# WEB USAGE MINING FOR UUM LEARNING CARE USING ASSOCIATION RULES

A project submitted to the Graduate School in partial fulfillment of the requirements for the degree Master of Science (Intelligent System)
Universiti Utara Malaysia

By:
Azizul Azhar bin Ramli

# PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirement for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor or, in her absence, by the dean of the Graduate School. It is also understood that due recognition shall be given to me and Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or make other use of materials in this thesis, in whole part, should be addressed to:

**Dean of Graduate School,**
**Universiti Utara Malaysia**
**06010 UUM Sintok**
**Kedah Darul Aman**
**Malaysia**

# ABSTRAK

Ledakan maklumat di dalam gerbang Web menjadikannya pilihan terbaik bagi penyelidikan perlombongan data. Aplikasi bagi teknik perlombongan data bagi gerbang Web dirujuk sebagai perlombongan Web dan telah digunakan dalam tiga pendekatan yang berbeza; Perlombongan Isi Web, Perlombongan Struktur Web dan Perlombongan Penggunaan Web. Pembelajaran Elektronik adalah salah satu aplikasi di dalam gerbang Web dimana ianya akan berhadapan dengan jumlah data yang sangat besar. Bagi menghasilkan corak penggunaan portal dan amalan pengguna, kajian ini akan mengimplimentasikan proses aras tinggi bagi perlombongan penggunaan Web menggunakan algoritma asas bagi Peraturan Kesatuan – algoritma *Apriori*. Perlombongan penggunaan Web mengandungi tiga fasa utama iaitu Preprosesan Data, Penjelajahan Corak dan Analisis Corak. Sumber utama iaitu data mentah adalah terdiri daripada fail-fail log pelayan dan ianya perlu melalui fasa-fasa dalam perlombongan penggunaan Web bagi menghasilkan keputusan akhir – set peraturan. Dengan keupayaan teknik perlombongan data, pendekatan perlombongan penggunaan Web telah digabungkan dengan asas Peraturan Kesatuan, algoritma *Apriori* bagi mengoptimumkan kandungan portal Pembelajaran Elektronik universiti. Akhir sekali, kajian ini akan membentangkan keputusan serta analisisnya supaya pihak pengurusan Web boleh menggunakan segala keputusan tersebut untuk tindakan bernilai yang sewajarnya.

# ABSRACT

*The enormous content of information on the World Wide Web makes it obvious candidate for data mining research. Application of data mining techniques to the World Wide Web referred as Web mining where this term has been used in three distinct ways; Web Content Mining, Web Structure Mining and Web Usage Mining. E-Learning is one of the Web based application where it will facing with large amount of data. In order to produce the university E-Learning (UUM Educare) portal usage patterns and user behaviors, this paper implements the high level process of Web usage mining using basic Association Rules algorithm – Apriori Algorithm. Web usage mining consists of three main phases, namely Data Preprocessing, Pattern Discovering and Pattern Analysis. Main resources, server log files become a set of raw data where it's must go through with all the Web usage mining phases to produce the final results – set of rules. With the powerful of data mining technique, Web usage mining approach has been combined with the basic Association Rules, Apriori Algorithm to optimize the content of the university E-Learning portal. Finally, this paper will present an overview of results with the analysis and Web administrator can use the findings for the suitable valuable actions.*

# ACKNOWLEDGEMENTS

First of all, I would like to express my appreciation to Allah, the Most Merciful whom granted me the ability and willing to start and complete this project. I pray to his greatness to inspire and to enable me to continue the work for benefits of my religion, Islam and country.

I would also like to express my gratitude to my supervisor, Encik. Mohd. Shamrie bin Sainin, lecturer of the Artificial Intelligent Department of Information Technology Faculty at the Universiti Utara Malaysia for her excellent guidance and advice through completing this project. Many special thanks to the Information System Officer from UUM Computer Centre, Cik Roslina Hanafiah who honestly prepared and provided the required data for this project, others lecturers with their informal guidance and friends for their helps and supports. Especially for my beloved family and my dearest one whose encourage me much, throughout this project. Thanks for every never-endings support and kindness. May Allah bless us, InsyaAllah.

