A.Noorhidawati, et al. (Eds.): ICOLIS 2012, Kuala Lumpur: LIS, FCSIT, 2012: pp 129-136

Kill the king but keep the tradition: web personalisation based document recommender system in academic library services

Daya Sridhar
Librarian SIWS N.R.S. College of Commerce and Economics, Mumbai, India
e-mail: daya.sridhar@gmail.com

'To carry knowledge to the doors of those that lack it and educate all to perceive the right. Even to give away the whole Earth cannot equal that form of service.' – Manusmrti

ABSTRACT

Academic libraries have been losing their library patrons, in part due to the advent of the Internet and World Wide Web which facilitate users' easy access to information from anywhere and anytime, and by ignoring the librarians and the physical library. Libraries are trying to woo the patrons back with alternatives, such as providing a place in the library for learning commons and digital commons. Technology intervention programmes are available for the librarians to obtain clickstream data and analyse it to determine the web negotiations of their users and find websites, blogs and other source web pages with high frequency use, as well as their success and failure rates. Basic statistical knowledge is sufficient to download the users' clickstream data from the server, clean the data and analyse it to obtain the total web information seeking behaviour of users, which can be used as a base for designing effective library information services. Librarians can then reorganize such prioritized sources in their library websites. Web personalization can bring back the lost users, the personalized reference service of yester years in a digitized format, and win more readerships. This presents a chance for librarians to get into the WWW and walk along with their readers in their information tour.

Keywords: Web Personalization; Recommender system; Web usage mining; Digital library service

INTRODUCTION

Ranganathan, the father of Indian librarianship cited *Manusmrti* in his book entitled *Reference Service*, published as early as 1957. He is a prophet of the library profession as he identified the basic concept of library service reflecting in the code of *Manu*. In accordance to his prophecy, libraries have developed the potential to deliver information to the users in their laptops/desktops without a necessity for them to come to the library. It cannot be an exaggeration if one concluded that, taking knowledge to the doors (laptops) of the users has become outreach programme today in general and library service mission in particular.

BRIDGING THE LIBRARIAN-USER-LIBRARY GAP

Librarians cannot continue to be sitting amidst books and periodicals in their tangible form, while leaving the subscribed databases to the independent opinions of the users. These users may have already switched over to the digital information environment, even though the success rate of their information retrieval and the subjective time involved in that process are not appreciable. It is time that the community of library professionals come out to the digital open.

The advent of the knowledge society over the development of information society in recent decades enabled capturing, filtering and storing large amount of digital data. These data need to be further processed to gain valuable information and knowledge of the user-information environment of their interactions which is the core of Library and Information systems. The scientific field dealing with extracting information and knowledge from data has evolved rapidly to cope with the extent and growth of information sources the number of which has geometrically increased with the appearance of the Internet and the World Wide Web. All traditional approaches in information retrieval, knowledge acquisition and data mining warrant for being adapted for the dynamic, heterogeneous and unstructured data regarding the user interaction for information in the web.

The resulting factor is the shaping of a fully-fledged research discipline namely web mining. Scharnhort and Wouters (2009) posit that, "The emergence of the Internet and the Web has modified the way that scientists and scholars search for and find information." The library also stands transformed in promoting library services with easy access from 'anywhere anytime'. "Digital library services have changed the way patrons seek and access the high-quality information traditionally found only in libraries" observes Srivastava et al (2000). Though the process of information access and retrieval sounds easier, hurdles are many between the user and his or her required information. The identification and retrieval of relevant information's entails technicalities requiring digital competencies which has an edge over digital literacy. Assessment and actualization of a wide range of information needs and different types of user behaviour form the basis for the design and implementation of information systems and services offered by the information service providers namely the libraries whether traditional or digital. Tracking, quantifying and analyzing the login statistics may facilitate more opportunities to the service providers to identify the changes and determine matching services promoting electronic delivery of information

LIBRARY INFORMATION SERVICE

Information is vital for any decision making in general. Current Awareness Service (CAS) is meant for groups, while Selective Dissemination of Information (SDI) is specific to individuals. So far as research in its academic environment is concerned, it is the SDI that is most relevant in the context of information needs of individual scholars. In order to fulfil the needs of groups or individuals, the ultimate goal of any library is designed to be reference service, now widely known as information service. Ranganathan (1987) defined reference service as "a process of establishing a contact between the right reader and the right document/information at the right time in the right personal way." Though Ranganathan defined the concept in 1957, it holds good today too as libraries exist for that sole purpose. In a traditional environment, they were the printed documents while in

modern context the documents are found to be more in digital format though printed materials are on the decline comparatively speaking.

