

## Abstract

Document searching in the Internet has special characteristic must be considered. Those are bandwidth or limited access speed and searching time spending much longer rather than desktop searching. Therefore, it needs to use indexing at Information Retrieval process that can increase speed and simply searching activities. More indexing terms mean more extra time to searching any term. It needs special methods to cut the indexing terms. One of them is document extraction with Hidden Markov Model. The method using in this extraction system is statistical approach and HMM Hedge for the HMM Model.

That method outputs results: tagging can reduce extraction and nnumber of indexed terms significantly, alpha parameter in decoding reach optimum value in 0,2 and 0,3, extraction can reduce indexing time and number of indexed terms, and corpus kinds influence extraction system accuracy.

**Kata kunci:** *extraction, Hidden Markov Model, indexing, Information Retrieval*