

Abstrak

Pengenalan Ucapan Kontinu Kamus Besar atau PUKKB merupakan sistem pengenalan suara yang paling mutakhir. Sistem ini mampu mengenali berbagai macam suara dan kata yang diucapkan orang. Kemampuan pengenalan tersebut didapat dengan melatih sistem menggunakan korpus suara membaca dan korpus suara spontan.

Korpus suara merupakan elemen penting dalam melatih sistem tersebut, terutama korpus suara spontan. Korpus ini menjadi referensi cara pengucapan bagi sistem tersebut. Beberapa bahasa seperti bahasa Inggris, sistem seperti ini mudah dikembangkan karena terdapat banyak korpus suara yang beredar, tetapi untuk beberapa bahasa seperti bahasa Indonesia, korpus suara yang beredar masih sedikit.

Dengan memadukan desain aplikasi pengumpul suara seperti Eyra, Woefzela dan Data Hound penulis mengembangkan aplikasi serupa untuk diaplikasikan dalam pembangunan korpus suara bahasa Indonesia, terutama dalam pengumpulan data suara.

Kata kunci: *Under-resourced*, korpus suara spontan, cakupan *triphone*, pembangkitan pertanyaan, *balanced sentence set*.