Web Usage Mining: Discovery and Use of AIOU Web Usage Patterns

Moiz Uddin Ahmed, Department of Computer Science, Allam Iqbal Open University, Islamabad, Pakistan
Amjad Mahmood, Department of Computer Science, College of IT, University of Bahrain, Zallaq, Bahrain

ABSTRACT

The technological revolutions have opened up new ways of information and communication. The Internet is growing as a vital source of information in this modern era of technology. The ever increasing volume of information through WWW is creating complexity in the design, development and deployment of WWW. It has become important for the organizations to analyze the usage of their websites. The web usage analysis may help the organizations not only to monitor the load on their websites and cater for the needs of their potential clients but also enhance their web services and restructure the organization to better serve their clients. Web mining has emerged as important research areas used to discover information which can be utilized for improvement of websites. Allama Iqbal Open University (AIOU) is one of the largest open and distant university of the world. Due to unique philosophy of open and distant learning, AIOU has been providing useful information online through its website. It is an active website which is flooded with huge flow of information. This paper presents web usage analysis of AIOU website and provides statistical analysis of the usage patterns. It presents how the results were used not only to enhance the web contents and services but also discusses how these results helped the university to allocate and reallocate its resources. The reallocation was used to improve efficiency and processes of the university in order to better serve its clients.

Keywords: Data Preprocessing, Pattern Analysis, Web Mining, Web usage, World Wide Web

1. INTRODUCTION

The advancements in Information and Communication Technologies (ICT) have led to new approaches in information exchange and communication. The growth of the Internet has increased exponentially and therefore playing a vital role in dissemination of information. The number of Internet users by June 30, 2012 was 2,405,518,376 which is 34.2% of the total world population (Internet World Stats, 2012). As a result, the World Wide Web (WWW) has evolved as an information hub on the Internet. It manages the enormous flood of information and presents it to the Internet users. However; the arrangement of information on WWW is...
important for its success. It should satisfy the needs of relevant users by providing right information in the right way with minimum hassle.

The WWW data is exchanged via client server communication over the Internet. The client makes use of web browser and sends the Hyper Text Transfer Protocol (HTTP) request to the web server. The web browser simply sends and receives the data. However the web server has more responsibilities to perform. It not only receives the request and sends the reply back but also stores access log details of different clients. The web server log file maintains history of page request per W3C standard including Internet Protocol (IP) address, date and time of request, the requested page, HTTP code, and bytes served (Phillip & Brian, n.d.). The log file analysis provides an opportunity to improve the website for better use and benefit of the organization. As a result, the field of Web Mining has emerged as an important area of research. Web mining is the discovery and analysis of web data. It is used to discover information which can be utilized for improvement of websites. Web Mining is divided into two major areas: Web content mining and Web usage mining. Web content mining is the automated search of useful information available from a large number of websites whereas web usage mining is the study and analysis of user access patterns through log files (Kavita, Gulshan, & Vikas, 2011). Web usage mining discovers useful knowledge from the web access log file. It aims to analyze the access records of client interaction and finds some interesting patterns. The results can be used to improve the structure of the website and performance of the web server (Navin, Solanki, & Manoj, 2010). It can also be used to improve the website recommendations, adaptive contents and personalized search on the website (Eirinaki, & Vazirgiannis, 2003).

Web usage mining is comprised of four steps (Srivastava, Sharma, & Kumar, 2011; Chitraa & Antony, 2010): data collection, data pre-processing, pattern discovery and the last pattern analysis. The data collection involves compilation of user log data from the web and proxy servers. The preprocessing involves data cleaning, user and session identification and path detection. The pattern discovery stage uses the data mining techniques to further refine the patterns. It includes classification, clustering, statistical analysis and machine learning algorithms. Finally the pattern analysis stage comes, which further process and filters the patterns.

For an educational institution, the website should be capable to provide the information to the students, teachers and parents. Its importance becomes more vital for a distance learning institution where the students, teachers and parents are separated by geographical distance. Web mining techniques can be applied to extract the web usage analysis to discover the hidden patterns. It can be used to gather the important information including number of visitors, popular hours and popular page hits. The information can be used to improve the website structure, navigation sequence and publish productive academic information by enhancing the web services.

Allama Iqbal Open University (AI OU) is one of the largest university of the World. It is a unique institution due to its philosophy and nature of academic programs. The University has a large number of students all over the country. It also offers programs for overseas students. The university is providing useful information through its website. It is an active website which provides academic and administrative information to the concerned stakeholders (VC Annual Report, 2012). The main objective of this research is to analyze the web usage of AI OU and discover important patterns which may be helpful for future improvements of web services and improvement of the overall efficiency of the university. It identifies interesting facts about the web usage which helped the academic and administrative departments to rearrange the existing contents and include some new information on the website. The paper also discusses that how the results were used to allocate and reallocate university resources to improve efficiency and better serve its clients.

The rest of the paper is organized as follows. Section 2 presents the related work in the field of web usage mining. Section 3 presents our web log analysis model and describe its components. Section 4 presents the results and
Problem-Based Online Learning in Indigenous Tertiary Education: Reflections on Implementation
[www.igi-global.com/chapter/problem-based-online-learning-indigenous/23545?camid=4v1a](www.igi-global.com/chapter/problem-based-online-learning-indigenous/23545?camid=4v1a)

Farmers' Access and Use of Mobile Phones for Improving the Coverage of Agricultural Extension Service: A Case of Kilosa District, Tanzania
[www.igi-global.com/article/farmers-access-and-use-of-mobile-phones-for-improving-the-coverage-of-agricultural-extension-service/163394?camid=4v1a](www.igi-global.com/article/farmers-access-and-use-of-mobile-phones-for-improving-the-coverage-of-agricultural-extension-service/163394?camid=4v1a)