Only God knows everything!
Thank you.

**AZIZUL AZHAR BIN RAMLI**
**Artificial Intelligent Department,**
**Information Technology Faculty, Universiti Utara Malaysia**

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

This chapter will discuss about the project background that contains the general overview of the project includes the brief description about the data mining technologies and Web usage mining as a part of Web mining approach. In addition, the project background sub chapters also describe the tools and software that was used for this project and also the approach that is used for the analysis purposes. The description of the project problem statement, lists of the project objectives and details of project scope and boundaries are also discuses in this chapter. Finally, the thesis organization that contains the structure of chapters that is included in this report.

## 1.1 Project Background

Finding useful patterns in data is known by different names (including data mining) in different communities (e.g., knowledge extraction, information discovery, information harvesting, data archeology and data pattern processing) (Fayyad *et al.*, 1996). Data mining is a burgeoning and promising research field in computer science field. Data mining is a technique used to deduce useful and relevant information to guide professional decisions and other scientific research (Chen, Han and Yu, 1996). The objective of data mining is to identify valid novel, potentially useful and understandable correlations and patterns in existing data (Chung and Gray, 1999). It is a cost-effective

way of analyzing large amounts of data, especially when a human could not analyze such datasets.

The growth of the use of the Internet has made automatic knowledge extraction from Web log files a necessity. Information providers are interested in techniques that could learn Web users' information needs and preferences. This can improve the effectiveness of their Web sites by adapting the information structure of the sites to the users' behavior. These factors give rise to the necessity of creating server-side and client-side intelligent systems that can effectively mine for knowledge both across the internet and in particular Web localities. However, it is hard to find appropriate tools for analyzing raw Web log data to retrieve significant and useful information.

Two situations above can be compiled as a one problem with the best alternative solution where Web mining approach that using data mining techniques to automatically extract and discover the hidden information from Web server. Web mining can be broadly defined as the discovery and analysis of useful information from the World Wide Web. This broad definition on the one hand describes the automatic search and retrieval of information and resources available from millions of sites and on line databases. A meta-analysis of the Web mining literature, categorized Web mining into three areas of interest based on which part of the Web is to be mined (Kosala and Blockeel, 2000; Srivastava *et al.*, 2000).

