

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Kualitas penelitian dan karya tulis ilmiah dalam dunia akademik sangat penting untuk diperhatikan. Bahkan kualitas karya tulis tersebut didasari dari berbagai aspek dan faktor. Salah satu yang paling penting adalah referensi literatur yang mendasari suatu karya tulis ilmiah tersebut.

Dalam menjaga kualitas karya tulis ilmiah, berbagai kendala pun dialami oleh para penulis. Dari mulai penyiapan judul penelitian yang matang, mencari latar belakang penelitian, mencari dan membandingkan literatur pendukung, mengumpulkan dan mengolah data pendukung, menjaga orisinalitas karya tulis dan lain sebagainya [1].

Masalah-masalah tersebut muncul disetiap tahapan penelitian. Dimulai dari proses pengumpulan ide dan pembahasan masalah yang ingin dijadikan fokus penelitian. Muncul permasalahan dimana para peneliti sulit mendapatkan pokok permasalahan yang ada. Pencarian literatur pendukung penelitian seperti penelitian terdahulu yang serupa dimana para peneliti harus mencari kata kunci yang cocok dalam penelitiannya. Proses penyusunan penelitian dan akhirnya setelah selesai penyusunan, pengecekan originalitas penelitian.

Untuk proses awal pengumpulan ide dan pembahasan masalah yang akan dijadikan fokus penelitian memang lebih menitik beratkan kepada masalah internal

peneliti. Namun untuk proses pencarian literatur, penyusunan dan pengecekan originalitas penelitian dapat dianggap sebagai faktor eksternal.

Penggunaan teknologi sudah dimanfaatkan oleh para penyusun karya tulis ilmiah dalam proses pemecahan masalah-masalah tersebut. Saat ini para penyusun/peneliti menentukan kata kunci yang menjadi pokok permasalahan pada penelitian mereka dan mencarinya di mesin pencari berbasis online untuk mendapatkan tulisan-tulisan yang serupa dan dapat mendukung penelitian mereka. Sementara untuk pengecekan originalitas penelitian, para peneliti biasanya menggunakan aplikasi pengecek tingkat kemiripan dokumen yang sudah dikembangkan.

Namun dari solusi yang sudah ada saat ini, masih ada beberapa kekurangan atau kelemahan. Yang pertama, para penyusun karya tulis ilmiah masih kesulitan menentukan kata kunci yang tepat dalam proses pencarian literatur pendukung. Lalu pada penggunaan aplikasi pengecek tingkat kemiripan dokumen, para peneliti tidak menemukan rekomendasi yang seharusnya diperbaiki dan dicantumkan oleh penulis.

Dari bidang sistem informasi salah satu solusi yang dapat dikembangkan adalah membangun model perbandingan dokumen yang dapat memudahkan para peneliti mengatasi masalah-masalah tersebut. Setelah para peneliti mengetahui latar masalah dalam penelitian, model ini diharapkan dapat memberikan rekomendasi literatur yang sejalan dengan pokok yang dibahas oleh peneliti. Dalam prosesnya juga dapat mengecek kemiripan dokumen dengan dokumen-dokumen yang ada dan

mengantisipasi adanya indikasi plagiarisme dan meningkatkan kualitas orisinalitas penelitian/karya tulis.

Model tersebut dapat dikembangkan dengan adanya metode-metode perbandingan dokumen. Seperti algoritma Doc2Vec yang mempelajari representasi vektor terdistribusi berkelanjutan untuk potongan-potongan teks [2]. Pengembangannya yang disebut Doc2VecC yang menambahkan komponen *corruption* untuk menghilangkan kata tak beraturan/acak dari data latih [3]. Dan algoritma Sent2Vec yang lebih menitikberatkan perbandingan kepada kalimat perkalimat [4].

Untuk meningkatkan akurasi dan meringankan tingkat komputasi yang tinggi dan memakan waktu yang cukup lama pada metode-metode *machine learning* tersebut [5], terdapat satu cara yang dapat menjadi solusi yaitu melakukan fragmentasi dokumen pada saat melakukan *pre-processing* dokumen-dokumen yang ada dengan membagi dokumen kedalam beberapa kalimat, paragraf dan bab untuk selanjutnya dibandingkan dengan metode yang ada.

Dengan memfragmentasi dokumen yang ada, metode ini dapat memberikan akurasi yang lebih baik karena konteks bahasan akan lebih terpusat dan dari sisi komputasi, jumlah vektor yang dibandingkan akan lebih terbagi sehingga membantu mengurangi tingkat komputasi yang tinggi tersebut.

## **1.2. Identifikasi Masalah**

Para peneliti dan penyusun karya tulis ilmiah membutuhkan rekomendasi dalam pencarian literatur dan perbandingan kemiripan dokumen. Sementara proses

perbandingan dokumen yang ada saat ini dilakukan secara *sequential* memerlukan proses komputasi yang besar dan memakan waktu cukup lama. Sementara kebutuhan bagi para penyusun, peneliti atau dosen yang ingin membandingkan suatu dokumen karya ilmiah dengan *knowledge base* yang sudah ada adalah kecepatan dan ketepatannya.

