

BAB I

PENDAHULUAN

1.1 Latar Belakang

Universitas Komputer Indonesia (UNIKOM) secara resmi berdiri pada hari Selasa, tanggal 8 Agustus 2000 berdasarkan Surat Keputusan Menteri Pendidikan Nasional nomor 126/D/0/2000. Selama mengikuti program pendidikan, mahasiswa wajib menyusun karya ilmiah yang meliputi skripsi dan tesis untuk peserta didik sarjana dan pascasarjana. Karya ilmiah yang akan dipublikasikan perlu diperiksa mengenai nilai ilmiahnya, yang tidak boleh ada plagiarisme di dalamnya. Pengecekan kemiripan dokumen karya ilmiah bisa dilakukan secara konvensional yaitu dengan membaca dan membandingkan karya ilmiah dengan karya ilmiah lainnya satu persatu, dan *model* seperti ini sangat tidak efektif dan efisien. Sehingga perlu dibuatkan suatu sistem komputerisasi yang dapat menggantikan proses pengecekan secara konvensional.

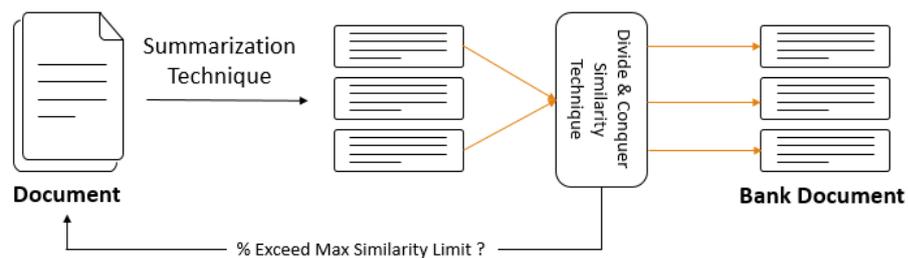
Beberapa penelitian telah membuat sistem pengecekan kemiripan karya ilmiah. Pengukuran kemiripan dokumen karya ilmiah dapat dilakukan dengan membandingkan jumlah kata kunci yang terdapat dalam sebuah dokumen karya ilmiah dengan dokumen karya ilmiah lainnya dalam waktu yang lebih cepat dengan Algoritma Genetika [1].

Dari semua metode dan algoritma perbandingan dokumen yang ada dan digunakan saat ini, proses perbandingan dokumen karya ilmiah dilakukan secara *brute force* yaitu membandingkan seluruh isi dokumen, sehingga mengakibatkan

tingkat komputasi sangat tinggi dan memakan waktu cukup lama jika bank dokumen yang ada sangat besar.

Terinspirasi dari J. Chen *and* Y. Saad tentang “*Divide and Conquer Strategies for Effective Information Retrieval*”, dimana strategi dalam proses komputasi dengan memecah-mecah proses yang ada menjadi beberapa bagian kecil sehingga lebih mudah untuk diselesaikan [2]. Dalam *Information Retrieval* suatu dokumen dapat diambil ringkasan informasinya sehingga tidak perlu melihat seluruh isi informasi yang tidak perlu. Dalam jurnal “*Abstractive Text Summarization Using Sequence-To-Sequence Rnns And Beyond*” dapat mengambil ringkasan dari suatu artikel dengan metode *Sequence-To-Sequence Rnns And Beyond* [3].

Dalam kasus ini penulis akan mencoba menggabungkan teknik *Text Summarization* dan *Similarity Document*. Dalam metode ini, dokumen akan dipecah menjadi ringkasan-ringkasan kalimat, dari ringkasan-ringkasan dilakukan perbandingan dengan ringkasan-ringkasan yang ada di bank dokumen sebelumnya. Dalam metode yang diusulkan ini dapat memberikan tingkat kemiripan suatu dokumen karya ilmiah dengan dokumen karya ilmiah yang ada di bank dokumen.



Gambar 1.1 Proses Perbandingan Dokumen yang Diusulkan

Dengan usulan tersebut, diharapkan dapat mempercepat proses identifikasi kemiripan suatu dokumen tanpa harus membandingkan seluruh isi dari dokumen.

1.2 Identifikasi Masalah

Proses perbandingan dokumen karya ilmiah yang ada saat ini dilakukan secara *brute force* yaitu membandingkan isi dari keseluruhan dokumen dengan bank dokumen sehingga proses perbandingannya memakan waktu yang cukup lama jika bank dokumen sudah sangat besar. Sementara kebutuhan saat ini adalah dapat mempercepat proses identifikasi kemiripan dokumen tanpa harus membandingkan seluruh isi dari dokumen.

