

Abstrak

Konversi *Indonesian grapheme-to-phoneme (G2P)* merepresentasikan sebuah tugas memetakan setiap grafem / symbol eja dalam sembarang kata yang dikenal dalam bahasa Indonesia ke representasi fonemik / simbol pelafalannya.

Pencarian metode terbaik yang dilakukan dalam tugas akhir ini memberikan hasil berupa penetapan sebuah model bernama *IG-tree* + strategi tebakan-terbaik sebagai metode yang terpilih untuk memecahkan permasalahan konversi *G2P*. Model tersebut pada dasarnya menggunakan struktur pohon-keputusan yang dibangun berdasarkan data *training*, dikonstruksikan menggunakan konsep *information gain (IG)* dalam menentukan kepentingan relatif atribut-atribut, dan dilengkapi dengan strategi tebakan-terbaik dalam mengklasifikasikan instan-instan baru. Akan tetapi sistem dalam tugas akhir ini dikembangkan lebih lanjut dengan properti-properti baru yang ditambahkan pada struktur asalnya untuk meningkatkan performansi sistem. Mekanisme *pruning* diusulkan untuk model dengan dua tujuan: (1) meningkatkan kemampuan generalisasi model, dan (2) meminimalkan ukuran model. Properti baru yang lain, *peng-handle* kasus homograf menggunakan metode kategorisasi teks, diusulkan untuk sistem untuk menangani kasus khususnya berupa beberapa himpunan kata yang sepenuhnya sama dalam representasi grafemik namun berbeda satu sama lain dalam representasi fonemik.

Ditunjukkan dalam tugas akhir ini bahwa model tersebut secara umum berperformansi bagus sementara properti-properti tambahan yang diusulkan memang memberikan keuntungan tambahan sebagaimana yang diharapkan.

Kata kunci: konversi *grapheme-to-phoneme*, bahasa Indonesia, *IG-tree*, strategi tebakan-terbaik, *pruning*, *peng-handle* kasus homograf