

# **BAB 1    PENDAHULUAN**

Bab ini menguraikan latar belakang, rumusan masalah, tujuan, ruang lingkup dan tahapan yang dilalui dalam penelitian.

## **1.1    Latar Belakang**

Sejak zaman dahulu kala, umat manusia sudah menyadari terhadap pentingnya suatu informasi. Jika kita melihat kembali buku-buku yang menceritakan sejarah peradaban umat manusia, mungkin akan timbul pertanyaan bagaimana para penulis buku-buku yang kita baca itu dapat mengetahui sejarah peradaban manusia dari zaman dahulu kala. Para penulis buku-buku sejarah tersebut tentunya tidak mendapatkan informasi melalui pendapat atau pemikiran mereka semata, melainkan melalui informasi-informasi yang dibuat sendiri oleh umat manusia sejak zaman dahulu kala. Hal ini membuktikan bahwa sudah sejak lama manusia mengenal informasi, dan sudah melakukan pengolahan terhadap informasi dengan tujuan informasi tersebut dapat digunakan oleh keturunan-keturunan mereka. Tentu saja representasi informasi pada masa itu tentu saja berbeda dengan masa kini, dimana pada masa tersebut umat manusia mengenal informasi dalam bentuk tulisan di atas sebuah daun, pahatan batu, batang kayu, arca, dan yang lainnya. Saat ini kita mengenal informasi dalam bentuk tulisan di atas kertas, berita televisi, surat kabar, internet, video dll, dimana yang membedakan dengan informasi pada masa lampau hanyalah medianya saja.

Informasi merupakan sumber dari segala ilmu pengetahuan, baik ilmu pengetahuan yang bersifat sains ataupun ilmu pengetahuan yang sifatnya sosial. Dengan terus berkembangnya zaman dan kebudayaan, ilmu pengetahuan manusia juga ikut berkembang mengikuti tuntutan zaman, dimana kebutuhan manusia terhadap informasi semakin besar. Ilmu pengetahuan berkembang dengan pesatnya, media penyimpanan informasi seperti batu atau batang kayu tidak lagi dapat memenuhi standar kebutuhan penyimpanan atau pengolahan informasi mereka, sehingga manusia mulai mencari alternatif penyimpanan informasi dalam bentuk kertas dan digital.

Masalah lain yang timbul selain media penyimpanan informasi adalah akses untuk mendapatkan informasi tersebut, yaitu bagaimana untuk mencari informasi yang jumlahnya tidak terbatas. Kemudian jika informasi-informasi tersebut sudah kita peroleh, maka salah satu masalah yang timbul selanjutnya adalah bagaimana caranya mendapatkan inti dari informasi-informasi tersebut dengan menggunakan waktu yang sesedikit mungkin. Contohnya adalah seorang karyawan yang gemar membaca surat kabar digital di internet pada waktu senggangnya akan menginginkan suatu bentuk lain dari penyajian informasi pada surat kabar tersebut sehingga ia bisa dengan cepat mendapatkan inti berita/informasi yang dibutuhkannya.

Peringkasan dokumen merupakan salah satu solusi yang dapat ditawarkan terhadap masalah di atas. Peringkasan dokumen akan menghasilkan suatu bentuk penyajian lain dari dokumen yaitu ringkasan, yang berisi intisari dari topik yang dibicarakan pada dokumen. Jika peringkasan dokumen ini dilakukan oleh manusia secara manual tentunya waktu dan tenaga yang dibutuhkan tidak akan sebanding dengan dokumen yang jumlahnya tidak terbatas. Oleh karena itu dikembangkanlah suatu metode peringkasan dokumen secara otomatis menggunakan sistem komputer.

Penelitian mengenai metode peringkasan dokumen awalnya hanya dilakukan untuk menghasilkan ringkasan dari sebuah dokumen saja. Hal tersebut dikarenakan sumber informasi pada saat itu belum sebanyak saat ini, sehingga peringkasan pada sebuah dokumen menjadi kebutuhan pada saat itu. Namun seiring dengan berjalannya waktu, dimana sumber informasi menjadi sangat banyak jumlahnya, sebuah peristiwa dapat diberitakan oleh banyak sumber dokumen. Hal ini mengakibatkan perolehan informasi melibatkan banyak sumber informasi. Hal-hal tersebut memicu penelitian mengenai metode peringkasan dokumen untuk ikut melibatkan banyak dokumen, yang disebut dengan peringkasan multi-dokumen. Peringkasan multi-dokumen merupakan metode peringkasan yang bertujuan untuk menghasilkan ringkasan dari banyak dokumen, dimana ringkasan tersebut berisi seluruh informasi utama/penting dari dokumen-dokumen tersebut.

