Modern way for Web Personalization through Web Mining

Rinku Singh[1], Dr. P. K. Rai[2], Kanu Priya Tiwari[3]

[1,2]*Department of Computer Science, A.P.S.University, Rewa (M.P) 486003India*
[3]*E&TC Department, Govt. Engineering College (REC), Rewa (M.P) 486003India*

Abstract— **In this network area, data has been available on the World Wide Web(W3) in an exponential manner and improve the quality of "recommendations". That is why, server concern over load data with over challenges for the users. This is very difficult for users to retrieve the most relevant data as per the user requirement on the web. If user doesn't know, the process to find the data then, it is very time consuming for the user. In this paper, we present a modern way for web personalization that provides a "Recommendation" to users by which they can retrieve their relevant data and save time.**

Keywords— **Semantic Web Personalization, Web Usage Mining, User profile, Collaborative Filtering, Web Usage Logs**

I. INTRODUCTION

Nowadays, the data is rapidly growing in various domains on the web. That is why, these days, the need for identifying and retrieving the data is based on exactly the needs of users in order to improve the usability of a website. To do this, web personalization is defined as any action that always adopts the contents or services provided by a web site to single user or a group of users based on their navigational behaviour, stored in the log file. This data is included with the content, structure as well as user's interests. The web personalization gives the output in the dynamic generation of suggestions to create web pages according to the needs of users, highlighted existing hyperlinks that are required by the users. Most of the earlier research efforts in web personalization deal with Web Usage Mining(WUM). Here, we present pure usage based personalization where insufficient data is available in order to extract pattern. When, the content of the web site is changed and new pages are added but that isn't included in the logs yet. These days, the user's aim is to find the content concerning a required content. Thus the underlined content semantics should be a dominant factor in the web personalization process. There have been many research studies that integrate the contents of web site in order to improve the process of web personalization. Most of the contents are characterized on the web by extracting features from the web pages. Keywords as features are used to retrieve similar content based on the interest of the user.

Semantic web personalization is more effective search response and user satisfaction in this modern generation because most of the users are working with internet using system or Smartphone.

II. BACKGROUND

Normally, the methodologies of web personalization are categorized into two complementary processes -

a) The collection of user information.
b) The inference of the collected data.

Here, in the first process, user profiles are used to enrich queries and to sort output at the user interface level [6] or in other techniques, they are used to inference relationships like the social based filtering [1] and collaborative filtering [9]. In the second process, information extraction on users navigations from system log files can be used [5]. Some information retrieval techniques are based on context extraction [2]. Information semantics are used to enrich the process of web personalization and queries can be enriched by adding new properties from the available domain Ontologies. As we know, the user modelling are based on ontology can be coupled with dynamic update of user profile using output of information filtering and Web Usage Mining (WUM) techniques. Data collected through search engines show that spatial information is pervasive on the web that many queries contain spatial specification, but it is a tough job to search relevant resources which could respond quickly to query including a spatial component. The personalized information can consider spatial property and link found in web documents. Here, three components are required in web applications which are:

a) User model and associated user preferences elicitation mechanisms.
b) Web personalization engine combining spatial and semantic criteria.
c) User interface enriched with spatial components.

Web personalization always requires the representation of features about user that are relevant to the domain. Similarity of semantic and spatial proximity measures as well as relevance function of ranking by the user.

It is the evaluation of semantic link which are existing between two concepts[12]. Classification algorithm can be used for measuring spatial proximity between two regions. Interactive adaptive map generation and visualization are another aspect of spatial web personalization that is concerned with web maps adaptation according to needs of user.

The presented personalization approaches have contributed to the improvement of information systems. These approaches have weakness and limitation despite their wide spread use. In fact, several approaches like, collaborative filtering presents the similar recommendations for all the users within the same cluster. Whereas, content based approaches facilitate items retrieval by recommending similar items to the user. Other approaches of web personalization are trying to determinate the behaviour of each & every user but they are limited by their items model that doesn't describe the differences between properties of items. In most web personalization approaches, spatial aspect isn't taken into consideration that requires an adaptation of approaches to be relevant, while applied to spatial information. The existing approaches are presented as an alternative that improves the quality of personalized systems[3].