WEB: THE BEACON LIGHT OF LIBRARY SCIENCE

In a traditional library, the librarian had total control of the library, its collection and services through the library catalogue with articulated entries. These entries were in accordance with procedures defined in cataloguing style manuals and a profile of users of his/her library with the description of individual user information needs precisely enunciated including keywords – all these providing possible matches between user defined needs and relevant information/document. Any kind of user-library transaction was confined to the records within the portals of a library. Those are the days of green past. Yet the personal touch in information service cannot be ruled out even in the ICT dominated digital environment. The only difference between the tradition oriented library reference service and the present digital environment is that the user gets all his information requirement supplied to his personal computer or e-mail box, including alerting services.

NEED FOR WEB PERSONALISATION

The number of patrons personally visiting library users have been on a constant decline. They have begun to use the Internet and the WWW to satisfy their information needs. The library has to bring them back into its fold of their websites.

Information service in general is based on a profile of the individual user and profiles of the documents, between which keywords function as the bridge. In a digital environment, the entire transaction involving information retrieval from the web is in intangible data. The librarian, to gain a thorough knowledge needs information about the user activities on the WWW and the websites or documents the users are accessing either with success or failures.

According to Bracke (2004) "The web has had a significant impact on libraries changing the formats and methods in which they present their resources and services to users. Most libraries use a website as a tool for organizing and providing access to resources in print and electronic formats. The impact of the web on libraries, however, has been more profound than providing a new access point for its users." One of the Association of Research libraries (ARL) (USA) white paper confirms that, "This change in the use of physical libraries is further demonstrated by a decline in foot traffic and the number of reference questions fielded by libraries in the past five years." This is true as found in the experience of many working librarians in India too. Web usage mining, tracking the user navigation and bringing prioritized websites into the library portal for easy and fast access can only bring back the alienated library users into the purview of library services.

WEB PERSONALIZATION: A SIMULATION TO LIBRARY TRADITION

Web personalization is a component of web mining. Srivastava, et al (2000) define web usage mining as "the process of applying data mining techniques to the discovery of usage patterns from web data, targeted towards various applications." Data mining efforts associated with the web, called web mining, can be broadly divided into three classes namely, content mining, usage mining and structure mining.

DATABASE APPROACH

Database approaches to web mining have focused on techniques for organizing the semi-structured data on the Web into more structured collections of resources and using standard database querying mechanisms and data mining techniques to analyze it.

Information Filtering/Categorization

A number of web agents use various information retrieval techniques and characteristics of open hypertext Web documents to automatically retrieve, filter and categorize them.

Personalized Web Agents

This category of web agents learn user preferences and discover Web information sources based on these preferences and discover Web information sources based on these preferences, and those of other individuals with similar interests (using collaborative filtering). Web personalization can be described as any action that makes the Web experience of a user customized to the user's taste of preferences. Principal elements of Web personalization include modelling of Web objects (such as pages or products) and subjects (such as users or customers), categorization of objects and subjects, matching between and across objects and /or subjects and determination of the set of actions to be recommended for personalization.

Web personalization in a digital environment is parallel to the concept of reference service of the traditional library and its various manually operated services. Dai and Mobasher (2000) state that, "Traditional approaches to personalization have included both content-based and user-based techniques. Content based techniques use personal profile of users and recommend other items or pages based on their content similarity to the items or pages that are in the user's profile. The underlying mechanism in these systems is usually the comparison of sets of keywords representing pages or item descriptions" According to Maurice, et al. (2000)" Personalization of search results list for particular information that users want or need exactly, without expecting from them to ask for it explicitly". Web personalization is one of the fastest growing segments of the Internet economy. Because it can help in reducing information overload and give users a more customized experience of a web site, search personalization can reduce waste time to find requested information on the web.

