

BAB 1

PENDAHULUAN

Pada bab ini akan dijelaskan beberapa hal mengenai latar belakang perlunya penelitian mengenai segmentasi dokumen beserta peranannya terhadap kemajuan teknologi informasi, perumusan masalah, tujuan penelitian, ruang lingkup penelitian, dan metodologi penelitian. Tujuan penelitian akan memberikan penjelasan mengenai hasil yang ingin diketahui setelah melakukan penelitian ini. Ruang lingkup penelitian akan menjelaskan mengenai batasan-batasan dalam proses pelaksanaan. Metodologi penelitian akan membahas langkah-langkah pelaksanaan tugas akhir ini. Bab ini pun akan menjelaskan mengenai sistematika penulisan laporan tugas akhir ini.

1.1 LATAR BELAKANG PENELITIAN

Dewasa ini, seiring dengan perkembangan teknologi dan meningkatnya kebutuhan informasi yang cepat dan tepat maka semakin berkembang pula berbagai penelitian-penelitian dibidang perolehan informasi. Beberapa contoh jenis penelitian yang dewasa ini semakin berkembang adalah *document summarization*, *document extraction*, *document classification*, *text-alignment*, *information retrieval*, dll. Minat penerapan topik-topik penelitian tersebut pada bahasa Indonesia pun turut berkembang seiring meningkatnya minat penelitian perolehan informasi untuk bahasa Indonesia. Perkembangan tersebut pun turut memacu berkembangnya jenis-jenis informasi yang tersedia, salah satu bentuknya adalah informasi dalam bentuk suara.

Salah satu contoh penelitian yang berkaitan dengan dokumen suara adalah penelitian tentang *topic detection and tracking* (TDT). TDT merupakan suatu penelitian yang berkaitan dengan pengolahan dokumen suara yang bertujuan untuk memonitor peristiwa-peristiwa yang terjadi di dunia melalui siaran berita. Pada penelitian tentang TDT ini, segmentasi dokumen merupakan suatu proses awal yang perlu dilakukan sebelum melakukan proses deteksi berita baru ataupun pencarian berita. Tujuan utama dari penelitian yang dilakukan penulis ini adalah

untuk mengembangkan sebuah sistem segmentasi dokumen bahasa Indonesia yang akurat. Dengan hasil segmentasi tersebut diharapkan dapat membantu menciptakan suatu sistem deteksi dan pelacakan berita suara bahasa Indonesia yang baik dan akurat.

1.2 PERUMUSAN MASALAH

Permasalahan yang melandasi dilakukannya penelitian ini adalah adanya kebutuhan untuk memproses dokumen-dokumen yang belum tersegmentasi secara jelas, sehingga dibutuhkan suatu sistem untuk melakukan segmentasi pada dokumen tersebut. Untuk mengatasi masalah tersebut, maka pada penelitian ini akan dilakukan pengujian pada salah satu metode baru yang dapat digunakan sebagai metode untuk melakukan segmentasi dokumen pada dokumen bahasa Indonesia, yaitu *genetic algorithm*. Diharapkan dengan metode yang cukup baru ini, hasil segmentasi yang diperoleh pun akan semakin baik.

1.3 TUJUAN PENELITIAN

Tujuan utama dari penulisan dan penelitian pada tugas akhir ini adalah untuk menguji coba metode *genetic algorithm* jika digunakan sebagai metode untuk melakukan segmentasi pada dokumen bahasa Indonesia. Hasil yang diharapkan dari proses ini adalah diperolehnya konfigurasi hasil segmentasi dari suatu dokumen yang belum tersegmentasi, sehingga batas-batas antar topik penyusunnya dapat diketahui dengan jelas.

1.4 RUANG LINGKUP PENELITIAN

Ruang lingkup penelitian pada tugas akhir ini meliputi:

- Segmentasi dokumen dengan menggunakan metode *genetic algorithm*. Kegiatan yang dilakukan meliputi implementasi sistem segmentasi dokumen, implementasi sistem untuk persiapan data-data percobaan, dan percobaan terhadap data-data yang telah disiapkan.
- Penelitian segmentasi dokumen ini menggunakan dua jenis data sebagai bahan percobaan, yaitu dokumen artikel dari media massa Indonesia

(www.kompas.com) dan dokumen abstrak tulisan ilmiah dari sistem Lontar yang dimiliki Fakultas Ilmu Komputer Universitas Indonesia.

- Perbandingan hasil segmentasi antara metode *genetic algorithm* dengan metode Texttiling.

