Preface

The World Wide Web has variety of information service centers, like news, sites, encyclopedias, education sites, e-commerce etc. Since this massive utility of web resources in recent scenario has turned to be an indispensable commitment for numerous reasons. The inconceivable boom of information available in the websites simultaneously throws the challenge of retrieving the precise and appropriate information at the time of need. Moreover, the web information is the mostly sought after powerful platform for working, studying, searching information, besides, being in touch with our friends. Apparently, the unpredictable amount of web information available becomes a menace of experiencing ambiguity in the web search. In order to successfully retain users in this rapidly developing environments, a website must be built in such a way that supports user personalization. To achieve this, an organization can keep track of user activities while browsing their websites. Besides the challenge to find relevant information, users could also find other difficulties when interacting with the web such as the degree of quality of the information found, the creation of new knowledge out of the information available (Lin & Huang, 2003).

Web Usage Mining is one of the fastest developing areas of web mining (Cooley, Mobeshar & Srivastava, 1997). Web usage mining is the application of data mining technologies to discover interesting usage patterns from web data, in order to understand and better serve the needs of web-based applications. Its attention in analyzing users behavior on the web after exploring access logs made its popularity very rapidly especially in E-services. Its direct application in these areas added its admiration and made it as an inevitable part in computer and information sciences (Lieherman & Letizia, 1995). Web usage mining provides the support for website design, personalization server, other business making decision etc.

Web Usage Mining is an aspect of data mining that has received a lot of attention in recent years. Commercial companies as well as academic researchers have developed an extension array of tools that perform several data mining algorithms on log files coming from web servers in order to identify user behavior on a particular website. Details like user log files, request for resources etc., are maintained in web servers, which is the core mining area of web usage. The analysis of these gives the user browsing patterns and that can be utilized for target advertisement, enhancement of web design, satisfaction of customers and making market analysis. Most of the e-service providers realized the fact that they can apply this tool to retain their customers (Fengrong, 2004).

Web Usage Mining provides better understanding for service the needs of web-based applications (Schafer, Konstan & Reidl, 2001). Site modification, business intelligence, system improvement, personalization and usage characterization are the areas in which the potentials of web usage mining have been recognized and extensively used.

This book, Web Usage Mining Techniques and Applications Across Industries, shall provide the support for web site design, personalization server, other business making decision etc. It focuses on

the systems, techniques and applications that could predict user's behavior while the user interacts with the web. This book shall offer significant contribution towards web information retrieval to the professionals for further research opportunities in this field. This book will provide the application of data mining techniques to the usage logs of large web data repositories in order to produce results that can be applied to many practical subjects, such as improving websites/pages making additional topic or product recommendations, user/customer behavior studies, etc.

The audience of this book will widely vary from individuals, researchers, scientists, academics, students, libraries, journalists and development practitioners. This book discovers the navigation patterns of the surfers from the web data and deals with the prediction of the surfer's behavior and interaction with the web which will be useful in furthering their research exposure.

ORGANIZATION OF CHAPTERS

The book has been divided into three sections; Concepts and Categorization, Applications and Analytics, and Methodologies and Technologies. Altogether there are fourteen manuscripts covering wider range of concepts, technologies and applications.

Chapter 1 aims to master web mining and Information Retrieval (IR) in the digital age, thus describing the overviews of web mining and web usage mining; the significance of web mining in the digital age; the overview of IR; the concept of Collaborative Information Retrieval (CIR); the evaluation of IR systems; and the significance of IR in the digital age. The chapter argues that applying web mining and IR has the potential to enhance organizational performance and reach strategic goals in the digital age.

Chapter 2 proposes a Markov chain based method to categorize the users into faithful, partially Impatient and completely Impatient user. And further, their browsing behavior is analyzed. We also derived some theorems to study the browsing behavior of each user type and then some numerical illustrations are added to show how their behavior differs as per categorization. Also the chapter extends the work by approximating the theorems.

Chapter 3 illustrates folksonomy based information retrieval by generating tag cloud. This model not only helps the industries to manage their electronic resources for retrieval but helps them by providing suggestions for tagging with the usage of similarity metrics. This suggestive mechanism also helps users to understand resources at specific and organizations at general. The authors also have implemented the model to demonstrate the experimental results followed by discussion.

Chapter 4 deals Web mining, Categories of Web mining, Web usage mining and its process, Applications of Web usage mining across the industries and its related works. This Chapter offers a general knowledge about Web usage mining and its applications for the benefits of researchers those performing research activities in WUM.

