Abstract Details

Title: Data Pre-Processing of Web Server Logs for Mining users Access Patterns

Authors: Mohammed Elhebir and Ajith Abraham

Abstract: Web Usage Mining (WUM) can be defined as the application of data mining techniques to extract the knowledge hidden in the Web log file, such as user access patterns from Web data in order to analyze users' behavioral patterns. However, the data stored in these log files do not present accurately a picture of the users' accesses to Web sites. The Web Usage Mining process goes through three phases: data pre-processing, patterns discovery and pattern analysis. Pre-processing is one of the most important phases in WUM which transforms a log into a set of web user sessions and makes them suitable for analysis. A sample Web log file was collected from the Web server in Sudan University of Science and Technology (SUST).This paper focuses on the pre-processing which reduces size of logs for mining user access patterns.

Keywords: Web Usage Mining, Web log data, Pre-processing, Access patterns.