Abstract—Web personalization is the process of customizing a Web site according to the needs of specific user. It takes advantage of the knowledge acquired from the analysis of the user’s navigational behavior (usage data) in correlation with other information collected in the Web context, namely structure, content and user profile data. Due to the explosive growth of the Web, the domain of Web personalization has gained great momentum both in the research and commercial area. A lot of research has been done in this field. This paper is comprised of three parts. First part gives introduction to Web Personalization. Second part tries to summarize the work done in this field and third part does research analysis. This paper tries to throw light on the research areas that are required to be explored.

Index Terms—Web Personalization, Customization, Recommendation System, Ranking System, Wisdom Web

I. INTRODUCTION

With the dramatic, quick and explosive growth of information on the Web, it seems that Web users are drowning in the ocean of information. They are facing the problem of information overload. Apart from the diversity in information, there is diversity in the way users navigate web. A major drawback of generic search engines is that they follow same model and are not adaptable to the needs of an individual user. Personalization of web is a solution to such problems as it can provide different search results based on preferences and information needs of the users.

II. WEB PERSONALIZATION

Web Personalization is a technique to provide users with the desired information they require, without expecting from them to ask for it. Personalization requires implicitly or explicitly collecting visitor information and leveraging that knowledge in Content Delivery Framework. It then analyzes this information and decides which information should be provided to user and how it should be presented [1].

Different users have different backgrounds and interests. They may have completely different needs and goal when issuing the same query. For example, the query issued “trees” by a Botanist requires totally different information as compared to the same query issued by a programmer. Such queries are termed as ambiguous queries. Mostly, these types of short and ambiguous queries are issued by the users. If the user does not get information related to his interest then he may get frustrated.

Personalization should not be confused with customization. It is different from customization. In case of customization user is given questionnaire and then based on his preferences web page is constructed. Hence customization is dependent on user’s input where as personalization does not expect from a user to give directions. Personalization mechanism is based on explicit preference declarations made by the user. Iteratively it monitors the user’s navigation, collects its request of ontological objects and stores them in its profile in order to deliver personalized contents [2]. Web Personalization is divided into four distinct phases as follows:

A. COLLECTION OF WEB DATA:

In this phase, Web data is collected. It can be in implicit or explicit form. Implicit data includes past activities/click streams which are recorded in Web server logs and/or via cookies or session tracking modules. Explicit data usually comes from registration forms or rating questionnaires. Apart from this data, Web content, structure and usage data can also be included to give extra knowledge that can be helpful in next phases.

B. PREPROCESSING:

In this phase, Web data that is collected from previous phase is preprocessed to put it into a format compatible with the analysis technique to be used in the next step. Preprocessing may include cleaning data of inconsistencies, filtering out irrelevant information and completing the missing links in incomplete click through paths.

C. ANALYSIS OF DATA:

In this phase, automatic user profile is created after analyzing data processed from the previous step. It is generally done offline so that no extra burden is added on the Web server. This step applies Machine Learning or Data Mining techniques to discover interesting usage patterns and statistical correlations between Web pages and user groups.

D. RECOMMENDATION PHASE:

The last phase in personalization makes use of the results of the previous analysis step to deliver recommendations to the user. The recommendation process typically involves generating dynamic Web content on the fly, such as adding hyperlinks to the last web page requested by the user. This can be accomplished using a variety of Web technology options such as Common Gateway Interface(CGI) programming.
Fig. 1 Phases of Web Personalization

Fig. 1 shows the different phases of Web Personalization. There are numerous strategies for Web Personalization. Developers can use any of these or in combination depending on the personalization requirements. Different strategies used for Web Personalization are as follows:

A. MEMORIZATION:
Using this strategy, user information such as name and browsing history is stored using cookies. It is later used to recognize and greet the returning user. It is usually implemented on the Web server.

B. CUSTOMIZATION:
This form of personalization takes as input a user’s preferences from registration forms in order to customize the content and structure of a web page. This process tends to be static and manual. So it is called semi-automatic. It is usually implemented on the Web server. Typical examples include personalized Web portals such as My Yahoo and Google.

