ABSTRACT

It is very often that an internet user wants to find some alternative websites which are similar or related with the website he or she is visiting. The process to find web pages which are related with a URL is very helpful to give more information from resulted web pages. An internet user can compare these web pages and find alternative web pages which are appropriate with his or her needs. The project explained in this paper uses link analysis approach to find related web pages. Link analysis ranks pages according to the relationship among those pages. There are two algorithms used, namely Companion and Cocitation algorithms. Both these algorithms start with creating a graph. Companion algorithm ranks web pages according to the highest authority value. Authority computation uses Imp algorithm which is based on HITS algorithm. Imp algorithm use edge weight when computing hub and authority values to solve HITS’ lack in TKC (Tightly-Knit Community). Assigning edge weights depends on host relationship among the nodes in the graph. Cocitation algorithm ranks web pages according to the highest degree of cocitation. The degree of cocitation is the number of same parents as with the web pages which are used as the query. From the test, it is proven that Companion algorithm performance is better than Cocitation algorithm. Companion algorithm works better in a graph which has many same hosts than Cocitation algorithm. The execution process of Cocitation algorithm is faster than that of Companion algorithm. Generally the URL results from these two algorithms are almost similar using Google for the comparison.

Keywords: companion, cocitation, related web pages, web structure mining, link analysis.