

# Rancang Bangun Sistem Transkripsi Otomatis Bahasa Indonesia Bersifat Luring Dengan Input Banyak Mikrofon Menggunakan Mozilla DeepSpeech = Development of an Offline Indonesian Language Automated Transcription System with Multiple Microphone Input Using Mozilla DeepSpeech

Qisas Tazkia Hasanudin, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20526379&lokasi=lokal>

---

## Abstrak

<p>Salah satu teknologi berbasis kecerdasan buatan yang kini semakin dibutuhkan adalah ASR (<em>Automatic Speech Recognition</em>), atau lebih sering disebut sebagai <em>speech-to-text</em>. Teknologi ini memiliki potensi untuk diterapkan di berbagai bidang, salah satunya adalah mentranskripsi naskah rapat atau persidangan. Tujuan dari penelitian ini adalah untuk mengembangkan sistem transkripsi otomatis Bahasa Indonesia yang dapat berjalan secara luring dan dapat memproses masukan dari beberapa mikrofon secara bersamaan.

Penelitian ini berhasil mengembangkan sistem transkripsi otomatis dengan mengkombinasikan teknologi ASR, pemrograman Python, aplikasi <em>word editor</em> seperti Microsoft Word, dan komputer yang terhubung dengan banyak mikrofon. Teknologi ASR pada sistem ini terdiri dari <em>acoustic model</em> yang dibuat menggunakan DeepSpeech dengan metode <em>fine-tuning</em> dan <em>language model</em> yang dibuat menggunakan KenLM. Sistem transkripsi otomatis dapat dijalankan pada komputer 64-bit dengan sistem operasi Windows yang di dalamnya terdapat Microsoft Word tanpa memerlukan spesifikasi <em>hardware</em> minimum tertentu.

Hasil pengujian terhadap performa sistem menunjukkan bahwa sistem hanya bersifat <em>CPU-intensive</em>, dan ini hanya terjadi apabila seluruh pembicara berbicara pada mikrofon secara sekaligus, yang mengakibatkan tingginya jumlah <em>thread</em> yang aktif. Hasil pengujian terhadap <em>acoustic model</em> menunjukkan bahwa model tersebut dapat menghasilkan WER terbaik sebesar 73,33% dan CER terbaik sebesar 23,59% apabila dilatih menggunakan learning rate sebesar 0,01 dan dropout rate sebesar 0,3. Hasil pengujian terhadap <em>language model</em> menunjukkan bahwa model yang dibuat dengan dataset teks bertopik umum dan berukuran besar dapat membantu <em>acoustic model</em> menghasilkan WER dan CER yang lebih baik lagi, yaitu 28,76% dan 14,68%.

.....One of the artificial intelligence-based technologies that is increasingly needed is ASR (Automatic Speech Recognition), or more commonly referred to as speech-to-text. This technology has the potential to be applied in various fields, one of which is generating transcripts for meetings or trials. The purpose of this research is to develop an Indonesian automatic transcription system that can run offline and can process input from multiple microphones simultaneously.

This study succeeded in developing an automatic transcription system by combining ASR technology, Python programming, word editor applications such as Microsoft Word, and computers connected to multiple microphones. The ASR technology in this system consists of an acoustic model created using DeepSpeech with a fine-tuning method and a language model created using KenLM. The automatic transcription system can be run on 64-bit computers with Windows operating system that has Microsoft Word installed on it. It does not require certain minimum hardware specifications.

Test results on system performance show that the system is only CPU-intensive, and this only occurs when all participants are speaking into all microphones at once, resulting in a high number of active threads. The test results on the acoustic model show that the model can produce the best WER of 73.33% and the best CER of 23.59% when trained using a learning rate of 0.01 and a dropout rate of 0.3. The test results on the language model show that the model made with a text dataset that has a large size and no particular topic can help the acoustic model produce better WER and CER, which are 28.76% and 14.68%, respectively.