Penelitian peringkasan multi-dokumen telah banyak dilakukan pada *domain* dokumen-dokumen yang berbahasa Inggris. Namun sumber-sumber informasi kini tidak hanya ditulis dalam Bahasa Inggris saja. Sumber-sumber informasi juga banyak

yang ditulis dalam Bahasa Indonesia. Hal inilah yang memotivasi penulis untuk melakukan penelitian dalam bidang perolehan informasi dengan topik peringkasan multi-dokumen untuk dokumen berbahasa Indonesia.

## 1.2 Rumusan Masalah

Penelitian yang dilakukan penulis memiliki fokus terhadap masalah-masalah sebagai berikut:

1. Bagaimana implementasi metode peringkasan multi-dokumen untuk dokumen berbahasa Indonesia?
2. Bagaimana pengaruh penggunaan teknik peringkasan multi-dokumen yang berbeda terhadap kualitas dari hasil ringkasan?
3. Bagaimana pengaruh kualitas dari kelompok dokumen terhadap kualitas hasil ringkasan yang diperoleh?

## 1.3 Tujuan

Tujuan yang ingin dicapai melalui penelitian ini adalah:

1. Mengimplementasikan metode peringkasan multi-dokumen untuk dokumen berbahasa Indonesia.
2. Melakukan analisis terhadap kualitas hasil ringkasan yang dihasilkan dua teknik peringkasan multi-dokumen yaitu teknik *centroid-based summarization* dan *k-means-based summarization*.
3. Mengevaluasi pengaruh kualitas dan ukuran kelompok dokumen terhadap kualitas dari ringkasan yang dihasilkan.

## 1.4 Ruang Lingkup Penelitian

Penelitian dilakukan dengan menggunakan koleksi dokumen berbahasa Indonesia. Peringkasan dokumen dilakukan dengan menggunakan data masukan berupa kelompok dokumen. Peringkasan dokumen untuk dokumen berbahasa Indonesia ini dilakukan berdasarkan teknik-teknik peringkasan yang sudah pernah digunakan sebelumnya yaitu teknik *centroid-based summarization* dan *k-means-based summarization* dalam *domain* bahasa yang berbeda.

Evaluasi akan dilakukan dengan membandingkan teknik peringkasan *centroid-based summarization* dan teknik peringkasan *k-means-based summarization* terhadap kualitas dari ringkasan yang dihasilkan oleh kedua buah teknik.

## 1.5 Metodologi Penelitian

Pada penelitian tugas akhir ini penulis menggunakan metodologi penelitian eksperimental yang terbagi ke dalam beberapa tahapan sebagai berikut:

1. Studi literatur. Penulis mempelajari berbagai pustaka seperti jurnal, makalah penelitian, dan literatur lainnya yang membahas penelitian mengenai perolehan informasi khususnya topik peringkasan multi-dokumen. Pustaka yang dipelajari tersebut berasal dari berbagai penelitian mengenai peringkasan multi-dokumen yang pernah dilakukan oleh orang lain sebelumnya.
2. Eksperimen. Penulis menyiapkan skenario serta koleksi dokumen untuk melakukan eksperimen terhadap peringkasan multi-dokumen untuk dokumen berbahasa Indonesia. Penulis membuat data masukan yang berupa kelompok dokumen yang diperoleh dari artikel surat kabar *online* dan situs-situs berita di internet. Eksperimen dilakukan terhadap teknik peringkasan *centroid-based summarization* dan *k-means-based summarization* dengan menggunakan kelompok dokumen tersebut sebagai masukan.
3. Analisis. Penulis menganalisis hasil eksperimen terhadap penerapan metode yang digunakan, melakukan perbandingan antara hasil eksperimen, kemudian menarik kesimpulan.

## 1.6 Sistematika Penulisan

Laporan penelitian tugas akhir ini terdiri dari 5 bab, yaitu:

- Bab 1 Pendahuluan, membahas mengenai latar belakang, rumusan masalah, tujuan, ruang lingkup, metodologi penelitian yang dilakukan dan sistematika penulisan laporan penelitian.
- Bab 2 Landasan Teori, membahas mengenai teori-teori yang menjadi dasar dalam penelitian penulis.
- Bab 3 Eksperimen, membahas mengenai objek dari eksperimen, skenario eksperimen, dan hal-hal lain yang terkait dengan eksperimen dari penelitian ini.

- Bab 4 Hasil dan Analisis Eksperimen, bab ini membahas hasil dari eksperimen yang dilakukan pada bab 3 beserta analisa penulis terhadap hasil dari eksperimen tersebut.
- Bab 5 Penutup, bab ini berisi kesimpulan yang diperoleh dari penelitian ini dan saran-saran dari penulis untuk pengembangan terhadap topik penelitian ini selanjutnya.

