Abstract

Address is an important part in a record that contains information about an individual's personal or organizational, therefore the address should be available in good condition before being processed by data mining or stored into data warehouse. Reality, there is usually the address in a form that is not consistent, still in the form of free-form. To produce accurate information, the address in the form of free-form would be better if there is in the form of a consistent or have a specific format, so the sistem *will be easier to process the address data into something that has more value.*

Hidden Markov Model (HMM) used in this final project to create a sistem *that can segmenting the address, the address which had been in the free-form will have a consistent form after processed by this* sistem*. The type of HMM used is ergodic because if illustrated, the pattern must accommodate the needs of address after moving to the right then it must be re-moved to the left. Before the address being processed by the* sistem *it should be passing the data pre-processing which includes data cleaning and data tagging. To obtain the state transition probabilities and observation symbols used supervised and unsupervised learning.*

From the results of a number of observations using the model HMM which has 11 number of the state and 10 number for observation symbols, the sistem *produces the best accuracy of 93.33% on testing set with 80 training set using supervised learning.*

***Keywords** : address, segmentation, Hidden Markov Model, supervised learning, unsupervised learning*