Here, web personalization frame work characterizes the usage profiles of a collaborative filtering system using ontology is represented by Dai and Mobasher[4]. The system profiles are transformed into domain level aggregate profiles by representing each page with a set of related ontology objects. Oberleet described, the idea of semantically enhancing the web logs using ontology concepts. This frame work is based on a semantic web site built on an underlying ontology. This site can contain static and dynamic both type of pages, to extract knowledge about group of users, preferences of users and rules. Authors present a general frame work where data mining can be performed on semantic web logs. Since the proposed frame work is built on a semantic web knowledge portal, the web contents are inherently semantically annotated exploiting the portals. This is discussed by authors that, how frame work can be extended using specializations of ontology terms, as well as for supporting the process of web personalization. General personalization frame work based on the conceptual modelling of the user's navigational behaviour is proposed by Acharya and Ghosh[6].

Here, the methodology involves mapping each and every visited page to a topic, imposing tree taxonomy on the concepts and then estimating the parameters of a process of semi-markov which is defined on this tree based on the generated user paths. In this model, the semantic characterization of the context is manually performed.

The paper of Myriam Hadjouni et.al. [12] is about a personalized information retrieval approach based on end user modeling. The proposed approach personalizes data retrieval using implicit user information and interests measurements.

Middleton et.al represented the ontologies uses in the process of user profile within collaborative filtering systems which focused on recommending paper. Here, authors represent the user profiles by the terms of research paper ontology. This process is known as hybrid recommender system which is based on collaborative and content based recommendation taxonomy, the contents are characterized with ontology terms by document classifiers and ontology is used to make a specialization of the user profiles. Kearney and Anand explored ontology to calculate the impact of ontology concept on the user's navigational behaviour. Here, they have suggested that, the impact values can be used to accept more accurate & determine distance between various users or preferences of users and other contents on the web site. K.Sridevi and Dr. R.Umarani have made a survey on the various approaches used by researchers to achieve Web Personalization in Web Mining[19].

III. ARCHITECTURE FOR WEB PERSONALIZATION

To address some of the open issues in the proposed architecture for personalizing the web are given in Fig.1. The architecture uses contents of web site, web logs created by observing the user's navigational behaviour and user profiles created according to the preferences of users to analyze and extract the information needed for the user to find the pattern expected by the user[20]. Recommendation has been created by this analysing that is presented to the user. For this work, Web Usage Mining (WUM) can be defined as automatic discovery of user navigational patterns. It can support decision making process of web site to understand the user completely. However, these taxonomies can be used for functions of web personalization. Characterizing the web content into semantic categories is done for predicting the web pages for single user or group of users.

Making user profiles would be done by collecting information to each user based on their interest or behaviour and other demographic information. To do this, web personalization can be worked to a group of interested users, based on the patterns obtained from Web Usage Mining(WUM), classification of web content and user profiles. Here, web personalization includes techniques such as, use of cookies, history and Machine Learning Strategies (MLS). It may be viewed as a type of recommender system, clustering, classification and prediction of web pages for a user or group of users. By the web personalization, the content of the web pages are modified in a better way for the needs of users. It involves, creating a unique pages each user or using the desires of a user to determine to retrieve the web document.
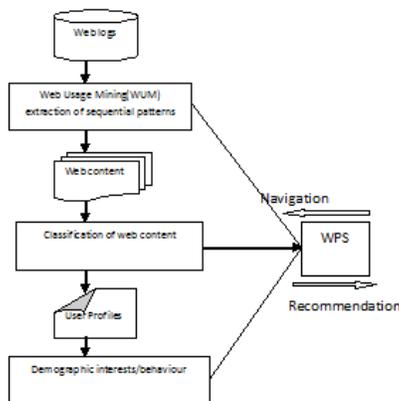

Fig. 1: Architecture of Web Personalization.