Chapter 5 compares the three factor based techniques viz. principal component regression (PCR), Generalized Least Square (GLS) Regression, and Maximum Likelihood Regression (MLR) method and explores their predictive ability on theoretical as well as on experimental basis. All the three factor based techniques have been compared using the necessary conditions for forecasting like R-Square, Adjusted R-Square, F-Test, JB (Jarque-Bera) test of normality. This study can be further explored and enhanced using sufficient conditions for forecasting like Theil's Inequality Coefficient (TIC), and Janur Quotient (JQ).

Chapter 6 explains the concepts of data mining and discusses at length about the landslide event. Further, the utility of data mining techniques in disaster management using a previous work was explained

and provides a brief note on the efficiency of web mining in creating awareness about natural hazard by providing refined information. Finally, a conceptual framework for landslide hazard assessment using data mining techniques such as Artificial Neural Network (ANN), Fuzzy Geometric Mean Model (FGMM), etc. were chosen for description.

Chapter 7 provides the details for a high quality website which is the one that provides relevant, useful content and a good user experience. The chapter also discusses all areas of website are thoroughly studied for analysing the quality of website. design.

Chapter 8 describes Query Recommendation, a technique to provide the alternate queries as a substitute of the input query to the user to frame the queries in future. The chapter also discusses about the methodology framed to identify the similar queries and they are clustered, which contains the similar queries which are used to provide the recommendations.

Chapter 9 presents the details about the kind of data getting generated through E-Governance initiative in India, which will help in opening up lot of opportunities for data analysts and mining experts to explore this data and generate insights out of them. The chapter also introduces areas in India where analytics can be applied for E-Governance related entities - citizens, corporate and government departments.

Chapter 10 discusses about the explosive growth in the amount of data in the field of biology, education, environmental research, sensor network, stock market, weather forecasting and many more due to vast use of internet in distributed environment has generated an urgent need for new techniques and tools that can intelligently automatically transform the processed data into useful information and knowledge. The chapter presents the new dynamic and scalable data mining approach has been discussed with educational data.

Chapter 11 explains in detail about the methodology of web usage mining which are data collection, data preprocessing, knowledge discovery and pattern analysis. The different Web Usage Mining techniques are described, which are used for knowledge and pattern discovery. These are statistical analysis, sequential patterns, classification, association rule mining, clustering, dependency modeling.

Chapter 12 demonstrates visual resolution retrieval with the help of an experiment using Discrete Wavelet Transform along with the discussion of various multiresolution techniques for visual information retrieval. The experiment explained in the chapter helps in explaining the important properties of multiresolution analysis and also provides future scope of research in this field.

Chapter 13 addresses the problem of how to support web usage mining techniques and applications across industries by combining language of web pages and algorithms that used in web data mining. Existing research in web usage mining techniques tend to focus on finding out how each techniques can apply in different industries fields. However, there is little evidence that researchers have approached the issue of web usage mining across industries. Consequently, the aim of this chapter is to provide an overview of how the web usage mining techniques and applications across industries can be supported.

Chapter 14 focuses on various web data extraction techniques available for different kinds of data rich pages, classification of web data extraction techniques and comparison of those techniques across many useful dimensions.

CONCLUSION

Web Usage Mining is becoming an active interesting field of research because of its prospective commercial benefits. Its attention in analyzing users behavior on the web after exploring access logs made

its popularity very rapidly. Web usage mining has many benefits which attract business and government agencies towards it. Government agencies utilized the classification and predicting capability of this technology to fight against terrorism and identifying criminal activities. Business sectors are benefited by personalized marketing, customer retention, and customer relationship, and even they got the opportunity to provide promotional offers to specific customers to retain them (Lee, LungLo & Fu, 2011).

A. V. Senthil Kumar
Hindusthan College of Arts and Science, India

REFERENCES

Cooley, R., Mohensher, B., & Srivastava, J. (1997). Web mining information and pattern discovery on the world wide web. In *Proceedings of the 9th IEEE International Conference on Tools and Artificial Intelligence.*IEEE. doi:10.1109/TAI.1997.632303

Fengrong, J. (2004). *Study of web usage mining and discovery of browse interest.* (Thesis). Beijing Science and Technology University, Beijing, China.

Lee, C. H., Lo, Y. L., & Fu, Y. H. (2011). *A novel production model based on hierarchical characteristic of website*. Elsevier.

Lieherman, R., & Letizia. (1995). An agent that assists web browsing. In *Proceedings of the 1995 International Joint Conference on Artificial Intelligence.* Montreal, Canada: IEEE.

Lin, J. G., & Huang, H. H. (2003). Web mining for electronics business application. In *Proceedings of the Fourth International Conference on Parallel and Distribution Computing, Application and Techniques* (pp. 872-876). Academic Press.

Schafer, J. B., Konstan, J. A., & Reidl, R. (2001). E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, *5*(5), 115–153. doi:10.1023/A:1009804230409