Web personalization, though could be traced during 1990s, and became popular in 2000 and after. Recently, researchers have used Artificial Immune System (AIS) as a web personalization technique to optimize many problems as illustrated in the papers by Acilar and Arslan (2009); Narzi et al.(2009), and Secker, et al (2008) Examples of such systems are

Letizia found in the papers by Lieberman,(1995) and Web watcher in the paper by Joachims, Freitag and Mitchell (1997). These systems perform well from the perspective of the end user who is searching Web for information. In a Web environment as Mulvenna et al., (2000) puts that "The ultimate goal of any user-adaptive system is to provide users with what they need without asking them for it explicitly". This concept exactly is identical to that of personalized reference service in its traditional sense. Automatic personalization therefore, is a central technology used in such systems in general and it should be relevant and meaningful in a virtual library information environment.

WEB PERSONALISATION BASED RECOMMENDER SYSTEM

In the context of the Web, personalization implies the delivery of dynamic content, such as textual elements, links, advertisement, product recommendations, etc., that are tailored to needs of interests of a particular user or a segment of users.

Deitel (2001) observes that "personalization uses information from tracking, mining, and data analysis to customize a person's interaction with a company's products, services, web site and employees. Consumers and companies can benefit from the unique treatment resulting from personalization. Providing content of special interest to your visitor can help establish a relationship that can build upon each time that person returns to your site." Mobasher, et al. (2004) illustrate personalization with clarity saying that, "The goal of personalization based on Web usage mining is to recommend a set of objects to the current (active) user, possibly consisting of links, ads, text, products, or services, tailored to the user's perceived preferences as determined by the matching usage patterns." This task is accomplished by matching the active user session (possibly in conjunction with previously stored profiles for that user) with the usage patterns discovered through Web usage mining. The usage patterns used in this context aggregate usage profiles since they provide an aggregate representation of the common activities or interests of groups of users. This process is performed by the recommendation engine which is the online component of the personalization system. If the data collection procedures in the system include the capability to track users across visits, then the recommendations can represent a long term view of user's potential interests based on the history of user's activity either within the site or WWW.

What is applicable to marketing strategies of web personalization and recommender system to the sales and services of their commercial products are applicable to also libraries which have adopted marketing strategies for many of their tailored services suiting individuals and groups of library users in their traditional environment already. Librarians have to turn to the Web as library users *enmasse* already have entered there.

MODUS OPERANDI

In a campus wide networked environment, the academic institutions ought to have main servers which automatically records and maintains the web log history of the click streams (series of clicks pressed by a user from the time of her/his opening the Internet till she/he logs out of the net session). Many may not know the existence of such enormous data

within their click away distance. The analysis of such data downloaded for Web personalization, (after cleaning the data) can be achieved by means of indigenous software programs with algorithms written or licensed software's or free downloads. When one may enter the keywords 'Web usage mining software free download' a number of packages both price tagged and free are displayed on the screen. The criteria for the appropriate choice should be based on budget provision, and the number of nodes on the campus to be covered. A librarian with basic statistical knowledge can do the whole thing just downloading the records of click stream and user sessions from the server itself.

WEB RECOMMENDER SYSTEM INITIATIVES

Web personalization has its roots at the practice of fixing 'hit logs' to websites in order to find out the visitor statistics which was later extended to commercial product and service websites in a matured form known as web mining. The same concept is applicable to study the library users their web negotiations for academic information based on the click streams captured through the campus main web server or the library server. This shall provide insight to the librarians of concerned institutions to know the web information transactions of the users besides the use their own library websites so that the librarian can intervene with his information service assistance with a personal digital touch.

Web personalization initiatives on the part of the library professionals provide opportunities for them to learn about their user information activities on the WWW so that they can understand the information seeking behaviour in the digital environment, bring their most preferred websites into the portals of the library website. MyLibrary at Los Alamos National Laboratory (LANL) provides digital library users (as individual or groups) with a personalised Web environment while offering all features like User education, current Awareness service and Selective Dissemination of information. MyLibrary @ LANL is the result of a project sponsored by LANL Research Library, which provides electronic services at the Research Library. Los Alamos National Laboratory (LANL) to its users that includes categories of scientists, students, staff and external users.

