WEBOMETRICS DASHBOARD IMPLEMENTATION USING DISTRIBUTED DOWNLOADER

Name : Ardi Imawan
NRP : 5106100059
Department : Teknik Informatika FTIf – ITS
First Supervisor : Ary Mazharuddin S., S.Kom, M.Sc
Second Supervisor : Diana Purwitasari, S. Kom, M. Sc

Abstract

Webometrics is a global university rank method that uses traces of a university academic in cyberspace as a parameter. In this thesis, will be made an application that can help the webometrics team of the university to monitor temporary rank of the university. This application is made by implementing a distributed method downloader which more than one downloader agent working in parallel on the downloader agent in this application. This method is used so that the time parameters of data retrieval process can be faster.

Distributed method requires a computer downloader to run the application manager and several computers to run the application agents. Application manager is used to generate URLs that need to be downloaded and distributed them to the computer agents in accordance with their capabilities. Communication between managers and agents are running asynchronously, so that when the manager communicates with an agent, communication with
other agents are not disturbed. An agent download a given URL manager and then retrieve a data that needed to be sent again to the manager if the destination server agents provide a successful response. After all data collected from the agents, the application’s data processing manager with a formula that represents the implementation of the method so as to form webometrics ranking of the universities listed on this application.

This application testing, found the facts that google does not allow a robot application sending an automatic query to access the server, so that Google sends out a captcha image as a response. Therefore this application is equipped with features appearances captcha image that the user can send the contents captcha manually so that the process can be resumed. Therefore, the necessary trial comparing the amount of data (URL), the number of agents, the number of times and the time required captcha. From the application test, we found the facts that more data, the more data to get, the more captcha appear and longer time required. While more and more agents are used, the less captcha that appear and the faster the time for this process. Thus we can conclude that the more the amount of data (URLs) that are downloaded then the more the agent would be required for the retrieval of data proceeded in a fast and smooth (not too many captcha that appears).

**Keywords : Webometrics, Downloader, Distributed**