

Abstract

Term frequency has been long used as a method of term weighting in text document. The method assumes that every term has importance value which is proportional to its frequency on document. It is the main weakness of term frequency which causes term becomes independent and disregards any dependencies that may exist between terms in the text. The problem of term frequency can be solved by applying the method of random-walk term weighting. The method is represented by graph-based ranking algorithm which is applied in textual graph that is able to integrates the dependencies of a term and its surrounding context.

The final project researches term weighting using the methods of term frequency and random-walk toward Indonesian news articles. There are two weighting schemes which will be used, *tf - rw* scheme and *tf.idf - rw.idf* scheme. Then, the datasets will be classified using Weka. The performance analysis of classification result is done by using the accuracy value and macro-average F-measure.

The experiments show that *random-walk* give better performance than *term frequency* especially on *tf.idf - rw.idf* scheme.

Key words : classification, term weighting, term frequency, random-walk, accuracy, macro-average f-measure