### **1.3. Premis dan Hipotesis**

Dari latar belakang permasalahan tersebut, dapat dinyatakan terdapat premis-premis seperti berikut:

- 1) Rekomendasi literatur berdasarkan perbandingan kesamaan dokumen diperlukan oleh para peneliti.
- 2) Teknik Doc2Vec yang unggul dalam metode state-of-the-art dari perbandingan kesamaan dokumen [5] membandingkan keseluruhan isi dokumen secara *sequential*.
- 3) Dokumen Karya Ilmiah memiliki karakteristik tiap bagian paragraf dan bab yang tidak terlalu berhubungan dengan bagian lainnya, maka tiap bagian dapat dibandingkan menjadi obyek vektor sendiri.

Berdasarkan premis tersebut, dapat dibentuk hipotesis yaitu “Penggunaan model fragmentasi dokumen dan algoritma Doc2Vec pada bagian-bagian dokumen yang tidak berhubungan dapat mempercepat proses perbandingan kesamaan antar dokumen karya ilmiah”.

#### **1.4. Tujuan**

Adapun tujuan dari penelitian ini adalah memberikan solusi dan rekomendasi bagi para peneliti dengan peningkatan akurasi dan kecepatan perbandingan dokumen dengan metode atau strategi baru yang menggunakan model fragmentasi dari suatu dokumen karya ilmiah.

Sehingga hasil perbandingan yang lebih cepat dan akurat ini diharapkan dapat membantu proses penyusunan, memberikan rekomendasi dan menjadi landasan pengambilan keputusan bagi peneliti/penyusun/dosen yang ingin membandingkan dokumen karya ilmiah seperti skripsi, thesis, disertasi dan lain sebagainya.

#### **1.5. Lingkup Penelitian**

Adapun dalam pelaksanaannya, lingkup atau batasan dari penelitian ini diantaranya adalah:

- 1) Penelitian ini membahas terkait perbandingan karya tulis ilmiah dengan pembuatan model pada objek karya tulis berbahasa Indonesia;
- 2) Perbandingan dokumen yang diteliti dalam penelitian ini adalah perbandingan bersifat kesamaan / kemiripan (*Document Similarity*);
- 3) Data yang digunakan sebagai data latih dan data uji berasal dari ELIB Unikom.

## **1.6. Kontribusi**

Dari sisi keilmuan, kontribusi yang dapat diberikan dari penelitian ini adalah dihasilkannya jalur baru penelitian model perbandingan dokumen secara fragmentasi dan paralel yang merupakan pengembangan dari model-model sebelumnya yang menggunakan cara *sequential*. Juga terbukanya wawasan penggunaan teknik Doc2Vec dalam perbandingan dokumen dengan model yang baru dan dengan bahasa baru yaitu Bahasa Indonesia.

Dari sisi penelitian, penelitian ini 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 perbandingan Dokumen.

## **1.7. Metodologi Penelitian**

Metodologi yang dilakukan pada penelitian ini adalah sebagai berikut:

- 1) Identifikasi, perumusan konsep dan pembentukan batasan

Pada tahap ini, dilakukan identifikasi masalah yang ada dari perbandingan dokumen secara *sequential* setelah itu ditentukan kerangka model yang akan dibuat. Selain itu, batasan-batasan dan spesifikasi dari model yang akan dibentuk juga ditentukan terlebih dahulu.

- 2) Perancangan dan pengembangan model

Pada tahap ini, dilakukan analisa tahapan-tahapan dari kerangka model yang akan dibuat, diperdalam dan ditentukan proses yang akan ada dalam model ini.

3) Pembuatan model dan implementasi

Setelah rancangan dan analisis model terbentuk, model kemudian dibuat dan di implementasikan sebagai sistem perbandingan kesamaan dokumen. Tidak lupa parameter batasan yang ada juga diukur dan diterapkan dalam model.

4) Pengujian model menggunakan sampling data

Pada tahap pengujian ini, model yang sudah dibentuk diuji dengan data sampling yang telah dibuat agar model yang telah dibuat dapat diukur dari segi kecepatan dan keakuratan perbandingan dokumennya sendiri.

5) Analisis dan penilaian konsep

Setelah model diuji, tahap terakhir ini menganalisa hasil uji dari model tersebut dan menimbang atau menilai konsep yang ditentukan apakah dapat memecahkan masalah yang ada dan bagaimana kontribusinya apabila diimplementasikan dalam studi kasus yang lebih nyata.

## **1.8. Sistematika Penelitian**

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 Studi Literatur**

Bagian studi literatur membahas tentang pengertian dari perbandingan dokumen, metode fragmentasi dokumen, penentuan prioritas pada bagian dokumen dan teknik-teknik yang akan digunakan dalam penelitian.

## **BAB III Perancangan Model**

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 Pengujian dan Analisis**

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 Penutup**

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.