

Elements of Web Personalization

Anuj Kumar Rathee¹, Umesh Gupta²

¹M.Tech Student, Dept. of ECE, MERI College of Engineering, Maharishi Dayanand University, Rohtak, Haryana, India

²Assistant professor, Dept. of ECE, MERI College of Engineering, Maharishi Dayanand University, Rohtak, Haryana, India

Abstract: Web personalization is a challenging strategy; this concept provides various techniques to make a website according to the requirements of the customer because the impact of the World Wide Web as a main source of information acquisition is increasing dramatically. The existence of such abundance of information leads to the information overload, in combination with the dynamic and heterogeneous nature of the web, makes web site exploration a difficult process for the average user. This thesis would address the requirement of clear understanding of the web personalization terms, classifications of the web personalization topics, various techniques of the web personalization, categorization of these techniques.

Keywords: Web Personalization, Web Mining, User Profiling, Log Analysis.

I. INTRODUCTION

Web personalization is a technique to personalize web pages according to the needs of customer, without asking them explicitly. It provides an organization, a variety of services to increase their sales, customer retention and to compete in the market. Web personalization is a strategy rather than it is a technique. It is a bundle of various techniques and tools. This paper studies the suggested web personalization techniques and evaluation methods and metrics to evaluate them. Web Personalization is simply defined as the task of making Web-based information systems adaptive to the needs and interests of individual users [3]. Web personalization is a process and a strategy to compete in the online market. This research work compares the performance of different web personalization techniques and suggests the evaluation techniques. Web site personalization can be defined as the process of customizing the content and structure of a Web site to the specific.

Personalization system requires knowledge to be represented in a machine interpretable form which is not available in web. In semantic web we can develop languages for expressing information in machine process able form therefore semantic web is the most appropriate platform for realizing personalization [1].

Principal elements of Web personalization include (a) the categorization and pre-processing of Web data, (b) the extraction of correlations between and across different kinds of such data and (c) the determination of the actions that should be recommended by such a personalization system [2]. The personalization of services offered by a Web site is an important step in the direction of alleviating information overload, making the Web a friendlier environment for its individual user and hence creating trustworthy relationships between the Web site and the visitor-customer [3].

Personalization is not limited to online shopping, and it is applied to search engines. In 2004, two personalized search engines, A9.com by an Amazon subsidiary and MyJeeves by Ask Jeeves, were launched to let users store individual search results and then provide personalized web searches. Recently, the My Yahoo service has been enhanced with personalized searching. That is, the users can save pages of search results to a 'personal web' and block URLs from appearing on the result list. Google generates personalized offers under the sponsored links shown on the top of the browser [8].

II. WEB USAGE MINING

Web mining is a complete process rather than an algorithm. In the case of Web usage mining this process results in the discovery of knowledge that concerns the behavior of users. Originally, the aim of Web usage mining has been to support the human decision making process and, thus, the outcome of the process is typically a set of data models that reveal implicit knowledge about data items, like Web pages, or products available at a particular Web site[7]. The purpose of Web usage mining is to apply statistical and data mining techniques to the pre-processed Web log data, in order to discover useful patterns. In the context of Web mining, we can distinguish two cases, user clusters and page clusters. Page clustering identifies groups of pages that seem to be conceptually related according to the users' perception. User clustering results in

groups of users that seem to behave similarly when navigating through a Web site. Such knowledge is used in e-commerce in order to perform market segmentation but is also helpful when the objective is to personalize a Web site [2].

III. LOG ANALYSIS

Log analysis tools (also called traffic analysis tools), take as input raw Web data and process them in order to extract statistical information. Such information includes statistics for the site activity (such as total number of visits, average number of hits, successful/failed/redirected/cached hits, average view time, average length of a path through a site), diagnostic statistics (such as server errors, page not found errors), server statistics (such as top pages visited, entry/exit pages, single access pages), referrers statistics (such as top referring sites, search engines, keywords), user demographics (such as a stop geographical location, most active countries/cities/organizations), client statistics (visitor's Web browser, operating system, cookies) etc[2].

IV. WEB LOG

Each access to a Web page is recorded in the access log of the Web server that hosts it. The entries of a Web log file consist of fields that follow a predefined format. The fields of the common log format are: remote host rfc931 authored date "request" status bytes where is not available, rfc931 is the remote log name of the user, authuser the username as which the user has authenticated himself, available when using password protected WWW pages, date the date and time of the request, "request" the request line exactly as it came from the client (the file, the name and the method used to retrieve it), status the HTTP status code returned to the client, indicating whether or not the file was successfully retrieved and if not, what error message was returned, and bytes the content-length of the documents transferred. If any of the fields cannot be determined a minus sign (-) is placed in this field[2,5].