C. RECOMMENDER SYSTEMS:
Recommender systems automatically recommend hyperlinks that are deemed to be relevant to user’s interest. It is usually implemented on the Web server. It relies on data that reflects the user’s interest implicitly or explicitly. Implicitly refers to browsing history as recorded in Web server logs. Explicitly refers to user profile as entered through a registration form or questionnaire.

D. TASK PERFORMANCE SUPPORT:
In case of Task Performance Support systems, personal assistant executes action on the behalf of the user in order to facilitate access to the relevant information. This approach requires heavy involvement on the part of the user, including access, installation, and maintenance of the personal assistant software.

III. PREVIOUS RESEARCH
Though Personalization of user data is a recent phenomenon due to the rise of Internet usage, but over the years Web mining techniques have developed a full fledged support to provide personalized experience for user. Lot of work is done to improvise Web Personalization. The two important surveys done by [1, 3], gave a complete overview of the work done in this sphere up to 2003. One of these survey [1] presented the uses of Web Mining and Web Personalization and other highlighted Web usage mining as a tool for personalization [3]. Another survey done by Chhavi Rana (2012), gave a precise and comprehensive understanding of the work done in this sphere from 2007 to 2012. This paper reviews usefulness of different approaches along with the projects associated with some of the techniques [4]. After that there has been lot of up heals that are encountered in this area of Web Personalization. This paper tries to summarize the work done in various subfields of Web Personalization after 2012.

With the dramatic increase in the number of websites on the Internet, tagging has become popular for finding related, personal and important documents. Onur Yilmaz (2013), presented a tag-based website recommendation method, where similarity measures are combined with semantic relationships of tags. This approach performs well in recommending new websites or catching user's current interests. However, there no control on the tags provided by users in this system. Although users do not intend to mislead the method while tagging websites, different purposes of tagging can create confusion [13].

Su, Chang and Tseng (2013) present music recommendation system that utilizes social media tags instead of user rating to calculate the similarity between music pieces. Through the proposed tag-based similarity, the user preferences hidden in tags can be inferred effectively. The empirical evaluations on real social media datasets reveal that proposed approach in this paper using social tags outperforms the existing ones which are using only ratings in terms of predicting the user’s preferences to music [7].

A. Vaishnavi (2011) proposed a technique for developing Web Personalization system using Modified Fuzzy Possibilistic C Means (MFPCM). The author claims that this approach raises the possibility that URLs presented before a user will be of his interest [6].

Another area that has been explored by many researchers is Research Paper Recommendation System. Different researchers have used different approaches. Hong, Jeon and Jeon (2013) proposed Personalized Research Paper Recommendation System PRPRS (PRPRS) that designed expansively and implemented a user Profile-based algorithm for extracting keyword by keyword extraction and keyword inference. PRPRS calculates the similarity between given topic and collected papers by using Cosine similarity which is used to recommend initial papers for each topic in the information retrieval [8].

Lee, Lee and Kim (2013) have proposed Personalized Academic Research Paper Recommendation System (PARPRS) using collaborative filtering methods, which recommends related articles, for each researcher that may be interesting to her/him [9]. Gao et al. (2014) have suggested novel collaborative filtering approach that will make item-based CF more robust [12]. Jiang et al. (2013) have proposed two novel boosting frameworks for collaborative filtering [15]. Jing Lu et al. (2013) investigated Online collaborative filtering techniques for building live
recommendation system [16]. Nikos and Katrien (2013) have discussed Layered evaluation applied for multi-criteria recommendation service. It could be deployed for paper recommendation using Mendeley dataset [14]. Mendeley dataset is in the form of Mendeley Research Catalog with a collection of 80 million research documents.

Bedi and Aggarwal (2013) presents Aspect-Oriented Trust Based Mobile Recommender System (AOTMRS) that uses the concept of trust and Aspect Oriented Programming for advice-seeking and decision-making process similar to real life. The proposed system AOTMRS builds a mobility aspect and generates the trustworthy recommendations based on the user preferences and his demographic information such as location, time, need etc [10].

