Chapter 13
Web Usage Mining:
Concept and Applications at a Glance

Vinod Kumar
Maulana Azad National Institute of Technology, India

R. S. Thakur
Maulana Azad National Institute of Technology, India

ABSTRACT

Websites have become the major source of information, and analysis for web usage has become the most important way of investigating a user's behaviour and obtaining information for website owners to use to make any strategic decisions. This chapter sheds light on the concept of web usage mining, techniques, and its application in various domains.

INTRODUCTION

World Wide Web has become the most popular platform for people. More than millions of users are interacting daily with the websites and visiting large numbers of websites, leaving behind a variety of information. Due to its various attractive and beneficial services web is getting more popular day by day. Now, No one is untouched with magic of World Wide Web service of internet technology. Website proves to be a popular means for information circulation among to the world. Today, almost every organization provides services and significant information to the targeted person through their websites. Such as re-source sharing, online shopping using e-commerce, online banking, e-learning, e-banking, online news broadcast, online rail ticket reservation, hotel's room booking and many more. Due to cloud computing and other supporting services, the World Wide Web is getting ubiquitous and a usual instrument for day to day life's activities of common man. Because of unprecedented and exponential growth in popularity of web, there have been great efforts by the researchers in development of techniques to deal with the web data. Initially, the data mining techniques were being used to retrieve, search and organize the information over the web. There was no distinct term for the area of web. It was Etzioni (Etzioni, 1996) who first coined the term web mining in his paper. Since, then this area of research is studied under this term "Web mining" (Zdravko, M. & Daniel, T.L., 2007) defined web mining as the application of data

mining techniques to discover the patterns in web content, structure and usage. Further, the web mining has been categorized into three major parts-Web content mining, web structure mining, and web usage mining. Figure 1 shows the taxonomy of web mining.

Web Content Mining (WCM)

As the name implies, it is the process of extraction of information from web document, Text, audio, video, structured records such as- list, tables. Web content mining (Adeniyi, D.A. & Wei, Z. Yangquan, Y., 2016) involves techniques for Summarization, Classification, and Clustering of information on over the World Wide Web. It also collects interesting patterns about user's needs and customer behaviour. It targets knowledge discovery in which it collects information from text documents, multimedia documents such as images and video which are embedded or linked in a web. Web content mining mines unstructured, semi-structured, structured, and multimedia data. These are the tools used to extract essential information that one needs- Screen scraper, Automation anywhere, Web content extractor, Web info, and extractor.

Web Structure Mining (WSM)

Web structure mining is also called link mining. Based on the topology of the hyperlink, web structure mining (Adeniyi, D.A. & Wei, Z. Yangquan, Y., 2016) will categorize the web page and generates the information. Such as similarity and relationship between different Website .It is a tool to extract patterns from hyperlinks in the web. It also generates structural summary about website and webpage by analyzing the link. Firstly, Hits and Page rank algorithm is the popular web structure mining algorithm where Hits algorithm ranks the web pages by processing in-links and out-links of the web pages. In this algorithm, the web page is named as authority if the web page is pointed by many hyperlinks and web pages is named as hub if the webpage point to various hyperlinks. Secondly, Page rank Algorithm, It is the most commonly used algorithm for ranking various pages. Working of PageRank Algorithm depends upon the link structure of the web pages. PageRank algorithm considers a back link in deciding the rank score.

Web Usage Mining (WUM)

Web Usage Mining (Mobasher, B., 2005) is defined as the application of data mining techniques to discover the useful pattern from web log data to know the user's behaviour as result of activities performed on the web. It is explained in detail in the subsequent section.

BACKGROUND

It has always been the point of interest to know about the visitors who are visiting the website of any organization or individual person. In the field of e-commerce it is inevitable to know the users behaviour and sentiments about a product or services provided by any company or organization. There have been tremendous efforts towards the development of the tools and techniques for discovery of the useful pattern and other significant information about the user's activity performed over the World Wide Web. Web usage mining which comes under the category of web mining is the area mining which deals with

Related Content

Leveraging UML for Access Control Engineering in a Collaboration on Duty and Adaptive Workflow Model that Extends NIST RBAC

Solomon Berhe, Steven A. Demurjian, Jaime Pavlich-Mariscal, Rishi Kanth Saripalleand Alberto De la Rosa Algarín (2021). *Research Anthology on Recent Trends, Tools, and Implications of Computer Programming (pp. 916-939).*

www.irma-international.org/chapter/leveraging-uml-for-access-control-engineering-in-a-collaboration-on-duty-and-adaptive-workflow-model-that-extends-nist-rbac/261061

Offshore Software Testing in the Automotive Industry: A Case Study

Tabata Pérez Rentería y Hernándezand Nicola Marsden (2021). *Research Anthology on Recent Trends, Tools, and Implications of Computer Programming (pp. 1574-1591).*

www.irma-international.org/chapter/offshore-software-testing-in-the-automotive-industry/261091

Open Innovation in Small and Medium Enterprises: Perspectives of Developing and Transitional Economies

Hakikur Rahman (2020). *Disruptive Technology: Concepts, Methodologies, Tools, and Applications (pp. 2030-2052).*

www.irma-international.org/chapter/open-innovation-in-small-and-medium-enterprises/231277

An Empirical Investigation of Decision Making in IT-Related Dilemmas: Impact of Positive and Negative Consequence Information

Chen Zhang, Judith C. Simonand Euntae "Ted" Lee (2021). *Research Anthology on Recent Trends, Tools, and Implications of Computer Programming (pp. 1671-1690).*

www.irma-international.org/chapter/an-empirical-investigation-of-decision-making-in-it-related-dilemmas/261096

Augmentative and Alternative Communication Technologies

Carol (Heins) Gonzales, Gondy Leroyand Gianluca De Leo (2012). *Computer Engineering: Concepts, Methodologies, Tools and Applications (pp. 1164-1180).*

www.irma-international.org/chapter/augmentative-alternative-communication-technologies/62504