

## CHAPTER: 13

# KNOWLEDGE DISCOVERY PROCESS FOR WEB ANALYSIS

**<sup>1</sup>Dr. SUDHAKAR RANJAN**

<sup>1</sup>School of engineering and technology  
Apeejay stya university Sohna-Palwal road, Gurugram

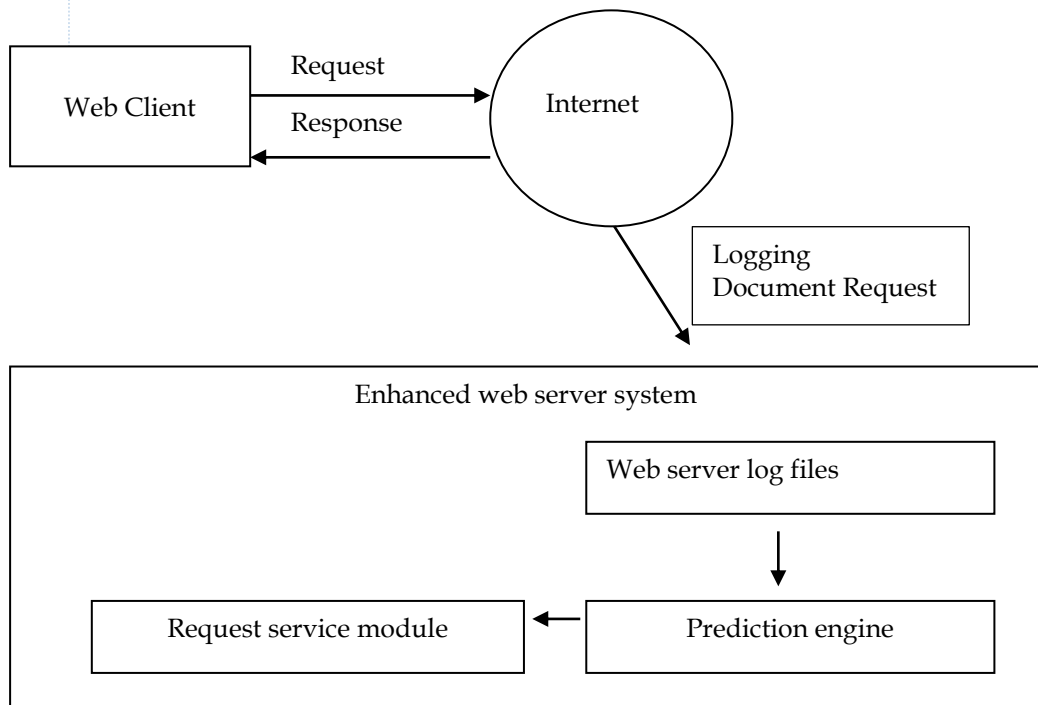
**<sup>2</sup>Dr. PARIKSHIT VASISHT**

<sup>2</sup>School of engineering and technology  
Apeejay stya university Sohna-Palwal road, Gurugram

## INTRODUCTION

Knowledge discovery process is a methodology to study, analysis data and finding useful information and pattern in data. The some common methods are visualization techniques and rule based systems. The critical steps in web server system are Web server log files, Prediction engine, Request service system, and web client shown in figure 1. The following types of types of data are available like

- Gather Information, Group Information, Generate Information and Give information to others



