Abstract

Web based applications are now increasingly becoming more popular among the users across the world. The user interactions with the applications are being tracked by the web log files that are maintained by the web server. For this purpose web usage mining (WUM) is being used. Web usage mining is the process of extracting user patterns from the web usage. In web usage mining, preprocessing plays a key role, since large amount of irrelevant information are present in the web. It is used to improve the quality and efficiency of the data. There are number of techniques available at preprocessing level of WUM. Different techniques are applied at preprocessing level such as data cleaning, data filtering, and data integration. In this paper, we present a survey on the various preprocessing techniques that have been used in order to improve the efficiency.

References

Advanced Preprocessing Techniques used in Web Mining: A Study


- Theint Theint Aye. , 2011, &quot;Web Log Cleaning For Mining Of Web Usage Patterns&quot;, 978-1-61284-2/11 © 2011 IEEE.
Keywords
   Web usage mining; log cleaning; User identification; sessionization

Index Terms
   Computer Science
   Data Mining