3.1 Sources of Data for Web Personalization

The Web Personalization System (WPS) utilizes web data in order to personalize the contents on the web site. These are classified into four categories [20] :

1) *Content data :-* These data can be simple text, images or structured data, such as, information retrieved from data base. They are presented to the end user appropriately structured.
2) *Structured data :-* These represent the way to organized the contents. It can be either data entities used within a web page like, HTML/XML tags or data entities used to set a web site together such as, hyper text or hyper links connect one page to another.
3) *Usage data :-* These data represent a web site's usage, like visitor's IP address, date/time access, path access and other attributes that are included in a web access log.

4) *User profile data :-* These represents information about the users of a web site. Here, user profiles contain demographic information like- name, age, education, interest, behaviour, etc. for each and every user of a web site. These types of information are acquired through questionnaires or registration forms. It can be inferred by analyzing Web Usage Logs(WUL).

3.2 Web Personalization using user profile

A user profile is a visual display or a collection of personal data associated to a specific user. It refers to the explicit electronic representation of a user's characteristics and it can be described as the system representation of a user model. User profile represents previous knowledge description of users and the concept models possessed by users to collect the information on the web. This model is implicitly possessed by users which are generated from their previous knowledge used to collect relevant information about a user preference. Here, user profiles are categorized into three groups [20]:-

1) Interview
2) Semi-Interview
3) Non-Interview

❖ *Interview:-* It considers a perfect user profiles. They are acquired by using manual techniques such as – questionnaires, getting interview to the users, etc. here, each and every user recommended to read each document and give their suggestions freely.
❖ *Semi-Interview:-* In this, user profiles are acquired by semi automated techniques with limited user.
❖ *Non-Interview:-* Here, users are not involved at all, but discover interest of user's. They acquire user profiles by observing activity of user and discovered previous knowledge of user.

User profiles models provide knowledge based on a globally or locally. The global analysis uses existing knowledge to produce effective performance. The generic ontology is included by common used knowledge base such as, wordnet, online libraries, etc. whereas, local analysis use behaviour of user. The user previous knowledge can be discovered by integration of local and global analysis because local analysis is used for analyzing the user behavior[20].

IV. TAXONOMIES USING USER PROFILES

A common way to store information about user profile that enables web personalization is represented in Fig. 2.

This is known as content based filtering which applied to a text document and evaluates the document relevance by the keywords stored in a user profile. It stored basically at the server to secure the profile transmitting through the network. Social or collaborative filtering [11] is another option to take advantage of user profiles. Here, this method collects the user profile of many people and generates recommendations based on the similar profiles. To implement collaborative filtering user profiles must be compared and therefore, the best storage location is available to centralize them at the server. In Fig.1, is representing a user profile that can be shared between various web personalized applications that require the same content of user profiles because web personalized application need to have access the complete set of profiles to store that at the browser.
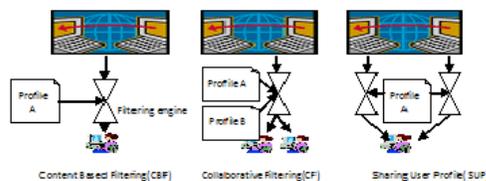

Fig.2 : Different Uses of Profiles

V. SEARCHING BASED ON SEMANTIC WEB PERSONALIZATION

Web personalization is used to search a subset of data on web that always matches the interested user profiles. This can be done by recommendation of web pages or websites to the users and filter option is used to achieved interested web pages by the users[9,11]. By analyzing the history/cookies data, users are able to access the web data and mining the relevant information to a user by clustering previous access web pages based on similar contents. If new web page is found to be similar in any cluster, it can be routed to the user. Personalized search always lakes advantage of semantic web standards to display the content and user profiles. Web data based on semantic personalization can be used effectively for improving the searching methods and by re-ranking the output based on the previous activities by the users. The core part of semantic approach on web personalization is used in the ontology. If web page annotates with ontology entity labels, than, the user can access the web pages with more effective content recommendation.

VI. CONCLUSION

Web personalization is an extremely important technique because it enables to better understand user and to develop online products offers and contents that correspond to what these users actually search. It directly linked to users need and preferences. In this paper, we have explored the modern way to research activity to improve the performance of web personalization process and the performance of Information Retrieval System (IRS) or different uses of profiles.

