International Journal of Innovative Computing, Information and Control
Volume **4**, Number **2**, February **2008**

ICIC International ©2008 ISSN 1349-4198
pp. 381–389

KNOWLEDGE DISCOVERY FROM WEB USAGE DATA: EXTRACTION AND APPLICATIONS OF SEQUENTIAL AND CLUSTERING PATTERNS - A SURVEY

G. T. Raju[1], P. S. Satyanarayana[2] and L. M. Patnaik[3]

[1]Department of Computer Science and Engineering
B. M. S. College of Engineering
Bangalore, Karnataka 560019, India
gtraju1990@yahoo.com; gtraju.cse@bmsce.ac.in

[2]Department of Electronics and Communication
B. M. S. College of Engineering
Bangalore, Karnataka 560019, India
pssvittala@gmail.com; pssvittala.ece@bmsce.ac.in

[3]Department of Computer Science and Automation
Indian Institute of Science
Bangalore, Karnataka 560012, India
lalith@micro.iisc.ernet.in

Received December 2006; Revised May 2007

Abstract. *Knowledge Discovery from the Web Usage Data (KDWUD) is that area of Web mining which deals with the extraction of interesting knowledge from the secondary data generated by the user interactions with the web that are stored in log files of the web servers. KDWUD has become very critical for effective and efficient managing of the activities related to e-business, e-services, e-education, personalization, web site management and so on. Mining web access logs that contain substantial data about user access patterns on one or more web localities is an emerging research area. In this paper, we present a survey of the recent developments in this area with a particular focus on Sequential Patterns and Clustering patterns in order to understand the navigational behavior of the users.*

Keywords: KDWUD, Clustering, Sequential patterns, Web usage mining

1. Introduction.

Online data continues to grow at an explosive pace, due to the Internet and widespread use of database technology. This has created an immense opportunity and need for methodologies of Knowledge Discovery in Databases (KDD). KDD deals with non-trivial extraction of implicit, novel, and potentially useful information from databases using advanced machine learning techniques. As more organizations rely on the WWW to conduct business, traditional strategies and techniques for market analysis need to be revisited. Organizations collect large volumes of data and analyze it to determine the lifetime value of customers, cross marketing strategies across products, and effectiveness of promotional campaigns. In the Web, such information is generally gathered automatically by Web servers and collected in server or access logs. Analysis of server access data can provide information on how to restructure a Web site for increased effectiveness, better management of workgroup communication, and analyzing user access patterns to target ads to specific groups of users.

The WWW is an immense source of data [1] that can come from web content (represented by the billions of pages publicly available), or from the web usage (represented by the log information daily collected by the servers around the world). Web mining is