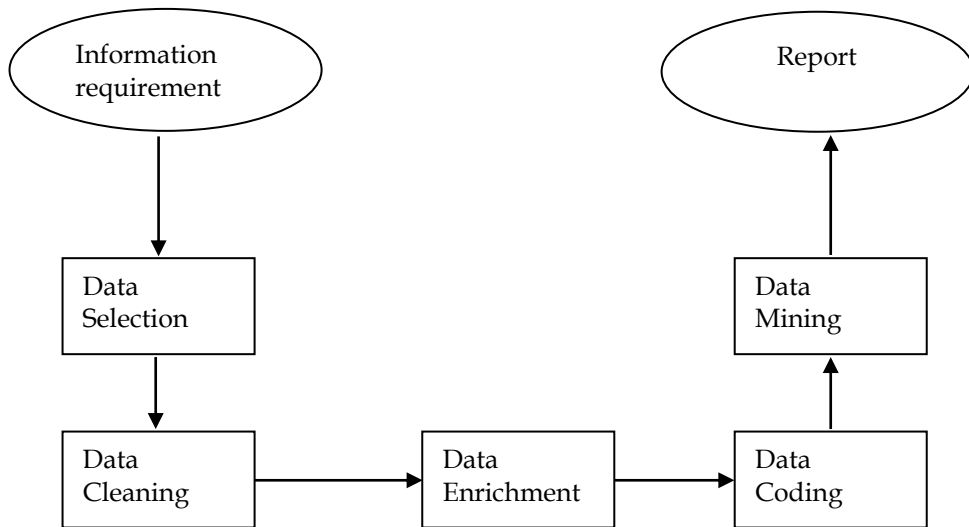
**Fig 13.1: Web server systems**

In general, a knowledge discovery process consists of an iterative sequence of the following steps and the task of data mining functions are given below:

- Classification
- Regression
- Time series Analysis

- Prediction
- Clustering
- Summarization
- Association Rules
- Sequence Discovery

The terms knowledge discovery in database and data mining are often used interchangeably.



**Fig 13.2: Position of Data Mining in the Knowledge Discovery process**

The KDD process consists of the five steps:

- Data selection
- Data cleaning.
- Data integration (*enrichment*)
- Data transformation (*coding*)
- Data mining.
- Knowledge representation

Visualization refers to the visual presentation of data. Visualization techniques include:

- Graphical
- Geometric
- Icon-Based
- Pixel Based
- Hierarchical
- Hybrid

### Data Mining Goals, Operations and Techniques

In general, data mining tasks can be classified into two categories:

- **Description:** The purpose of descriptive model identifies patterns, association or relation of the data.
- **Prediction:** The Categories of prediction as well as description are associated with the six basic operations, as presented in Figure 3. A predictive model makes a predictive about values of data using known result.

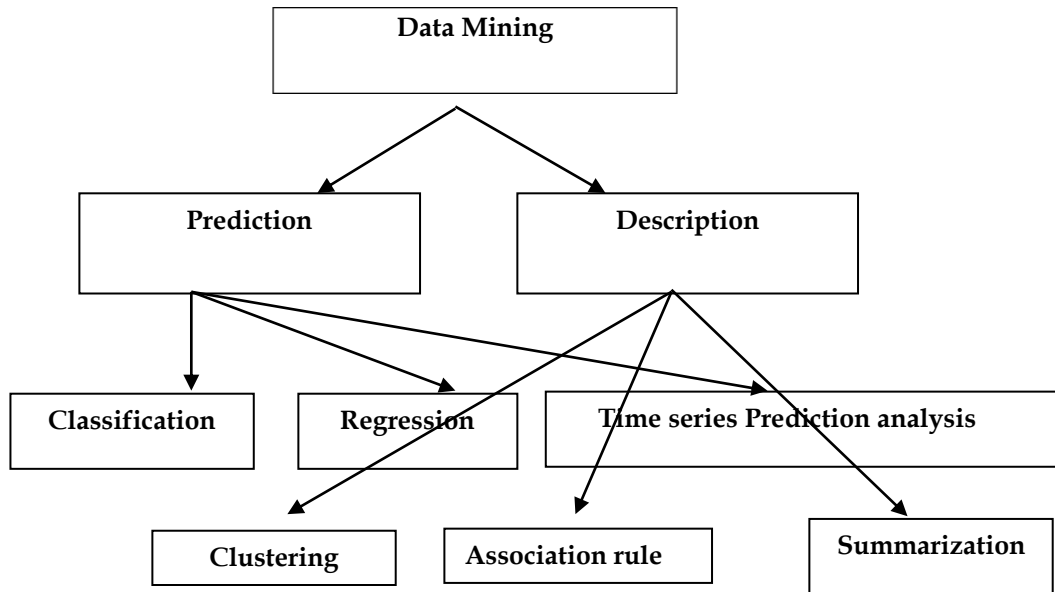


Fig 13.3: The connection between data mining goals and operation

### Access Logs Processing Architecture

A general architecture for web access mining (see Figure 4) using the site's Web server logs as data source.

