

1. Pendahuluan

Bagian ini menjelaskan mengenai latar belakang, perumusan masalah, tujuan, ruang lingkup, batasan masalah penelitian, hipotesa, metodologi penyelesaian masalah, serta jadwal kegiatan untuk menyelesaikan masalah Tugas Akhir ini.

1.1 Latar Belakang

Pada saat ini, kemajuan dibidang teknologi dan informasi semakin berkembang cepat. Sehingga penyebaran data digital pun semakin luas dan mudah. Dalam penyebaran data digital misalnya yang berbentuk dokumen teks digital dengan menggunakan internet, data dapat dimanipulasi sehingga kemungkinan data tersebut tidak asli (hasil manipulasi orang lain). Kebanyakan dokumen yang dipublikasikan diinternet merupakan jurnal-jurnal ilmiah. Artikel-artikel tersebut merupakan karya dari seseorang yang perlu dilidungi hak ciptanya agar terhindar dari segala macam penyalahgunaan [1].

Salah satu bentuk pencegahan dalam penyalahgunaan dokumen teks digital yaitu dengan *watermarking*. *Watermarking* tidak mencegah terjadinya manipulasi dokumen teks, tetapi akan ditelusuri asal dan kepemilikan dari dokumen tersebut. *Watermarking* pada dokumen teks berbeda dengan *watermarking* pada gambar dan video. Pada teks penyimpanan data *watermark* pada bit-bit data dari dokumen sehingga akan mengubah isi dari dokumen [9].

Natural Language Watermarking mempunyai prinsip yang sama dengan *watermarking* biasa. *Natural Language Watermarking* bertujuan untuk menanamkan informasi dalam dokumen dengan cara memanipulasi struktur semantik dan atau sintaksis pada dokumen teks digital. *Watermarking* pada dokumen teks dilakukan dengan mengubah format atau font teks seperti memodifikasi spasi antar-kata atau antar-kalimat [1].

Pada tugas akhir sebelumnya metode yang digunakan dalam mengimplementasikan *Natural Language Watermarking* adalah *Syntactic Alteration*. Metode tersebut dicapai melalui beberapa hal yaitu *tagging*, *treebank sentence* dan mentransformasi kalimat. Namun, metode tersebut memiliki kekurangan dalam performansi pengujian *coverange*, tidak tahan terhadap *watermarking* ulang, dan pengurangan kalimat [2].

1.2 Perumusan masalah

Perumusan masalah dari Tugas Akhir ini adalah adanya kekurangan dari metode *Syntactic Alteration* yaitu

1. Performansi *coverage* yang kurang baik
2. Tidak tahan terhadap *watermarking* ulang
3. Tidak baik dalam pengurangan kalimat.

1.3 Batasan Masalah

Karena waktu dan sumber daya yang terbatas dalam pembuatan tugas akhir ini maka dibuatlah beberapa batasan masalah, antara lain:

1. Bahasa yang digunakan adalah hanya bahasa Indonesia baku.
2. Data yang akan digunakan dalam implementasi Tugas Akhir ini berasal dari artikel online surat kabar dan PAN Localization.

1.4 Tujuan

Tujuan dari Tugas Akhir ini adalah membuat atau merancang *Natural Language Watermarking* pada dokumen teks digital berbahasa Indonesia yang tahan terhadap *watermarking* ulang, pengurangan kalimat dan peningkatan tingkat *coverange*.

1.5 Hipotesa

Untuk mencapai tujuan tersebut maka digunakan metode *Morphological Division* dan *Insertion*. Dengan metode ini dapat meningkatkan tingkat *coverage* yang lebih baik dari pada sistem referensi [2] yang hanya sebesar 50%. Hal ini dikarenakan penyisipan yang dilakukan per kata sedangkan pada sistem referensi [2] penyisipan dilakukan per kalimat. Selain itu, metode ini diharapkan juga bisa diterapkan pada bahasa Indonesia, ketahanan terhadap *watermarking* ulang dan pengurangan kalimat untuk disisipi *watermark*.

1.6 Metodologi penyelesaian masalah

Untuk melakukan proses *watermarking* pada kasus ini, metode yang digunakan adalah *Morphological Division* dan *Insertion*. Tahap penyelesaian dalam Tugas Akhir ini adalah studi literatur, perancangan, implementasi sistem, pengujian dan pengumpulan data, serta analisa hasil.

Studi Literatur mencakup pencari referensi berkaitan dengan Tugas Akhir ini yaitu *natural language watermarking*, *insertion of words* dan *morphological division*.

Perancangan dilakukan dengan metode diagram blok yang menggambarkan sistem secara keseluruhan mulai dari teks diinputkan, *watermarking* disisipkan sampai dengan analisis hasil pengujian.

Implementasi sistem yang digunakan untuk menyusun tugas akhir ini adalah dengan metode *morphological division* dan *insertion*, maka diperlukan tahap *preprocessing watermarking*. Selanjutnya tahap-tahap yang dilakukan adalah *watermarking selector*, *embedding watermarking*, *watermarking extraction*, dan *watermark tester*.

Pengujian dan pengumpulan data yang digunakan untuk pengujian merupakan teks digital berbahasa Indonesia yang berisi kalimat-kalimat pernyataan atau kalimat berita. Untuk kalimat-kalimat sendiri diambil dari surat kabar digital dan PAN Localization[3,4]. Strategi pengujian untuk mengukur *naturalness* adalah dengan membuat *tools* untuk pengujian sistem setelah didapatkan outputnya dan berdasarkan *human judgment* atau pengujian dari beberapa ahli tentang bahasa Indonesia, seperti guru, dosen dan lainnya.

Analisa hasil yang mencakup performansi dari sistem yang berupa *coverage* dari kata dan kalimat yang belum dan telah tersisipi watermark serta *naturalness* dari bahasa yang dimodifikasi.