Preeti, Ankit and Purnima (2014) emphasize visualization in recommendation systems. It proposes an argument based recommender system which uses hybrid approach in which two results: trustworthy users and arguments occurred between user agent is visualized using D3 tool [11].

Pan and Chen (2013) propose a new Group Bayesian Personalized Ranking (GBPR) system for one-class collaborative filtering or collaborative ranking which models user’s ranking-related preferences more directly [17].

S.Geetha rani (2013) suggests click count as well as link-click based Ranking algorithm. In this algorithm the count of click of each query concept can be evaluated as well as the link is also evaluated in the submitted query [18]. Wang et al. (2013) explained General Ranking Model Adaptation Framework for personalized search is explained [19]. Global RankNet model which is widely adopted ranking model for web search tasks is discussed in another paper [20]. They have suggested improvements in this model.

One of the subfield of web personalization is e-learning. Personalization of e-learning is considered as a solution for exploiting the richness of individual differences and the different capabilities for knowledge communication. In particular, to apply a predefined personalization strategy for personalizing a course, some learner’s characteristics have to be considered. Furthermore, different ways for the course representation have to be considered too. Fathi, Leila, Mohammad (2013) have studied solutions to the question: How to automate the E-learning personalization according to an appropriate strategy? The study is about finding an answer to this original question by integrating the automatic evaluation, selection and application of personalization strategy [21].

D.Suresh, S.Prakasam (2013) suggest the instructors to use the combination of e-learning system using Rank-Based Clustering Algorithm (ESURBCA) to get consistency in content delivery, quality content in learning materials, students self-learning concept and performance improvement in their examination [24].

Diana Butucea (2013) presents a theoretical framework for visually impaired persons, followed by a technical implementation of the concept in relation with the e-learning context. The solution proposes an analytical approach over the computer aided learning mechanism. It defines the concept of personalized learning and provides an example of implementation for a software system. This example subsequently offers support and assistance for visually impaired computer users [25].

Yarandi, Jahankhani & Tawil (2013) present an ontology-based approach to develop adaptive e-learning system based on the design of semantic content, learner and domain models to tailor the teaching process for individual learner’s needs. The proposed new adaptive e-learning has the ability to support personalization based on learner’s ability, learning style, preferences and levels of knowledge [22].

Cakulaa and Sedleniececa (2013) have aimed to identify overlapping points of KM and e-learning phases to improve the structure and transfer of personalized course knowledge using effective methods of ontology and metadata standards. This research offers a theoretical background of knowledge management principle implementation for the development of a practical personalized e-learning model [23].

Another major application area of personalization is e-commerce. Personalized recommendation technology in e-commerce is widespread to solve the problem of product information overload. However, with the further growth of the number of ecommerce users and products, the original recommendation algorithms and systems face several new challenges such as modeling of user’s interests more accurately, providing more diverse recommendation modes and supporting large-scale expansion. To address these challenges from the actual demands of e-commerce applications, Dong et al. (2013), designed and implemented a personalized hybrid recommendation system, which can support massive data set using Cloud technology [26].

World Wide Web has undergone three generations from Information Web to Social Web to Semantic Web. It has started its journey towards the fourth generation which expects wisdom from the Web and so termed as the Wisdom Web. In present era, where computers and Internet has become inseparable parts of our life, user wants the Web to sense their requirements and interests and serve the contents accordingly. Search engines play major role in information extraction and delivery and present models of search engines that are still struggling for providing personalized information to the users. Aarti Singh and Basim Alhaididi [27] present knowledge oriented personalized search engine framework which can provide personalized contents to its users. This framework provides a direction for the next generation of WWW and contributes towards Wisdom Web. In another research paper Aarti Singh and Dr. Singh (2013) highlight the technologies contributing towards the next generation of WWW and also suggest future direction for Web Personalization [28].

Another upcoming area is being suggested by Hanak et al. (2013). They have suggested methodology for measuring personalization in Web search results. They have also applied this methodology to 200 users on Google Web
Search and have deducted the results. The causes of personalization on Google Web Search are also investigated. There effort is a step towards understanding the extent and effects of personalization on Web search engines today [29].