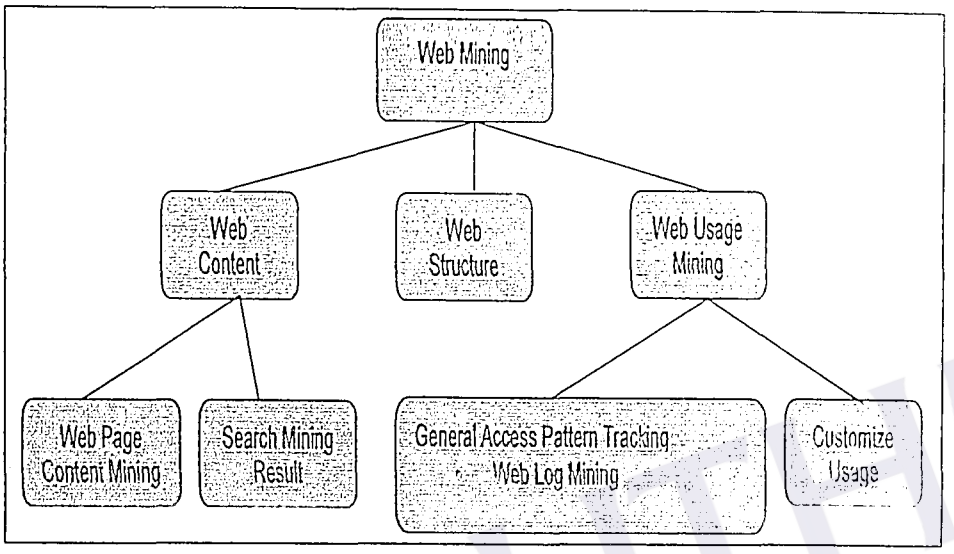


**Figure 1.1:** Taxonomy of Web Mining

Recently, the advent of data mining techniques for discovering usage pattern from Web data (Web usage mining) indicates that these techniques can be a viable alternative to traditional decision making tools (Srivastava *et al.*, 2000). Web usage mining is the process of applying data mining techniques to the discovery of usage patterns from Web data and is targeted towards applications (Srivastava *et al.*, 2000). Web usage mining mines the secondary data (Web server access logs, browser logs, user profiles, registration data, user sessions or transactions, cookies, user queries, mouse clicks and any other data as the result of interaction with the Web) derived from the interactions of the users during Web sessions.

This project will describe the use of Web usage mining techniques to analyze Web log records collected from E-Learning portal (UUM Educare) by applying Association Rules (AR) technique. AR is a statement about relationships between the attributes of a known group of entities and one or more aspects of those entities that enable predictions to be made about aspects of other entities who are not in the group, but who possess the same attributes.

Using commercial Web mining tools (*WebLog Expert 3.0 and Sawmill 6*) and *ARunner 1.0* (prototype of GUI Christian Borgelt *Apriori* tool by Mohd. Shamrie Sainin, FTM, UUM), it have identified several Web access pattern by applying well known data mining techniques (*Apriori* algorithm) to the access logs of this educational portal. This includes descriptive statistical analysis and Association Rules for the portal including support and confidence to represent the Web usage and user behavior for UUM Educare. The findings and results of this experimental analysis can be use by the Web administration and may be upper level in the UUM community in order to improve Web services and performance through the improvement of Web sites, including their contents, structure, presentation and delivery.

Web usage mining can help in addressing some of the shortcomings of the standard approaches for Web personalization. However, the discovery of patterns from usage data is not by itself sufficient for performing the personalization tasks. Nevertheless, it is an important component for effective derivation of "actionable" user profiles derived from Web access patterns. The learned knowledge can also be used for other applications, such as improving site usability, business intelligence and usage characterization.

## 1.2 Problem Statement

With the explosive growth of information sources that available on World Wide Web (WWW), it has become increasingly necessary for Web administrator to utilize automated tools to find the desired information resources and to track and analyze their Web usage pattern. For the university E-Learning portal (UUM Educare), normally there is about 1000 to 1500 of users including lecturers and students that accessing this site portal within certain period of time. The complete data for this particular information are placed in sever log file and Web administrator is the person who are responsible for monitoring the task. Without any specific tools in order to assist this complex task, Web administrator cannot easily analyze the pattern of Web site usage for certain valuable action.

Web usage mining taxonomy Web mining tree is a one of the Web mining approach. With this approach, Web administrator that facing with a large size of log files will easily extract all the required data, preprocessing the data and generate the Web usage patterns of university E-Learning portal. The final results that produced in the implementation of this Web usage mining process will become the best inputs for improvement, enhancement and maintaining the performance of university E-Learning (UUM Educare) portal as a main online resource for education purposes.

## 1.3 Project Objectives

The general objective of this project is to perform World Wide Web mining with the main task of ***Web usage mining - process mining of the user browsing and access pattern***. Specifically, the objectives of this project are:

i. To preprocess data for UUM Educare server logs files in order to determine and discover the user access pattern from university E-Learning Web servers (UUM Educare).

ii. To perform the basic Association Rules – *Apriori* algorithm for implementation the Web usage mining process for producing usage pattern in order to determine the most user interest sites based on the options that being provided by the university E-Learning portal (UUM Educare).

iii. To analyze the usage patterns output and user behaviors for UUM Educare from the Web usage mining implementation process.

## 1.4 Project Significant

Generally, this project will produce the useful guideline for analyzing the Web usage pattern for E-Learning and more specific:

i. This study will become the first step for the analyzing university E-Learning portal by implementing Web usage mining approach using with Association Rules – *Apriori* algorithm.

ii. The outcomes from this study can be used by the Web administrator to plan necessary improvement and enhancement of the Web services and performance through the improvement of Web sites, including their contents, structure, presentation and delivery to the university E-Learning portal.

iii. The implementation of Web usage mining process for university E-Learning portal may becomes the guideline for the system development purposes.

