Abstract
An adaptive language model is an important module to develop an automatic speech recognition system. This paper focuses on developing an adaptive language model for Bahasa Indonesia. First, a corpus text of 10 million sentences is developed by crawling some Indonesian websites of news, magazines, personal blog, and writing forums. The text corpus is then used to construct an adaptive language model using Latent Dirichlet Allocation (LDA) with Collapsed Gibbs Sampling training method. The adaptive language model gives a better performance in the word selection to produce the best sentence.

Keywords: adaptive language model, Bahasa Indonesia, collapsed gibbs sampling, corpus text, latent dirichlet allocation