Web personalization research is more akin to Computer Science research domain and is yet to enter into library field widely. Library initiatives are lacking much in this aspect.

CONCLUSION

Web personalization is the process of customizing a web site to the needs of specific users, taking advantage of the knowledge acquired from the analysis of the user's navigational behaviour (usage data), in correlation with other information collected in the Web context, namely structure, content, user profile data and the web authentication of the user. Recommending refined keywords with precision exactly matching the user required information, suitable websites, blogs, high informative source documents, more high productive journals similar but not accessed by the user- all these can bring the user back to the library web track and see that the library service 'Paradise regained' Switch over to the digital environment along with users, identify their navigating movements and transform gradually your tradition based library services into web based personalization and recommender system. Within the wide construct we focus on information literacy 'the ability to evaluate critically the intellectual, human and social strengths and

weaknesses, potentials and limits, benefits and costs of information technologies' (Jeremy, 1966)

Instead of lamenting over the declining rate of library visitors and their ignoring physical documents based services, the librarians can adopt web personalization and recommend ways and means for enhancing the users' hit rates and through user education promote digital competence and digital literacy among users. Web personalization and recommender system is a boon to the community of librarians' to win back their users. Kill the king (of manually operated systems) but keep the tradition (by extending personalized digital information services and library services extended to digital information services) so that the library fulfils its objective.

REFERENCES

Acilar, M. and Arslan, A. 2009. A Collabarative Filtering Method Based on Artificial Immune Netwoek. *Expert Systems With Applications*, Vol.36, no.4: 8324-8332.

Association of Research Libraries. *Research Library Trends: An Introduction*. Available at: http://www.arl.org/stats/arlstat/01pub/intro.html.

Bracke, P. J. 2004. Web usage mining at an academic health sciences library: an exploratory study. *Journal of Medical Association* : 421-428.

Dai, H. and Mobasher, B. 2003. Integrating Semantic Knowledge with Web Usage Mining for Personalization. Available at: http://maya.cs.depaul.edu/~mobasher/papers/DM04-WM-Book.pdf.

Deital, H.M. et al. 1987. *E-Business and e-Commerce for Managers*, New York: Prentice Hall.

Joachims, T., et al. 1997. Webwatcher: Atour Guide for the World Wide Web. *Proceedings of the 15th International Conference on Artificial Intelligence.* Nagoya, Japan.

Srivastava, J. 2000. *Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data.* Available at: http://nlp.uned.es/WebMining/Tema5.Uso/srivastava2000.pdf

Lieberman, H. 1995. Letizia: An Agent that Assists Web Brousing. *Proceeding of the 1995 International Joint Conference on Artificial Intelligence.* Montreal, Canada.

Maurice, D.M., Sarabjot, S.A, and Alex, G.B. 2000. Personalization on the Net Using Web Mining: Intraduction. *Commun ACM*, Vol.43, no.8: 122-125.

Mobasher, et al. 2004. Infering User's Information Context: Integrating User Profiles and Concept Hierarchies. *Proceedings of the 2004 Meeting of the Internation Federation of Classification Societies.* Chicago: IL, 2004.

Mulvena, M., Anand, S.S., and Bucher,A.G. Personalization on the Net Using Web Mining. *Communication of ACM*, 2002: 43(8) 122-125.

Nazri, M., et al. 2009. A Hybrid Approach for Learning Concept Hierarchy from Malay Text Using GAHC and Immune Network. *Artificial Immune Network* : 315-328.

O'Dell, J. 2009. Libraries and the Future of Search. *Library Philosophy and Practice.*

Ranganathan, S. R 1987. *Reference Service.* Bangalore: Sarada Ranganathan Endownment.

Scharnhorst, A. and Wouters, P. 2012. Web indicatiors-a new generation of S&T indicators? *International Journal of Scientometrics, Informetrics and Bibliometrics*, Vol.5, no.20: 10(1).

Secker, A., Freitas, A.A. and Timmis, J. 2008. An Artificial Immune System for Interesting Information Discovery on the Web. *Applied Soft Computing*, Vol.8, no.2: 885-905.

Shapiro, J. and Shelley Hughes, K. 1966. Information Literacy as a Liberal Art, Enlightenment Proposals for a New Curriculum. *Education Review* : Vol.31, no.2.