

BAB 5 PENUTUP

Bab ini membahas kesimpulan dari penelitian peringkasan multi-dokumen untuk dokumen berbahasa Indonesia yang dilakukan oleh penulis, serta saran-saran yang dapat digunakan untuk kelanjutan dari penelitian di bidang ini.

5.1 Kesimpulan

Pada penelitian yang dilakukan penulis, yaitu peringkasan multi-dokumen untuk dokumen berbahasa Indonesia, dapat ditarik beberapa kesimpulan yang mengacu kepada hasil eksperimen dan evaluasi. Teknik peringkasan *centroid-based summarization* dan *k-means-based summarization* ternyata dapat diterapkan untuk peringkasan dokumen berbahasa Indonesia.

Kualitas dan besar kecilnya kelompok dokumen sangat mempengaruhi kualitas dari ringkasan yang dihasilkan. Terbukti dari eksperimen yang dilakukan pada penelitian ini, hasil evaluasi perbandingan terhadap ringkasan referensi menunjukkan bahwa kelompok dokumen yang memiliki nilai kualitas tinggi akan menghasilkan ringkasan yang berkualitas tinggi pula. Begitu pula sebaliknya, kualitas ringkasan akan menjadi lebih buruk jika kelompok dokumen memiliki nilai kualitas yang rendah.

Hal ini disebabkan semakin banyak kalimat-kalimat pada kelompok dokumen yang membawa informasi yang relevan terhadap topik, maka kemungkinan kalimat-kalimat tersebut muncul pada ringkasan juga akan semakin besar. Begitu pula sebaliknya, semakin banyak kalimat-kalimat pada kelompok dokumen yang tidak relevan terhadap topik, maka kemungkinan kalimat-kalimat yang tidak relevan tersebut untuk muncul pada ringkasan juga akan semakin besar. Hal tersebut terjadi terutama pada ringkasan yang dihasilkan oleh teknik *centroid-based summarization*, dimana kualitas ringkasan sangat dipengaruhi oleh kualitas kelompok dokumen.

Besar kelompok dokumen juga mempengaruhi kualitas dari ringkasan yang dihasilkan. Kelompok dokumen yang besar berarti memuat banyak informasi di dalamnya, sebaliknya kelompok dokumen yang lebih kecil berarti memuat lebih sedikit informasi di dalamnya. Terbukti dari hasil eksperimen penelitian, kelompok

dokumen yang lebih besar menghasilkan ringkasan yang lebih bagus, karena mengandung lebih banyak cakupan informasi dibandingkan dengan ringkasan dari kelompok dokumen yang kecil. Hal tersebut terjadi terutama pada ringkasan yang dihasilkan oleh teknik *k-means-based summarization*, dimana kelompok dokumen yang lebih besar menghasilkan ringkasan yang lebih bagus dibandingkan dengan kelompok dokumen yang lebih kecil.

Nilai evaluasi yang diberikan oleh juri terhadap ringkasan menunjukkan ringkasan yang dihasilkan oleh teknik *centroid-based summarization* memiliki nilai rata-rata di antara 3 sampai dengan 4.33. Hal tersebut mengindikasikan bahwa para juri menilai ringkasan yang dihasilkan oleh teknik *centroid-based summarization* memiliki kualitas yang bagus. Ringkasan yang dihasilkan oleh teknik *k-means-based summarization* pada *10% compression rate* memiliki nilai rata-rata antara 2.33 sampai dengan 4.33. Sedangkan pada *20% compression rate*, 6 dari 7 ringkasan memiliki nilai rata-rata sama dengan atau di atas 3. Hal tersebut mengindikasikan bahwa pada *10% compression rate* para juri kurang puas terhadap ringkasan yang dihasilkan, namun pada *20% compression rate* para juri menilai ringkasan yang dihasilkan cukup bagus.

Namun untuk kelompok dokumen yang sangat besar, contohnya seperti kelompok dokumen G (lihat Tabel 3.1 bagian 3.1.1), ringkasan yang dihasilkan akan menjadi sangat panjang. Hal tersebut membuat kalimat-kalimat yang membawa informasi yang berulang ikut masuk ke dalam ringkasan, sehingga banyak terjadi pengulangan informasi pada ringkasan dan membuat kualitas ringkasan menjadi tidak bagus. Hal tersebut dapat dihindari dengan menggunakan *compression rate* yang lebih kecil. Dengan penggunaan *compression rate* yang lebih kecil, kalimat-kalimat berulang yang nilainya telah dikurangi dengan nilai penalti menjadi tidak terpilih untuk masuk ke dalam ringkasan

Dari hasil eksperimen pada penelitian ini, dapat pula disimpulkan bahwa kualitas ringkasan yang dihasilkan oleh teknik peringkasan *centroid-based summarization* lebih ditentukan oleh kualitas kelompok dokumen dan kualitas ringkasan yang dihasilkan oleh teknik peringkasan *k-means-based summarization* lebih ditentukan oleh besar/kecilnya kelompok dokumen. Dari hasil eksperimen juga menunjukkan

bahwa ringkasan yang dihasilkan oleh teknik *centroid-based summarization* lebih baik hasilnya jika dibandingkan dengan ringkasan yang dihasilkan oleh teknik *k-means-based summarization*.

5.2 Saran

Saran-saran yang diberikan penulis untuk pengembangan pada penelitian dengan topik peringkasan dokumen yang selanjutnya adalah sebagai berikut:

- Penerapan teknik-teknik atau pendekatan yang lain ke dalam metode peringkasan multi-dokumen. Hal tersebut dimaksudkan agar dapat menghasilkan ringkasan yang lebih baik, khususnya dalam *domain* peringkasan untuk dokumen berbahasa Indonesia.
- Pembuatan korpus standar dengan mengumpulkan berbagai macam jenis dokumen berbahasa Indonesia. Hal ini perlu dilakukan agar penelitian-penelitian dengan topik peringkasan dokumen yang selanjutnya sudah memiliki korpus yang besar yang dapat digunakan sebagai data penelitian.
- Penggunaan berbagai macam metode evaluasi yang lain untuk menilai kualitas ringkasan yang dihasilkan oleh peringkasan dokumen, sehingga kualitas ringkasan yang dihasilkan dapat diukur lewat berbagai macam penilaian.