

# Implementing Web Mining into Cloud Computing

C Mayanka

**Abstract**— In this paper different types of web mining and their usage in cloud computing (CC) have been explained. Web mining can be broadly defined as discovery and analysis of useful information from the World Wide Web [1]. It is divided into three parts that is Web Content Mining, Web Structure Mining and Web Usage Mining. Each Mining technique is used in different ways in cloud i.e. Cloud Mining.  
**Index Terms**— Cloud Computing, Data-Centric Mining, Web Mining, Web Content Mining, Web Structure Mining, Web Usage Mining.

## 1 INTRODUCTION

Web Mining is type of data mining which is used for web. Web Mining has become an emerging and an important trend because of the main reason that today storage of data has become enormous that retrieving and processing them has become an overhead. Two different approaches were taken in initially for defining Web mining they were 'process-centric view' and 'data-centric view' [2] [1]. Process centric accounted the sequence of tasks whereas data centric accounted for the types of web data that was being used in the mining process. The second approach is taken into consideration widely in recent times.

Web Mining has been categorized into three major distinct categories: Web Content Mining, Web Usage Mining and Web Structure Mining. The implementation of these categorises on World Wide Web have been well reviewed. But usage of these on the CC is newly clubbed technology.

## 2 WEB USAGE MINING

Web Usage Mining is technique of data mining which is used to retrieve the web usage information of the users. The mining is majoring used to study the behavioural pattern of the usage of web. These patterns are analysed to improve the web services given to the users.

Usage mining is done in three phase approach i.e. data preparation, pattern discovery and pattern analysis phase. [Fig. 1][3]

1. **Data Preparation/Pre-Processing of data:** The usage data is mined from the Web clients, proxy servers and servers. This data is processed identifying the users, user sessions and so on. The information is generally found in user log available in the browsers. This involves three steps i.e. Data Cleansing, Data Transformation and User/Session Identification.
  - a. **Data Cleansing:** This is process of obtaining useful information and removing the unwanted from the log data.
  - b. **Data Transformation:** Using data mining techniques like clustering, grouping together several requests. All these requests are analysed based a

session period.

- c. **User/Session Identification:** This is most difficult task in which user and session are identified from log file. This is difficult because of presence of any users on the same computer, proxy servers, dynamic addresses etc.
2. **Pattern Discovery:** This data is analysed to find out the patterns in them.
3. **Pattern Analysis:** These patterns are understood to figure out the sequence of data being accessed.

These three phases are adopted by two different approaches in usage mining. The first approach is mapping of usage data into relational tables and then using them for data mining techniques. The second approach is based on pre-processing techniques applied on the log data of the user.

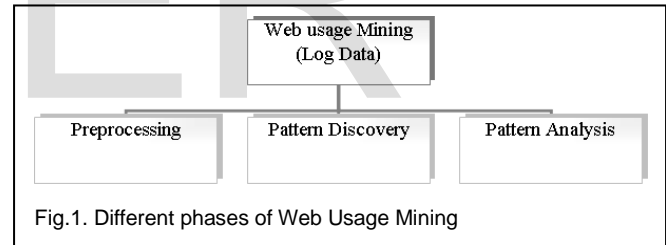


Fig.1. Different phases of Web Usage Mining

### 2.1 CC and Web Usage Mining

CC is most impressive technology because it is cost efficient and flexible. Cloud Mining's Software as Service (SaaS) is used for implementing Web Mining, as it reduces the cost and increases the security. Compared to all the other web mining techniques, Web usage mining is vastly used have proven to have given productive results.

Web Usage Mining is a major business tool for CC has become attraction in field of e-commerce, e-education, e-services etc. The following are the ways how the usage patterns are identified.

1. **HTTP Log Files:** These are recording of HTTP request and HTTP response between the user and the web server. The log file gives he details of web activity.
2. **Web Usage Data:** The usage data are stored in Web Usage Mining Warehouse, where the details like orientation, subject, integration and history are stored.