#### Site File

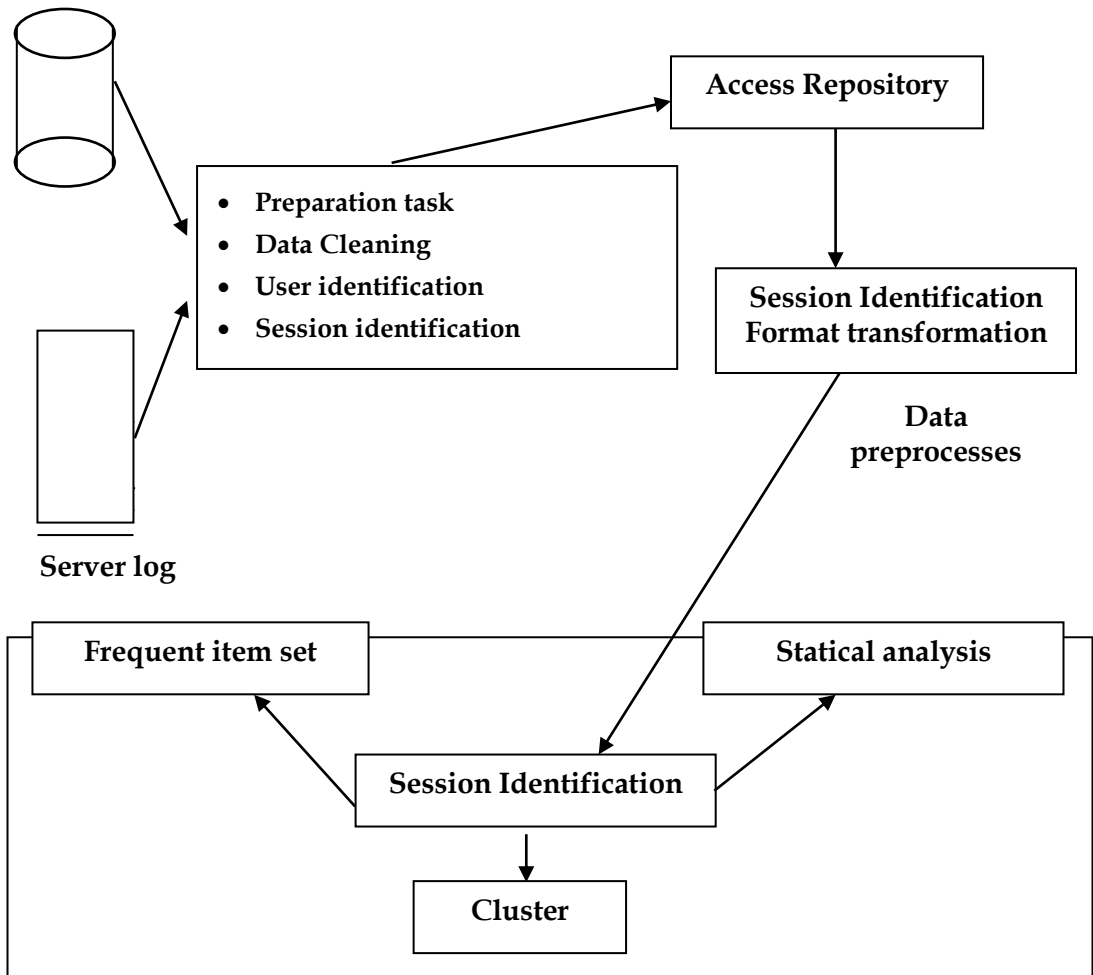


Fig 13.4: Web Access Pattern

## CONCLUSION

A major goal of KDD is to be able to describe the results of the KDD process in meaningful manner. The Web and its techniques and volume of the data increasing rapidly, the users have less time to retrieve relevant information.