1.3 Pembatasan Masalah

Pembatasan masalah membahas titik awal penelitian dan ruang lingkup kajian sehingga penelitian ini dapat lebih terarah, berikut pemaparan batasan-batasan masalah yang dikemukakan:

1. Data yang dicek adalah data skripsi unikom jurusan teknik informatika.
2. Format data yang dibaca sistem adalah format pdf.
3. Dalam pengecekan dokumen akan dipecah menjadi beberapa bab, yaitu bab1, bab2, bab3, bab4, dan bab5.
4. Pengecekan kemiripan akan dilakukan perkalimat yang sudah dilakukan teknik *Text Summarization*.
5. Dalam penelitian ini hanya mengecek tingkat kemiripan dokumen, tidak sampai mengecek plagiarisme.

6. Karya ilmiah yang akan dicek adalah karya ilmiah dalam bahasa indonesia.
7. Algoritma yang digunakan untuk teknik *text summarization* adalah algoritma *Ant Colony Optimization*.
8. Algoritma yang digunakan untuk teknik *similarity document* adalah algoritma *smith-waterman*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menggunakan teknik *Text Summarization* dengan teknik *Similarity Document* dalam mengecek tingkat kemiripan dokumen.
2. Mengukur tingkat akurasi dari hasil proses pengecekan kemiripan dokumen.
3. Mengukur tingkat komputasi dari hasil proses pengecekan kemiripan dokumen.
4. Memberikan informasi tingkat kemiripan karya ilmiah dan memberikan rekomendasi perubahan kalimat untuk mengurangi persentasi tingkat kemiripan.

Diharapkan dari tujuan penelitian ini, dapat membantu proses pengambilan keputusan bagi penyelenggara/dosen yang ingin melihat atau mengecek tingkat orisinal suatu dokumen karya ilmiah.

1.5 Manfaat Penelitian

Pada penelitian ini, akan menghasilkan suatu kontribusi pada sisi keilmuan, yaitu menghasilkan suatu *model* yang dapat membandingkan ringkasan dari suatu dokumen dimana prosesnya dibagi-bagi ke dalam beberapa proses kecil, yang merupakan pengembangan dari *model-model* sebelumnya. Bagi kegiatan penelitian diharapkan dapat membuka jalur penelitian-penelitian baru lainnya dimana pengembangan teknik-teknik yang lebih baik untuk mencapai tingkat kecepatan dan ketepatan yang efektif dan efisien dalam pengecekan kemiripan Dokumen.

1.6 Premis dan Hipotesis

Premis dan Hipotesis dalam penelitian ini dari hasil-hasil penelitian terdahulu yang dapat memperkuat penarikan jawaban sementara yang kebenarannya akan dibuktikan melalui penelitian ini, sebagai berikut:

1.6.1 Premis

1. Teknik *Text Summarization* digunakan untuk mendapatkan ringkasan dari suatu dokumen karya ilmiah.
2. Teknik *Similarity Document* digunakan untuk membandingkan tingkat kemiripan dari sebuah dokumen karya ilmiah dengan dokumen karya ilmiah lainnya.

1.6.2 Hipotesis

Berdasarkan premis tersebut, dapat dibentuk hipotesis yaitu “Penggunaan teknik *Text Summarization* dan *Similarity Document* dapat lebih mempercepat proses deteksi kesamaan suatu dokumen karya ilmiah”

1.7 Sistematika Penulisan

Laporan penelitian ini disusun dengan sistematika sebagai berikut:

BAB I : Pendahuluan

Bagian pendahuluan ini menjelaskan latar belakang penelitian, identifikasi masalah, premis dan hipotesis, tujuan, metoda yang digunakan dan sistematika penelitian.

BAB II : TINJAUAN PUSTAKA

Bagian landasan teori membahas tentang pengertian dari perbandingan dokumen, dan teknik-teknik yang akan digunakan dalam penelitian.

BAB III : METODE PENELITIAN

Bagian ini membahas tahap desain *model* yang diusulkan, tahapan-tahapan proses yang akan dilakukan dalam *model*, implementasi *model* dalam kasus perbandingan dokumen karya ilmiah dan batasan-batasan yang harus ditetapkan dalam *model* terkait.

BAB IV : HASIL DAN PEMBAHASAN

Bagian pengujian dan analisis akan berisi hasil uji komputasi dari *model* yang telah dikembangkan menggunakan sampling data, analisis kelebihan dan

kekurangan dari *model* yang telah dibuat dibandingkan dengan *model* yang ada saat ini.

BAB V : KESIMPULAN DAN SARAN

Bagian penutup menyimpulkan hasil-hasil yang diperoleh dari penelitian ini dan kecocokan *model* yang digunakan dalam perbandingan dokumen dalam kasus tertentu. Selain itu juga dibahas jalur pengembangan penelitian tahap berikutnya.