Abstract

Lot of information which is designed both off-line and on-line (on the internet) makes users experience difficulties in discovering the main details. Generally, those information has no framework. This causes users to find useful details based on their needs. One of the best ways to arrange those details is textual content classification which helps users in gathering the accurate and appropriate one.

Usually, classification textual content uses the same reference in organizing the words how often the terms appear in one document. Based on that condition, users can use the distributional features according to categorize the document such as compactness of the appearances and position of the first appearance of the word. The distributional features will help the users to discover its performance with a little additional calculations cost.

This study concern on some aspects such as TFIDF and LOGTF which can be combined along with the distributional features. To prove the effects the distributional features, the classic method namely k-nearest neighbor (kNN) was used.

Keywords: *distributional features*, TFIDF, LOGTF, kNN