## 1.5 Project Scope and Boundaries

Basically this project is done by extracting the server log files from the E-Learning domain server host, www.e-web.uum.edu.my that provided for the university students. The E-Learning portal consists of learning material such as complete lecture notes, tutorials question, quizzes, assignment, mid semester and also final examination questions. Beside that, the E-Learning portal also provides certain important facilities such as work schedule, email and forum. The results for the particular subjects also will be placed on the university E-Learning portal. Therefore, this project is focusing on the usage of E-Learning portal. The details of project scopes are:

| | | |
|---|---|---|
| i. | **Research Organization** | : University E-Learning portal (UUM Educare). |
| ii. | **Focus of Project** | : Extract the server log file from university E-Learning server on certain of weeks within a semester, preprocessing the set of raw data, select the data that contribute for pattern analysis, implement the pattern mining using basic association rule, *Apriori* algorithm, in order to produce the final results (*rules*). |

## 1.6 Thesis Organization

As a forerunner of this project specified in Web usage mining study for E-Learning portal (UUM Educare), the paper is organized as following topic:

i. Chapter 1 : INTRODUCTION
ii. Chapter 2 : LITERATURE REVIEW
iii. Chapter 3 : SERVER LOG FILES
iv. Chapter 4 : PROJECT METHODOLOGY AND IMPLEMENTATION
v. Chapter 5 : FINDINGS AND RESULTS
vi. Chapter 6 : RECOMMENDATION AND CONCLUSION

# CHAPTER 2

# LITERATURE REVIEW

This chapter will describe the details about related literature review of this project. Literature review chapter start with a discussion about the data mining technologies. It follows by Web mining, a part of data mining technologies and Web usage mining as a type of Web mining technologies. The Web usage mining sub section exclusively explains details about the high level process for this Web usage mining approach includes the important of each particular phase and also their sub tasks. The basic description of server log files and Web server log files analysis also discussed. The most important sub chapter is about the Association Rules and its popular technique which is *Apriori* algorithm. In the end of this chapter, implementation of Web usage mining in the educational environment is explained.

## 2.1 Data Mining

The objective of data mining is to identify valid novel, potentially useful, and understandable correlations and patterns in existing data (Chung and Gray, 1999). Finding useful patterns in data is known by different names (including data mining) in different communities (e.g., knowledge extraction, information discovery, information harvesting, data archeology, and data pattern processing) (Fayyad *et al.*, 1996). The term "data mining" is primarily used by statisticians, database researchers and the MIS and

business communities. The term Knowledge Discovery in Databases (KDD) is generally used to refer to the overall process of discovering useful knowledge from data, where data mining is a particular step in this process (Fayyad *et al.*, 1996; Peacock, 1998; Han and Kamber, 2001). The additional steps in the KDD process, such as data preparation, data selection, data cleaning and proper interpretation of the results of the data mining process, ensure that useful knowledge is derived from the data.

Data mining is an extension of traditional data analysis and statistical approaches in that it incorporates analytical techniques drawn from a range of disciplines including, but not limited to:

- Numerical analysis,
- Pattern matching and areas of artificial intelligence such as machine E-Learning,
- Neural networks and genetic algorithms.

While many data mining tasks follow a traditional, hypothesis-driven data analysis approach, it is common place to employ an opportunistic, data driven approach that encourages the pattern detection algorithms to find useful trends, patterns, and relationships.

Essentially, the two types of data mining approaches differ in whether they seek to build models or to find patterns. The first approach, concerned with building models is, apart from the problems inherent from the large sizes of the data sets, similar to conventional exploratory statistical methods. The objective is to produce an overall summary of a set of data to identify and describe the main features of the shape of the distribution (Hand, 1998). Examples of such models include a cluster analysis partition of a set of data, a regression model for prediction, and a tree-based classification rule. In model building, a distinction is sometimes made between empirical and mechanistic models (Hand, 1998). The former (also sometimes called operational) seeks to model relationships without basing them on any underlying theory. The latter (sometimes called substantive or phenomenological) are based on some theory or mechanism for the underlying data

generating process. Data mining, almost by definition, is primarily concerned with the operational task.

