

Abstrak

Model bahasa adaptif adalah modul penting untuk mengembangkan sistem pengenalan suara otomatis. Laporan ini berfokus pada pengembangan model bahasa adaptif untuk Bahasa Indonesia. Diawali dengan pembangunan korpus teks 10 juta kalimat dengan melakukan crawling pada beberapa situs berita, majalah, blog pribadi, dan forum penulisan di Indonesia. Korpus teks tersebut kemudian digunakan untuk membangun model bahasa adaptif menggunakan Latent Dirichlet Allocation (LDA) dengan metode pelatihan Collapsed Gibbs Sampling. Model bahasa adaptif memberikan kinerja yang lebih baik dalam pemilihan kata untuk menghasilkan kalimat terbaik.

Kata kunci : model bahasa adaptif, Bahasa Indonesia, collapsed gibbs sampling, korpus teks, latent dirichlet allocation