

Abstract

A variety of large amounts of information can be obtained easily from the Internet. However, it is not easy to find useful information among the many sources of information that we get. Some search engines can perform full text search, but the results are less satisfactory. Efficiency search of the number of documents that are very much very much influenced by the quality of the keywords provided by each document. Therefore, the necessary quality required keywords are retrieved automatically extracted from documents in order to obtain high efficiency values. In this Final Project, will discuss "Keyword Extraction using Conditional Random Fields". Conditional Random Fields (CRF) are probabilistic models for segmenting and labeling sequence data. CRF using different techniques in performing preprocessing in which sentences are to be segmented and labeled, so that will give effect to the results of keyword extraction. Test results showed that the extraction of keywords with CRF have a better accuracy on the document type of a general nature (general topics), or in other words having high data variation. In terms of its functions, POS features (without normalization) gives significant effect on system performance, while Len feature does not have a significant influence on system performance. In addition, training was based on the number of documents, extraction of keywords with CRF can be applied effectively even though the amount of training was a bit of data.

Keywords: *Keyword extraction, Conditional Random Fields, preprocessing, POS feature, Len feature, F-Measure.*