The second type of data mining approach, pattern detection, seeks to identify small (but nonetheless possibly important) departures from the norm, to detect unusual patterns of behavior. Examples include unusual spending patterns in credit card usage (for fraud detection), sporadic waveforms in EEG traces and objects with patterns of characteristics unlike others. It is this class of strategies that led to the notion of data mining as seeking "nuggets" of information among the mass of data. In general, business databases pose a unique problem for pattern extraction because of their complexity. Complexity arises from anomalies such as discontinuity, noise, ambiguity and incompleteness (Fayyad, Piatetsky-Shapiro and Smyth, 1996). And while most data mining algorithms are able to separate the effects of such irrelevant attributes in determining the actual pattern, the predictive power of the mining algorithms may decrease as the number of these anomalies increase (Rajagopalan and Krovi, 2002).

## 2.2 Web Mining

Data mining efforts associated with the Web, called *Web mining*, can be broadly categorized into three areas of interest based on which part of the Web to mine; Web content mining, Web structure mining, and Web usage mining (Kosala and Blockeel, 2000). Web content mining focuses on techniques for searching the Web for documents whose content meets Web user's queries. Web structure mining is related to the analysis of the link structure of the Web, in order to identify relevant documents. Web usage mining is defined as the process of applying data mining techniques to the discovery of usage patterns from Web logs data, to identify Web user's behavior (Srivastava *et al.*, 2000).

In Web mining, data can be collected at the server-side, client-side, proxy servers or a consolidated Web/business database. Srivastava *et al.*, (2000), presents a more detailed description of these data sources. To summarize, (i) Web server logs explicitly records browsing behavior of site visitors, (ii) Client-side data collection can be implemented by using a remote agent or by modifying the source code of an existing browser (iii) and Web proxies act as an intermediate level of caching between client browsers and Web servers.

The information provided by the data sources described above can be used to construct several data abstractions, namely users, page-views, click-streams, and server sessions. A *user* is defined as a single individual that is accessing the Web servers through a browser. In practice, it is very difficult to uniquely and repeatedly identify users. A user may access the Web through different machines or use more than one browser at one time. A *page-view* consists of every length that contributes to the display on a user's browser at one time and is usually associated with a single user action such as a mouse-click. A *click-stream* is a sequential series of page views requests. Note that any page view accessed through a client or proxy level cache will not be recorded on the server side. A *server session* (or *visit*) is the click-stream for a single user for a particular Web site. The end of a server session is defined as the point when the user's browsing session at that site has ended.

The process of Web usage mining can be divided into three phases: Data Preprocessing, Pattern Discovery and Pattern Analysis (Srivastava *et al.*, 2000). The details about Web usage mining is explained on the next section.

## 2.3 Web Usage Mining

The earlier usage of a Web site was measured by the number of hits on the Website. But recent research has proved that the metric of number of hits doesn't reveal any profitable information. The analysis using hits is s.milar to estimating the quality of music by the loudness.

The term Web usage mining was discovered by (Cooley *et al.*, 1997), when the first attempt of taxonomy of Web mining was done. Web usage mining focuses on techniques that could predict user behavior while user interacts with the Web. Usage Data for Web usage mining can be obtained from the Web clients, proxy servers and servers (Srivastava, Cooley, Deshpande and Tan, 2000). Data, which is of interest in Web usage mining, is:

- **Usage:** Data that describes the pattern of usage of Web Pages, such as IP addresses, page references, and the date and time of access.
- **User Profile:** Data that provides demographical information about users of the Web Site. This includes registration data and customer profile information.

