World Wide Web has gained popularity because of the fact that it acts as an effective communication medium between business and end users. Company needs to have a web site which satisfies the intended needs of their end users. Users like to revisit a web site which is usable in nature. Web usage patterns of end users must be identified to improve usability on any web site. It is done with analyzing web server log files. Web logs contain noisy, redundant and incomplete data in huge volume which restricts to identify precise usage pattern from it. So, the effective data pre-processing techniques are required. In this paper algorithms are proposed and implemented for pre-processing tasks includes Data Cleaning, User identifications and Session Identification. Pre-processing algorithms are implemented on web log files of two websites and results of these algorithms are useful to study usage pattern of end users.

References


Index Terms

Computer Science  Information Sciences

Keywords

Preprocessing, Web Server log data, Web Usage Mining, Sessions, Users, Data Cleaning