1.5 METODOLOGI PENELITIAN

Metodologi penelitian yang diterapkan pada tugas akhir ini adalah:

1. Studi Literatur – Tahap pencarian informasi mengenai segmentasi dokumen, metode-metode yang sudah ada, metode-metode yang sedang berkembang, dan pembelajaran serta pemahaman terhadap metode-metode tersebut.
2. Perancangan – Tahap melakukan pembuatan rancangan percobaan, mempersiapkan data-data yang akan digunakan, penentuan variabel percobaan, dan perancangan sistem segmentasi dokumen dengan menggunakan metode *genetic algorithm*.
3. Implementasi – Tahap realisasi dari apa yang sudah dirancang pada tahap perancangan, dimulai dengan implementasi sistem segmentasi dokumen dengan metode *genetic algorithm* sampai dengan melakukan percobaan-percobaan yang telah dirancang.
4. Analisis Hasil – Tahap melakukan perbandingan hasil segmentasi dengan metode *genetic algorithm* yang ditinjau dari aspek *fitness function*, metode penghitungan *similarity*, jumlah iterasi, ukuran populasi, tipe *crossover*, probabilitas mutasi, jumlah segmen pada dokumen, dan kemiripan antar dokumen penyusun, lalu pada percobaan terakhir dilakukan percobaan untuk membandingkan hasil segmentasi menggunakan metode *genetic algorithm* dengan metode lain (Texttiling).

1.6 SISTEMATIKA PENULISAN

Laporan tugas akhir yang disusun oleh penulis ini terdiri atas enam bab, dengan rincian sebagai berikut:

- **BAB 1 PENDAHULUAN**

Pada bab ini akan dijelaskan beberapa hal mengenai latar belakang perlunya penelitian mengenai segmentasi dokumen beserta perannya terhadap kemajuan teknologi informasi, perumusan masalah, tujuan penelitian, ruang lingkup penelitian, dan metodologi penelitian. Tujuan penelitian akan memberikan penjelasan mengenai hasil yang ingin diketahui setelah melakukan penelitian ini. Ruang lingkup penelitian akan menjelaskan mengenai batasan-batasan dalam proses pelaksanaan. Bab ini pun akan menjelaskan mengenai sistematika penulisan laporan tugas akhir ini.

- **BAB 2 LANDASAN TEORI**

Pada bab ini akan dijelaskan mengenai landasan teori dan metode-metode yang digunakan pada penelitian mengenai segmentasi dokumen ini. Pembahasan akan dimulai dengan penjelasan mengenai segmentasi dokumen itu sendiri, penjelasan mengenai *topic detection and tracking* yang merupakan salah satu aplikasi yang menggunakan proses segmentasi dokumen dan penjelasan mengenai metode segmentasi Texttiling. Selanjutnya akan dijelaskan mengenai metode *genetic algorithm* (beserta hal-hal yang terkait) dan penelitian segmentasi dokumen menggunakan metode *genetic algorithm* yang pernah dilakukan.

- **BAB 3 PERANCANGAN**

Pada bab ini akan dijelaskan mengenai rancangan percobaan pada penelitian segmentasi dokumen ini. Pembahasan akan dimulai dengan penjelasan mengenai gambaran umum proses segmentasi yang dilakukan, lalu akan dijelaskan mengenai data yang digunakan pada percobaan beserta pengolahan yang dilakukan sebelum data tersebut dapat digunakan sebagai data percobaan. Selanjutnya akan dijelaskan mengenai metode-metode segmentasi dokumen yang digunakan pada percobaan ini beserta proses pengukuran hasil segmentasi.

- **BAB 4 IMPLEMENTASI**

Pada bab ini akan dijelaskan secara terperinci proses implementasi dari hasil perancangan penelitian segmentasi dokumen yang telah dijelaskan

Universitas Indonesia

pada Bab 3. Proses implementasi yang akan dijelaskan terdiri dari implementasi proses pengolahan data yang dilakukan sebelum data tersebut siap digunakan pada percobaan dan implementasi *genetic algorithm* sebagai metode yang digunakan pada penelitian segmentasi dokumen ini. Keseluruhan proses implementasi ini akan dilakukan dengan menggunakan bahasa pemrograman JAVA dan *library* ECJ 18 yang juga berbasiskan JAVA untuk implementasi *genetic algorithm*. Hasil yang diperoleh dari implementasi proses pengolahan data adalah dokumen-dokumen yang siap digunakan pada percobaan. Hasil yang diperoleh dari implementasi *genetic algorithm* adalah sistem yang dapat melakukan segmentasi dokumen secara otomatis dan hasil segmentasi dari masing-masing *test case* yang diujikan.

- **BAB 5 HASIL DAN PEMBAHASAN**

Pada bab ini akan dijelaskan mengenai hasil dari percobaan segmentasi dokumen yang telah dilakukan. Pembahasan yang dilakukan mencakup perbandingan hasil dari variabel eksperimentasi (*fitness function*, metode penghitungan *similarity*, *genetic operator*) dan variasi *test case*. Pada pembahasan terakhir akan dibahas perbandingan hasil segmentasi antara metode *genetic algorithm* (menggunakan konfigurasi *genetic operator* terbaik) dan metode Texttiling. Setelah membahas semua hasil percobaan maka akan dilakukan pembahasan mengenai analisa kesalahan dan rangkuman hasil percobaan.

- **BAB 6 PENUTUP**

Bab ini merupakan penutup dari laporan tugas akhir yang berisi kesimpulan, kendala-kendala yang ditemui selama menyelesaikan tugas akhir ini. Selain itu disertakan pula pada subbab terakhir saran-saran untuk pengembangan lebih lanjut dalam penelitian segmentasi dokumen dengan menggunakan *genetic algorithm* dikemudian hari.