ABSTRACT

ADE OFIK HIDAYAT. The evaluation of Breadth First, Best First, and Page Rank algorithms on Web Crawler. Under the guidance of SONY HARTONO WIJAYA, M.Kom.

Web crawling is a process by which we gather pages from the web, in order to index them and support a search engine. The objective of crawling is to quickly and efficiently gather as many useful web pages as possible, together with the link structure that interconnects them.

The objective of this research is to evaluate Breadth First, Breadth First with Time Constraint, Best First and Page Rank algorithms. The process of collecting web pages consists of the initialization process (keywords and starting URL), inserting a link to the frontier, stopping crawling, taking the link from frontier, fetch, parsing and indexing. The focus of web crawler algorithm is determines the next link that will be visited.

Based on the precision and opportunity of keywords per algorithm, the result of the evaluation indicates that the page rank algorithm is better than three other algorithms. While based on the complexity of algorithm, the result of the evaluation indicates that the page rank algorithm has higher complexity as compared to three other algorithms. In addition, based on the average fetch time, the result of the evaluation indicates that the best first algorithm is more stable than three other algorithms.

Keywords : web crawler, web crawling, breadth first, breadth first with time constraint, best first, page rank, cosine similarity, web crawler evaluation, precision values of web crawler, complexity algorithm of web crawler.