Here, we have also presented, Architecture to retrieve the personalized web document and different ways to filter the user profiles which shows the filtered contents to the users.

REFERENCES

[1] [1999].Mladenic D,"Text learning and Related intelligent Agents: a survey", IEEE Intelligent systems, 14(4):44-54.

[2] [2000]. B.Mobasher, H.Dai, T.Luo, Y.sung, J.zhu, "Integrating web usage and content mining for more effective personalization", in proc. Of the international conference on ecommerce and web technologies.

[3] [2002]. Burke, R,"Hybrid recommender systems: Survey and Experiments, user modelling and user adapted interaction", 12(4): 331-370.

[4] [2002]. H.Dai, B.Mobasher, "Using ontologies to discover domain level web usage profile", in proc. Of the 2nd workshop on semantic web mining.

[5] [2003]. M.Eirinaki, M.vazirgiannis, "Web mining for web personalization", ACM transactions on internet technology.

[6] [2003]. S.Acharyya, J.Ghosh, "Context sensitive modelling of web surfing behaviour using concept trees", in proc. Of the 5th WEBKDD workshop, Washington.

[7] [2004]. M.Albanese, A.picariello, C.sansone, L.sansone, "A web personalization system based on web usage mining technologies".

[8] [2004]. Paulakis, S.Lampos, C.Eirinaki, M.vazirgiannis, M.sewep, "IT web mining system supporting semantic personalization", in "15th European conference on machine learning and 8th European conference on principles and practice of knowledge discovery in data bases, LNCS, vol.3202,pp.552-554.

[9] [2005]. H.dai and B.Mobasher, "Integrating semantic knowledge with web usage mining for personalization", in web mining: Application and techniques, A.scime, ed.hershey, pa, sa: IGI global, pp.276-306.

[10] [2006]. Jiawei Han and Micheline kamber,"Data mining:Concepts and techniques", 2nd ed, Morgan Kaufmann publishers, ISBN 1-55860-901-6.

[11] [2009]. Xiaoyuan su and Taghi M. Khoshgoftaar, "A survey of collaborative filtering techniques", Hindawi publishing corporation advances in artificial intelligence, article id 421425,19 pages.

[12] [2009]. Myriam Hadjouni, Mohamed Ramzi Haddad, Hajer Baazaoui, Marie-Aude Aufaure, and Henda Ben Ghezala, "Personalized Information Retrieval Approach", Proceedings of WISM 2009

[13] [2010]. WANG yong.gui and JIA zhen, " Research on semantic web mining", international conference on computer design and applications.

[14] [2011]. C,Ramesh, Dr.K.V.chalapati rao, Dr. A.Goverdhan, "A semantically enriched web usage based recommendation model", international journal of computer science & information technology(IJCSIT), vol.3.

[15] [2011]. Mahendra thakur, yogendra kumar jain, Geetika silakari,"Query based personalization in semantic web mining", International journal of advanced computer science and applications(IJACSA), vol.2.

[16] [2012]. Bhaganagare ravi Shankar, Dharmadhikari dipa, "Web personalization using ontology: a survey", IOSR journal of computer engineering(IOSRJCE) ISSN:2278-0661,vol.1.

[17] [2012]. Chhavi rana, "Trends in web mining for personalization", IJCST vol.3.

[18] [2012]. Sneha.y, S.G Mahadeven, Madhura prakash, "A personalization product based recommendation system using WUM and semantic web", international journal of computer theory and engineering vol.4.

[19] [2013]. K.Sridevi and Dr. R.Umarani, "Web personalization approaches: A survey", International journal of advanced research in computer and communication engineering, vol.2., issue 3, pp 1533-1539

[20] [2015]. Doddegowda B J, Dr. G T Raju, "A Novel Approach for Web Personalization through Web Mining Techniques", COMPUSOFT, An international journal of advanced computer technology, 4 (5), May-2015 (Volume-IV, Issue-V), pp 1753-1759