

MEMBANGUN SISTEM PENGENALAN UCAPAN OTOMATIS BAHASA INDONESIA MENGGUNAKAN METODE HIDDEN MARKOV MODEL

ASKARYA QAULAN SYADIDA

(Pembimbing : Dr. MULJONO, S.Si, M.Kom)

Teknik Informatika - S1, FIK, Universitas Dian Nuswantoro

www.dinus.ac.id

Email : 111201307960@mhs.dinus.ac.id

ABSTRAK

Dengan semakin majunya teknologi menuntut kemudahan dalam pengaksesan informasi dan penyampaian data. Menjawab tantangan tersebut, melalui perangkat komputer masukan berupa ucapan atau ujaran dapat mewujudkan tuntutan itu. Sinyal ucapan yang berupa analog di ubah ke dalam sinyal digital melalui perantara microphone. Sinyal digital kemudian di transkripsikan ke dalam suatu bilangan biner sehingga komputer dapat membaca dan menampilkan hasilnya. Teknologi ini menjadi dasar dari Sistem Pengenalan Ucapan Otomatis. Berdasarkan pendekatan yang dilakukan peneliti sebelumnya untuk membentuk model Sistem Pengenalan Ucapan Otomatis berbasis Bahasa Indonesia dengan menggunakan transisi dari model atau pendekatan tata bahasa Bahasa Inggris menghasilkan akurasi Word Error Rate yang bernilai besar. Secara garis besar untuk menghasilkan nilai akurasi yang baik dalam implementasi Sistem Pengenalan Ucapan Otomatis diperlukan pendekatan terhadap aturan tata Bahasa Indonesia. Sistem Pengenalan Ucapan Otomatis berbasis Bahasa Indonesia dapat mengenali ucapan berbentuk berkas rekaman yang diberikan, dengan format berkas waveform audio format dengan besar bit rate 8000 Hz. Pengolahan data rekaman menggunakan toolkit dari CMUSphinx, dan pembangunan aplikasi menggunakan library Sphinx4, serta memanfaatkan algoritma Hidden Markov Model. Setelah dilakukan proses testing terhadap model akustik yang terbentuk, didapatkan nilai word error rate sebesar 23,0 % dan sentence error rate sebesar 32,8 %. Di mana pada implementasinya nilai kedua variabel ini dinamis dan selalu berada pada lingkup nilai testing tersebut. Semakin rendah kedua variabel tersebut maka semakin baik pengenalan terhadap inputan berkas ucapan yang diberikan.

Kata Kunci : sistem pengenalan ucapan otomatis, model akustik, hidden markov model, CMUSphinx toolkit, sphinx4.

BUILDING AUTOMATIC RECOGNITION SYSTEM FOR INDONESIAN LANGUAGE USING HIDDEN MARKOV MODEL METHOD

ASKARYA QAULAN SYADIDA

(Lecturer : Dr. MULJONO, S.Si, M.Kom)

*Bachelor of Informatics Engineering - S1, Faculty of Computer
Science, DINUS University*

www.dinus.ac.id

Email : 111201307960@mhs.dinus.ac.id

ABSTRACT

The advancement of technology requires ease in accessing information and delivering data. Answering these challenges, through computer input devices such as speech can realize the demands. The analogue speech signal is converted into a digital signal through an intermediary microphone. The digital signal is then transcribed into a binary number so that the computer can read and display the results. This technology is the basis of Automatic Speech Recognition System. Based on the approach of previous researchers to establish the model of Automatic Speech Recognition System on Indonesian language using transition from English language model or grammatical approach resulted in great value of Word Error Rate accuracy. Broadly speaking to produce a good accuracy value in the implementation of Automatic Speech Recognition System required approach to the rules of grammar Indonesia language. The Indonesian-based Auto Speech Recognition System can recognize speech in the form of a given recording file, with a file waveform audio format which this file have a large bit rate of 8000 Hz. Recording data processing uses CMUSphinx toolkits, and application development using the Sphinx4 library, and utilizes the Hidden Markov Model algorithm. After the testing of the acoustic model, the value of word error rate is 23.0% and the sentence error rate is 32.8%. Where in the implementation of that two value is dynamic and always be on the scope of the testing value. The lower the two variables, the better the introduction to the input given speech files.

Keyword : automatic speech recognition system, acoustic model, hidden markov model, CMUSphinx toolkits, sphinx4.