## REFERENCES

1. Navin kumar,arya gangopadhyay,Discovery valuable pattern through internet web log access analysis.
2. Cooley R., Mobasher B., and Srivastava J., Data preparation for mining World Wide Web browsing patterns , Knowledge and Information Systems, V1(1), 1999
3. Bartoloni G., Web usage mining and discovery of association rules from HTTPserverlogs,2001 <http://www.prato.linux.it/~gbartolini/html/wum.html>
4. Jiawei Han, Man Xin, Osmar Zaiane, Discovering web access patterns and trends by applying OLAP and data mining technology on web logs, In Proceedings of the Fifth IEEE Forum on Research and Technology Advances in Digital Libraries, 1998.
5. White paper on Speed Tracer: A Web usage mining tool (<http://www.research.ibm.com/journal/sj/371/wu.html>)
6. Bamshad Mobasher, Namit Jain, Eui-Hong Han, Jaideep Srivastava (1996). Web mining: pattern discovery from World Wide Web Transactions.
7. Robert Cooley, Bamshad Mobasher, Jaideep Srivastava(1997). Grouping Web page references into transactions for mining World Wide Web browsing patterns.
8. Robert Cooley, Bamshad Mobasher, Jaideep Srivastava (1997). Web Mining: information and pattern discovery on the World Wide Web.
9. Masegla, P. Poncelet, M. Teisseire (1999). Using data mining techniques on Web access logs to dynamically improve Hypertext structure.
10. Robert Cooley, Bamshad Mobasher, Jaideep Srivastava (1999). Data preparation for mining World Wide Web browsing patterns.
11. Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang- Ning Tan (2000). Web Usage Mining: Discovery and applications of usage patterns from Web data.

12. Sudhakar Ranjan, Komal Kumar Bhatia "Query interface integrator for domain Specific Hidden web" *International Journal of Computer Engineering and Applications (IJCEA)* ISSN 2321-3469, Vol. IV, Issue I/III, Oct.-Dec. 2013.
13. Sudhakar Ranjan, Komal Kumar Bhatia "Indexing for Vertical Search Engine: Cost Sensitive" *International Journal of Emerging Technology & Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal)*, Volume 3, Issue 10, October 2013.
14. .Sudhakar Ranjan, Komal Kumar Bhatia "Web Log for Domain Specific Hidden Web", *International Journal of Computer Engineering and Applications (IJCEA)* ISSN 2321-3469, Vol VII, Issue I, July 2014.
15. Sudhakar Ranjan, Komal Kumar Bhatia "Indexing for Domain Specific Hidden Web" *International Journal of Computer Engineering and Applications (IJCEA)* 2321-3469, Vol VII, Issue I, July 2014.
16. Sudhakar Ranjan, Komal Kumar Bhatia "Transaction in Hidden Web" *International Journal of Computer Engineering and Applications (IJCEA)* 2321-3469, Vol VIII, Issue II, Nov 2014.
17. .Sudhakar Ranjan, Komal Bhatia: Design a Least Cost Vertical Search Engine based on DSHWC" selected for publication in *IJIRR* ,2017
18. Sudhakar Ranjan, Sarim Moin, Parikshit Vasisht, Abdus Samad Moin Uddin, "10 Role of Cyber-Security in Smart Energy Management Systems", *Smart Energy Management Systems and Renewable Energy Resources*, AIP publishing, Sepetember 2021 [https://doi.org/10.1063/9780735422827\\_010](https://doi.org/10.1063/9780735422827_010)