IV. RESEARCH ANALYSIS

The wealth of Research projects described in the previous section indicates that the demand for better Web personalization is high. However, many challenging research problems must be addressed if this demand is to be fully met. Issues that cut across all of the applications are henceforth described, where progress will consequently have the broader impact. The goal of personalization is to provide users with what they want or need without requiring them to ask for it explicitly. Following are listed some of the major issues that needs to be catered to provide a better personalized system.

Research on recommendation system is an emerging field. Although it has obtained very good results, but there are still many problems that need to be handled. One of the problems faced is Data acquisition. It is mainly dependent on the user's explicit evaluation. If user does not know about the product it will not be helpful. Another problem that is being faced is Cold start problem also called first evaluation problem. If a new product has no evaluation, then the product will not get a recommendation, hence recommendation system will lose action. Ping, Xiang and Ming (2012) have suggested that the future recommendation research should use the web mining technology to collect the user’s implicit browsing information rather than depending on explicit evaluation information supplied by the user for some product. At present, the recommendation system needs the user to make preference assessments for a product but this method has low degree of automation and many users do not evaluate the product. Web mining technologies can improve the automation degree of the collected information.

Another angle for future research has been suggested by Xujuan et al. (2012). Social Networking Sites contain a warehouse of information. It can be mined and analyzed to expand user profiles. It can be used to build complex diagrams and maps of user-to-user and user-to-interest relationships. A lot can be understood about the user using these websites. The research issues on how to make breakthrough on the current recommender system for social networking environment or how to build the trust-based Web personalized recommender systems need to be explored further [31].

Another area which has been researched is Research Paper Recommendation System. Different authors have given different approaches. Some have used Content Based Filtering and some have used Collaborative Filtering. Stereotyping is one of the oldest approaches that are successfully applied by Yahoo. Some authors have used item centric approach and some have taken help of graphs. One generalized approach is required which could handle various issues related to research paper recommendation. Despite the large number of research articles, there is only a hand full of active recommender systems, and most of them apply simple recommendation approaches that are not based on any recent research results. So there is a need of research paper recommendation system which incorporates all the latest research work done in this field.

Another future scope can be to develop frameworks for research paper recommender system evaluations. This should include an analysis of how suitable offline evaluations are, to what extend datasets should be allowed to be pruned, how many participants user studies shall have at minimum and which factors influence the results of evaluations (e.g. user demographics).

One of the latest areas under Web Personalization is Wisdom Web. WWW is on a journey of wisdom and maturity, where it can sense what user wants and serve relevant contents on its own. Transformation of WWW to Wisdom Web requires change in the way web is accessed and therefore it requires new architecture. A new model of web personalization search engine, which is capable of providing knowledge, based recommendation to an individual is suggested by Singh & Allhadidi (2013). Its implementation and evaluation is still in pipeline. Further new approaches can be adopted and developed [27].

Another area of measuring personalization of web search can be explored. Over the past few years, we have witnessed a trend of personalization in numerous Internet-based services, including Web search. While personalization provides obvious benefits for users, it also opens up the possibility that certain information may be unintentionally hidden from users. Despite the variety of speculation on this topic, to date, there has been little quantification of the basis and extent of personalization in Web search services today. Efforts should be made to provide transparency for users of Web search and other Web based services. Measurement of personalization is in its budding stages. New methodologies can be investigated, generalized and applied on various Web Search engines and Web sites.

Personalized e-learning model using effective methods of ontology can also be explored. Dynamic user modeling is another upcoming research area. Few systems have attempted to handle the dynamics within the user profile. The behavior of users varies over time and it may affect the construction of models. In nutshell a Web Personalization system should be able to adapt to the user’s behavior.

V. CONCLUSION

Web personalization is the process of customizing a Web site to the needs of a specific user. Web personalization is a very important, if not necessary, part of WWW today with many of its application areas. This paper tries to present a comprehensive description of the various web personalization efforts made in the last three years. Though, a lot of research has been done since the last decade in this area and many researchers have addressed the problem of information bombardment, yet a system that can effectively integrated various diverse requirements of the users has not yet been proposed. This paper tries to identify potential directions, and provides at the same time a quest for exploration in the area of Web Personalization.
REFERENCES


