

ABSTRAK

NORMALISASI TEKS PADA CHATBOT SISTEM INFORMASI AKADEMIK MENGGUNAKAN ALGORITME DAMERAU– LEVENSHTEIN DISTANCE DAN PREFIX TREE (STUDI KASUS: UNIVERSITAS TEKNOKRAT INDONESIA)

*Text Normalization In Academic Information System Chatbot Using
Damerau–Levenshtein Distance Algorithm And Prefix Tree (Case Study:
Teknokrat University of Indonesia)*

Oleh

Muhammad Thomas Fadhila Yahya

15312574

Penelitian ini dilakukan atas dasar permasalahan pada *chatbot* yang sulit memahami dan merespons pesan dengan tepat karena terdapat kesalahan pengetikan, tata bahasa dan penggunaan bahasa yang buruk dalam pesan pengguna. Oleh karena itu dibutuhkan sistem normalisasi teks yang dapat mengubah bentuk pesan pengguna ke dalam bentuk baku. Dengan sistem ini diharapkan dapat mengatasi permasalahan *chatbot* saat memahami dan merespons pesan.

Sistem ini terdiri tujuh tahapan normalisasi yaitu normalisasi garis baru, normalisasi huruf kecil, normalisasi karakter berulang, normalisasi spasi, tokenisasi, normalisasi kata dasar, dan pengecekan ejaan. Pada tahap pengecekan ejaan, menggunakan algoritme Damerau–Levenshtein Distance untuk menghitung jarak *string* dan fungsi Perhitungan Kedekatan Huruf. Prefix Tree digunakan untuk mengubah data korpus ke dalam bentuk *node*.

Berdasarkan hasil penelitian dan implementasi, diketahui bahwa menggunakan algoritme Damerau–Levenshtein Distance dan fungsi Perhitungan Kedekatan Huruf menghasilkan nilai *mean average precision* sebesar 0,86. Dan menggunakan Prefix Tree menghasilkan waktu proses sebesar 0.004 detik untuk kata dengan panjang mulai dari 3 karakter, bertambah 0.002 detik untuk setiap karakter. Untuk kata dengan panjang mulai dari 11 karakter, peningkatan waktu proses berubah menjadi 0.003 detik untuk setiap karakter.

Kata Kunci : *Chatbot, Normalisasi, Damerau–Levenshtein Distance, Prefix Tree*