Abstract

This paper discusses the analysis of data from log files and presents novel methods and ideas of analyzing these log files. Huge amount of data is generated on daily basis in every IT or non-IT organization. This data is stored in logs. These logs contain data which can prove valuable. These log files can be analyzed for different reasons. This data is initially collected and then this raw data is organized in such a way that frequent patterns can be recognized from this data set. Various techniques and algorithms are used for finding such patterns, but many of them don’t prove useful on huge data sets. A number of products are available in the market. A number of algorithms have been proposed so far for mining frequent patterns. The data in these logs if used properly can prove useful in improving system performance and generating various reports on the usage of data. This information provides insights into user behaviors as well. To make this analysis easier, we need a tool which will extract the data, organize it, analyze it and generate suitable reports. Historical reports having older data also can prove vital if analyzed for drawing certain conclusions and predicting future use. Thus, our aim should be to provide with a platform-independent, low cost tool which can be affordable by smaller organizations as well. It should also be able to analyze the huge data sets generated by large organizations efficiently.
References

- Raymond T. Ng, Efficient and effective clustering methods for data mining.
- Xindong Wu, Xingquan Zhu; Gong-Qing Wu; Wei Ding. "Knowledge and Data Engineering." Sch. of Comput. Sci. & Inf. Eng., Hefei Univ. of Technol., Hefei, China, Jan 2014.

Index Terms

Computer Science

Information Science

Keywords
Clustering  User Profile  Algorithm PAM