

Abstract

Multi-document summarization is needed in order information retrieval can be done effectively dan efficiently. This summarization system uses extraction framework that has been widely used to summarize the documents, which is the results of the summarization use sentences that have been extracted from documents. Multi-document summarization can be done by looking at the information value of the sentences from document based on the features that are applied to the existing sentences. The features include the linguistic and statistical feature such as position of the sentence. Conditional Random Fields (CRF) is a probabilistic model to overcome the segmentation and labeling sequence data. CRF is used to combine the features to get a model which is used to predict the value of importance of each sentences. Feature that used are basic feature and complex feature. Accuracy of the system is measured by using F-Measure, by comparing summaries created by the system and summaries created by expert. The test results showed that the average accuracy of the system is around 62,5% with value of threshold set 0.45 where that values used to classificating summary sentences and non-summary sentences.

Keywords : *Multi-document summarization, Conditional Random Fields, Basic Feature, Complex Feature, F-Measure*