Web usage mining is the type of Web mining activity that involves the automatic discovery of user access patterns from one or more Web servers. As more organizations rely on the internet and the World Wide Web to conduct business, the traditional strategies and techniques for market analysis need to be revisited in this context. Organizations often generate and collect large volumes of data in their daily operations. Most of this information is usually generated automatically by Web servers and collected in server access logs. Other sources of user information include *referrer logs* which contains information about the referring pages for each page reference and user registration or survey data gathered via tools such as CGI scripts.

Analyzing such data can help these organizations to determine the life time value of customers, cross marketing strategies across products and effectiveness of promotional campaigns among other things. Analysis of server access logs and user registration data can also provide valuable information on how to better structure a Web site in order to create a more effective presence for the organization. In organizations using intranet technologies, such analysis can shed light on more effective management of workgroup communication and organizational infrastructure. Finally, for organizations that sell advertising on the World Wide Web, analyzing user access patterns helps in targeting ads to specific groups of users.

Jiang, (2003) have been concluding that the application area of the Web usage mining can be divided into two main groups and they are:

- **Personalized**: discover the preference and needs of individual Web users in order to provide personalized Web site for certain types of users.
- **Impersonalized**: examine general user navigation patterns in order to understand how general users use the site.

The general architecture divides the Web usage mining process into two main parts. The first part includes the domain dependent processes of transforming the Web data into suitable transaction form. This includes preprocessing, transaction identification and data integration components. The second part includes the largely domain independent application of generic data mining and pattern matching techniques (such as the discovery of Association Rules and sequential patterns) as part of the system's data mining engine.

Given its application potential, particularly in terms of electronic commerce, interest in Web usage mining, increased rapidly in both the research and practice communities. As shown in Figure 2.1, three main tasks are performed in Web usage mining; preprocessing, pattern discovery and pattern analysis The following sub chapter explains brief description about the main task of Web usage mining process.

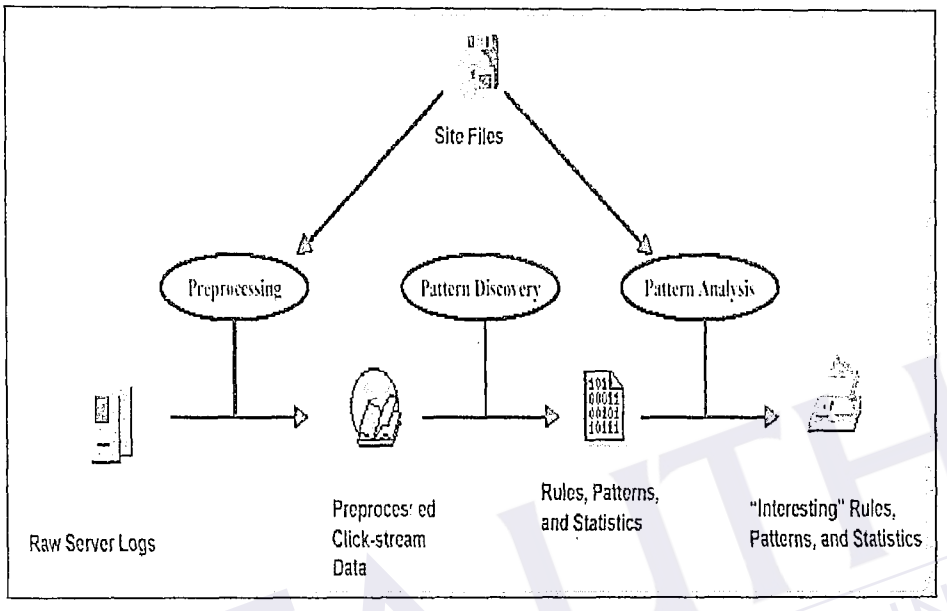


**Figure 2.1:** A High Level Web Usage Mining Process
(Adapted from Srivastava *et al.*, 2000)

## 2.4 High Level Web Usage Mining Process

### 2.4.1 Data Preprocessing

The inputs of the preprocessing phase may include the Web server logs, referral logs, registration files, index server logs and optionally usage statistics from a previous analysis. The outputs are the user session file, transaction file, site topology and page classifications. It's always necessary to adopt a *data cleaning* techniques to eliminate the impact of the irrelevant items to the analysis result. The usage preprocessing probably is the most difficult task in the Web usage mining processing due to the incompleteness of the available data (Cooley, 2000).