*C Mayanka is currently pursuing her final year in Integrated Master degree program in Computer Science and Technology in Women's Christian College, Chennai, India  
E-mail:mayankachandrashekar@gmail.com*

### 3 WEB CONTENT MINING

Web Content Mining is the technique of extracting the contents or the information present on the web. Content of web means all the open pages available on the web. This technique is used to find out the contents accessed by the user. The result of this technique is defined as page content mining. [4]

Web Content Mining mines out just the contents of the web without any specific grouping or pattern. For actual usage of content mining, two approaches are used based on the content present i.e. Unstructured text mining approach and Semi-Structured and Structured mining approach. [5]

- a. Unstructured text mining approach: This is also called as text data mining or text mining. The research on the mining techniques on the unstructured text data is termed as Knowledge Discovery in Texts.
- b. Semi-Structured and Structured mining approach: The structured data on the web are easier to extract compared to unstructured texts. Semi-structured is a combination of the Web dealing with documents and database communities dealing with data.

#### 3.1 CC and Web Content Mining

The web content mining mines large information on the web which is contained in regularly structured data objects. The Web data records often represent the important information on the web. For CC, web content mining is used in applications like comparative shopping, meta-search, query etc.

### 4 WEB STRUCTURE MINING

Web structure mining is analysis of hyperlinks with in the web. This is also called as Link Mining, which is a combination of link analysis (old area of research), hypertext and web mining as well as graph mining. The tasks that are possible through link mining are [6]

1. **Link-based Classification:** This task is to find out the category of the web page depending on factors like word occurrence, links and anchor texts.
2. **Link based Cluster Analysis:** The data is segmented into groups where similar ones are grouped and others are grouped into different ones.
3. **Link Type:** This predicts the existence of links, type of link as well as the purpose of link.
4. **Link Strength:** This tells about the weights of the links.
5. **Link Cardinality:** This predicts the number of links between two entities.

#### 4.1 CC and Web Structure Mining

Web Structure Mining helps in predicting the importance of the links available on web. Through this mining technique unused service on the cloud can be removed and the services which are on demand by the users can be increased. This also helps in understanding the structural usage of links on the

web, so that the service providers can place the most profitable service or most used service on the link which is accessed the most.

### 5 CONCLUSION

This paper discusses about the different types of web mining that are encountered in the data analysis on Web. The upcoming trend: The usage of Web mining in CC to manage data is also a topic of discussion here. Further research can be done by using different techniques of web mining on CC technologies for improvising the services.

### ACKNOWLEDGMENT

The author wishes to thank Women's Christian College for the opportunity provided to do research. The author also thanks Ms.Serin.J for her support and guidance, Dr.Savithri for inspiring to publish journals and all the professors of Department of Computer Science and Technology, Women's Christian College for their encouragement.

### REFERENCES

- [1] R. Cooley, J. Srivastava, and B. Mobasher. "Web Mining: Information And Pattern Discovery On The World Wide Web" In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), 1997
- [2] O. Etzioni. "The World-Wide Web: Quagmire or Gold Mine?" Communications of the ACM, 39(11):65-68, 1996.
- [3] Chhavi Rana. "A Study of Web Usage Mining Research Tools" Int. J. Advanced Networking and Applications. Volume:03 Issue: 06 Pages: 1422-1429 (2012) and References there in.
- [4] Aishwarya Rastogi et al. "Web Mining: A Comparative Study" International Journal of Computational Engineering Research (Mar-Apr 2012) Volume: 2. Issue: 2 Pages: 325-331.
- [5] Abdelhakim Herrouz et al. "Overview of Web Content Mining Tools" The International Journal of Engineering and Science (IJES) (2013) Volume: 2 Issue: 6 and References there in.
- [6] Miguel Gomes da Costa Júnior and Zhiguo Gong. "Web Structure Mining: An Introduction" Proceedings of IEEE International Conference on Information Acquisition (2005) Page: 590- 595 and References there in