V. PRIVACY ISSUES

The most important issue that should be encountered during the user profiling process is privacy violation. Many users are reluctant to giving away personal information either implicitly as mentioned before, or explicitly, being hesitant to visit Web sites that use cookies or avoiding to disclose personal data in registration forms. P3P (Platform for Privacy Preferences) is a W3C proposed recommendation [P3P] that suggests an infrastructure for the privacy of data interchange. This standard enables Web sites to express their privacy practices in a standardized format that can be automatically retrieved and interpreted by user agents. [2]. Therefore, the process of reading privacy policies will be simplified for the users, because key information about what data are collected by a Web site can be automatically conveyed to a user, and discrepancies between a site's practices and the user's preferences concerning the disclosure of personal data will be automatically flagged. P3P, however, does not provide a mechanism for ensuring that sites actually act according to their policies[5]. While web personalization has been shown to be effective in attracting new users, its use should be balanced by taking a proactive approach towards protecting data privacy on the users' side. Consistent with prior IS work, our findings showed that online shoppers are concerned about their privacy (Awad and Krishnan 2006). Threats to user privacy involve tracking users without their consent, divulging users' profiles online, buying (or selling) users' profiles from (or to) other firms, or falsifying and misinterpreting the mining results to support a predetermined marketing goal [8].

VI. TOOLS AND APPLICATIONS

In this section we present some of the most popular Web sites that use methods such as decision tree guides, collaborative filtering, and cookies in order to profile users and create customized Web pages [5]. It is evident that Web usage mining is a powerful tool for corporations that invest in the e-business sector. The application of Web usage mining techniques to data gathered from customers' online activity helps them to acquire business intelligence by providing high-level knowledge in the form of rules and patterns that describe consumer navigational and purchasing behavior [Buchner and Mulvenna 1998] [5]. Popular Web sites such as Yahoo! [YAH], Excite [EXC] or Microsoft Network [MSN] allow users customize home pages based on their selections of available content, using information supplied by the users and cookies thereafter. In that way, each time the user logs in the site, what she/he sees is a page containing information addressed to their interests [2].

VII. CONCLUSION

Though a lot of research work has been done in the area of web mining and hence in the field of web personalization, lot of techniques have been developed and used too to tailor the contents and structure of the websites according the preferences of the users. In order to get all these things done a series of steps is followed. In this research paper I have tried to explore

the basic and new techniques of web personalization. Personalization of a web site may be performed by the provision of recommendations to the users, highlighting/adding links, creation of index pages, etc

The overall process of Web personalization consists of five modules, namely: user profiling, log analysis and Web usage mining, information acquisition, content management and Web site publishing. Web personalization is a domain that has been recently gaining great momentum not only in the research area, where many research teams have addressed this problem from different perspectives, but also in the industrial area, where there exists a variety of tools and applications addressing one or more modules of the personalization process.

Web usage mining is the process of applying statistical and data mining methods to Web log data in order to extract useful patterns concerning the users' navigational behavior, user and page clusters, as well as possible correlations between Web pages and user groups. User profiling is the process of gathering information specific to each visitor to a Web site either implicitly, using the information hidden in the Web logs or technologies such as cookies, or explicitly, using registration forms, questionnaires, and the like.

The main component of a Web personalization system is the usage miner. Log analysis and Web usage mining is the procedure where the information stored in the Web server logs is processed by applying statistical and data mining techniques such as clustering, association rules discovery, classification, and sequential pattern discovery, in order to reveal useful patterns that can be further analyzed.

Web personalization is a domain that has been recently gaining great momentum not only in the research area, where many research teams have addressed this problem from different perspectives, but also in the industrial area, where there exists a variety of tools and applications addressing one or more modules of the personalization process

I started from the meaning of web personalization, its characteristics, issues related to the web personalization, viz viability of the web personalization, user's privacy issues, process of web personalization, and classification of web personalization techniques according to their applicability.

References

- [1]. Ayesha Ameen, Khaleel Ur Rahman Khan, B.Padmaja Rani, Vol 2, No.6, 2012, Semantic Web Personalization: A Survey, Information and Knowledge Management, ISSN 2224-5758 (Paper) ISSN 2224-896X (Online), www.iiste.org
- [2]. Magdalini Eirinaki, Michalis Vazirgiannis, ACM Transactions on Internet Technology, { Vol. 3, No. 1, February 2003 },
- [3]. Web Mining for Web Personalization, <http://doi.acm.org/10.1145/643477.643478>
- [4]. R.Khanchana, Dr.M.Punithavalli, Volume 11 Issue 7 Version 1.0 May 2011, Web Page prediction for Web Personalization: A Review, Global Journal of Computer Science and Technology, Global Journals Inc. (USA)
- [5]. MAGDALINI EIRINAKI and MICHALIS VAZIRGIANNIS, February 2003, Pages 1–27, Web Mining for Web Personalization, Vol. 3, No. 1, ACM Transactions on Internet Technology
- [6]. A.JEBARAJ RATNAKUMAR, AN IMPLEMENTATION OF WEB PERSONALIZATION USING WEB MINING TECHNIQUES, Journal of Theoretical and Applied Information Technology © 2005 - 2010 JATIT. www.jatit.org
- [7]. DIMITRIOS PIERRAKOS¹, GEORGIOS PALIOURAS¹, CHRISTOS PAPA THEODOROU² and CONSTANTINE D. SPYROPOULOS, 2003 Web Usage Mining as a Tool for Personalization: A Survey, User Modeling and User-Adapted Interaction 13: 311–372, 2003. 311# 2003 Kluwer Academic Publishers. Printed in the Netherlands.