# REFERENCES


Abd. Wahab, M. H, Siraj, F and Yusoff, N. (2004). *Log Mining Using Generalize Association Rules*. In Proceedings of Master Final Project 2004 Presentation, UUM, Malaysia.

Agrawal, S, Agrawal, R., Deshpande, P.M. Gupta, A. Naughton, J., Ramakrishna, R and Sarawagi, S. (1996). *On the Computation of Multidimensional Aggregates*. In Proc. of the 22nd VLDB Conference, Mumbai, India. Pp 506-521.

Agrawal, R., Imielinski, T. and Swami, A. (1993). *Mining Association Rules between Sets of Items in Large Databases*. In Proceedings of the International ACM SIGMOD Conference, Washington DC, USA, pages 207–216.

Agrawal, R. and Srikant, R. (1994). *Fast Algorithm for Mining Association Rules*. Proc. of the 20th VLDB Conference. Pp 487-499.

Agrawal, R., and Srikant, R. (1995). *Mining Sequential Patterns*. In Proc. of the Eleventh International Conference on Data Engineering (ICDE), Taiwan. Pp 3-14.

Bertot, J. C., McClure, C. R., Moen, W. E., and Rubin, J. (1997). *Web Usage Statistics: Measurement Issues and Analytical Techniques*. Government Information Quarterly. 14 (4). Pp 373-395.

Borgelt, C. (2004). *Apriori: Finding Association Rules/Hyperedges with the Apriori Algorithm*. School of Computer Science, University of Magdeburg.

Boon Lay, C, Khalid, M and Yusof, R. (1999). *Intelligent Database by Neural Network and Data Mining*. In Proc. of Artificial Intelligent Applications in Industry, Kuala Lumpur. Pp 201-219.

Buchner, A. G., and Mulvenna, M. D. (1998). *Discovering Internet Marketing Intelligence Through Online Analytical Web Usage Mining*. SIGMOD Record, 27 (4), Pp 54-61.

Chen, M.-S., Jan, J., Yu, P.S. (1996). *Data Mining: An Overview from a Database Perspective*. IEEE Transactions on Knowledge and Data Engineering, (8:6). Pp 866.883.

Chung, H. M., Gray, P. (1999). *Special Section: Data Mining*. Journal of Management Information Systems, (16:1). Pp 11-17.

Cooley R. (2000). *Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data*. PhD thesis, Dept. of Computer Science, University of Minnesota.

Cooley R., Mobasher B., and Srivastava J. (1997). *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).

Cooley, R., Mobasher, B. and Srivastava, J. (1997). *Grouping Web Page References into Transactions for Mining World Wide Web Browsing Patterns*. Technical Report TR 97-021, University of Minnesota, Dept. of Computer Science, Minneapolis.

Cooley, R, Mobasher, B. and Srivastava, J. (1999). *Data Preparation for Mining World Wide Web Browsing Patterns*. Knowledge and Information Systems, 1(1).

Dereson, C. (1997). *Using an Incomplete Data Cube as a Summary Data Sieve*. Bulletin of the IEEE Technical Committee on Data Engineering. Pp 19-26.

Edelstein, H., A. (2001). *Pan for Gold in the Clickstream.* Informationweek.com, March12. 2001, Pp 77-91.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, R. (1996). *The KDD Process for Extracting Useful Knowledge from Volumes of Data.* Communications of the ACM, (39:11). Pp 27-34.

Gray, J., Bosworth, A., Layman, A and Pirahesh, H. (1996). Data Cube: A Rational Aggregation Operator Generalizing Group-By, Cross-Tab and Sub-Totals. In IEEE 12th International Conference on Data Engineering. Pp 152-159.

Han, J., Kamber, M. (2001). *Data Mining: Concepts and Techniques.* Morgan-Kaufmann Academic Press, San Francisco.

