan dibahas *review* aplikasi-aplikasi yang sejenis dengan perangkat lunak peringkasan dokumen dari banyak sumber berbasis web menggunakan *sentence scoring* dengan metode TF-IDF, analisis penggunaan

metode TF-IDF, dan gambaran tentang perangkat lunak peringkasan dokumen dari banyak sumber berbasis web menggunakan *sentence scoring* dengan metode TF-IDF.