Hand, D. J. (1998). *Data Mining: Statistics and More*". The American Statistician. May (52:2). Pp 112-118.

Harinarayan, V, Rajaraman, A. and Ullman J.D. (1996). *Implementing Data Cubes Efficiently.* In Proc. of 1996 ACM-SIGMOD Int. Conf. Management of Data. Montreal, Canada. Pp 311-322.

Jiang, Q. (2003). *Web Usage Mining: Process and Application.* Presentation for CSE 8331.

Kosala, R., Blockeel, H. (2000). *Web Mining Research: A Survey.* ACM SIGKDD (Special Interest Group on Knowledge Discovery and Data Mining) Explorations. June, (2:1). Pp 1-10.

Mannila, H., Toivonen, H. and Verkamo, A. I. (1994). *Efficient Algorithms for Discovering Association Rules.* In AAAIWorkshop on Knowledge and Discovery in Databases, Seattle, Washington, USA, Pp 181–192.

McLaughlin, M., Goldberg, S. B., Ellison, N., and Lucas, J. (1999). *Measuring Internet Audiences: Patrons of an On-line Art Museum.* In S. Jones (Ed.), Doing Internet Research: Critical Issues and Methods for Examining the Net. Thousand Oaks, CA: Sage. Pp. 163-178.

Peacock, P. R. (1998). *Data Mining in Marketing: Part 1.* Marketing Management, Winter. Pp 9-18.

Pitkow, J., and Bharat, K. (1994). *Webvis: A Tool for World Wide Web Access Log Analysis.* In First International WWW Conference.

Rajagopalan, B., Krovi, R. (2002). *Benchmarking Data Mining Algorithms.* Journal of Database Management, Jan-Mar. 13, Pp 25-36.

Brin, S., Motwani, R., Ullman, J. D. and Tsur, S. (1997). *Dynamic Itemset Counting and Implication Rules for Market Basket Data.* In Proceedings of the International ACM SIGMOD Conference, Tucson, Arizona, USA, Pp 255–264.

Shukla, A., Deshpande, P.M., Naughton, J and Ramaswamy, K. (1996). *Storage Estimation for Multidimensional Aggregates in the Presence of Hierarchies.* In Proc. of the 22nd VLDB Conference. Mumbai, India. Pp 522-531.

Spiliopoulou M. (1999). *Data mining for the Web.* In Proceedings of Principles of Data Mining and Knowledge Discovery, Third European Conference, PKDD'99, Pp588-589.

Srivasta, J., Cooley, R., Deshpande, M., and Tan P. N. (2000). *Web Usage Mining: Discovery and Application of Web Usage Pattern from Web Data.* Department of Computer Science and Engineering, University of Minnesota.

Tang, C.; Lau, R.W.H.; Li, Q.; Yin, H.; Li, T.; and Kilis, D.(2000). *Personalized Courseware Construction Based on Web Data Mining*. In Proc. of the First International Conference on Web Information Systems Engineering (WISE 2000) vol.2, Pp. 204-211.

Tang, Y. T. and McCalla, G. (2001). *Student modeling for a Web based Learning Environment: a Data Mining Approach*. Department of Computer Science, University of Saskatchewan, Canada.

Webopedia. (2001). *Web Server*. http://webopedia.internet.com/TERM/W/Web_server.html Date Accessed : 06 April 2004.

Wilson, T. (1999). *Web Trafic Analysis Turns Management Data to Business Data. TechWeb*. http://www.internetk.com/story/INW19990402S0006 Date Accessed : 24 March 2004.

Xue, G. R., Zeng, H. J., Ma, W. Y and Lu, C. J. (2002). *Log Mining to Improve the Performance of the Methods from statistic, Neural Nets, Machine Learning and Experts System*. Morgan Kaufman.

Zaiane, O. and Luo, J. (2001). *Towards Evaluating Learners' Behavior in a Web-based Distance Learning Environment*. In Proc. of IEEE International Conference on Advanced Learning Technologies, Madison, WI. Pp 357-360.