



International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 10 • Number 14 • 2017

A Survey on Web Personalization Approaches through Web Mining

Kamika Chaudhary¹, Neena Gupta¹ and Santosh Kumar²

¹ Department of Computer Science Kanya Gurukula Campus(Gurukula Kangri Vishwavidyalaya)Haridwar, Uttarakhand India, Email: kamika.agrohi@gmail.com , neena71@hotmail.com

² Department of Computer Science & Engineering Krishna Institute of Engineering & Technology, Ghaziabad, Uttar Pradesh, India, Email: santoshg25@gmail.com

Abstract: Web personalization is the process that tailors the content of the web site according to user requirement by using the knowledge acquired from user navigational behavior on the site. Web personalization is found to be advantageous in decreasing the information overload which is one of the most vital problems. This has been seen through the prism of providing an aid to search engine for efficient information retrieval. Nowadays this has become an important domain and is being examined from both research front and commercial front. A survey that introduces how web mining and its various forms assist in achieving personalization along with its various approaches, functionalities and applications is presented in this paper.

Keywords: Web Personalization, Web Mining, Information Retrieval

1. INTRODUCTION

The ceaseless development in the size and utilization of World Wide Web forces new techniques for outline and advancement of online data administrations. The substance on the Web in different fields is quickly expanding and the requirement for distinguishing and recovering the content precisely in light of the requirements of the clients is more than required. An efficient tool is required to access the web page which matches with the request of the user. Search engine is such kind of information retrieval tool that acts as a link between user request and response provided to him by World Wide Web. As the user hits a query in web search engine, it produces the results in the form of hyperlinks which ranges from several hundreds to thousands and arranged in the form of search engine result pages. Now it becomes a tedious task for user to scroll down through all the web pages in search of his most relevant document matches with the query. To minimize all these difficulties personalization came as a rescuer. Most Web structures are extensive and entangled and clients frequently miss the objective of their request, or get uncertain results when they attempt to explore through them. In this way, the necessity for foreseeing client needs keeping in mind the end goal to enhance the ease of use and client maintenance of a Web webpage can be tended to by personalizing it.

Personalization is defined as a task that reorganizes the information provided by the web site according to the requirement of the user by utilizing the information gained from user navigational behavior along with content and structure of web site. Personalization important step is to collect the user information explicitly or implicitly and then using that information to predict which content will be of importance to user and the deliver accordingly [12]. Personalization process modifies the content of the web site according to the user preference. The outcomes of the process of personalization may be seen in the form of suggestions provided by the pages to the user, spotting the important hyperlinks and creation of user specific web pages [20].

Most of the time the term customization and personalization are used interchangeably. It is important to distinguish between them at this point of time. Customization of web pages handles the reorganization of the web sites according to user preference. In this process users are the master of the process as customization is performed manually or semi automatically. Whereas on the other hand personalization is a complete automatic process that provides the user tailored content dynamically. As a result customization is considered as a static process and personalization is dynamic in nature [11].

The rest of the paper is organized as follows. Section II is literature review that describes about the work done in the past about personalization Web Personalization and its approaches are explained in section III. Various functions performed by personalization are explained in section IV and conclusion is given in section V.

2. LITERATURE REVIEW

The personalization of web focuses on the individual interest of the user. The content present on the web is moulded according to the taste of the web user. Web personalization has gained a lot of importance in E-commerce. As nowadays the business has turned itself into electronic form so the competition among organisation has also increased. In order to deliver the most valuable customer satisfaction, personalization of the website acts as a trump card. It has become beneficial for both the user and the organisation. From user perspective it helps in improving the user experience on the web site and from organisation it helps in learning the customer requirements in much better ways as well as also plays an important role in knowing about the future trends and increasing the customer loyalty towards the products and services. The importance of web personalization in E-Government services has been described by the authors in [1] and they have also emphasized on how the users will react to acceptance of information technology in E-Government domain. Behavioural intention of user is conjoined with the personalization and its influence on the user is studied by incorporating certain modifications in the UTAUT2 model. A theoretical model that studies various success measures in web personalization is proposed in [2]. Web developer used three strategies to provide the personalized content on the web that includes: gaining data through customized Site, giving individual points of interest on the customized Website (e.g. filling on the web frame for enlistment on web based business Website), and exploring through customized Site. One feature that the personalized system needs to be compliant with is the privacy feature of user. As personalization deals with personal information of the user, sometime user did not feel comfortable in providing their private details. So in this situation they are not able to take the full advantage of personalization. A novel approach has been proposed in [3] which are more flexible, scalable and systematic towards privacy concerns of the users. The model proposed is PLA based and concentrates on the architectural aspect of the privacy and does not hold any control on the method of collecting the user data. Another personalization model specific to work on personalizing the user query results has been introduced in [4]. Authors have presented a formal scoring model based on generated languages that can incorporate the user interest in various kinds of probabilistic scoring model. A semantic search engine model, NAGA has been used as standard model and enhancements are made on this model. The evaluation performed showed the enhanced results. The concept of ontology has been introduced in the process of web personalization [5] to make its results more semantically enhanced and improve the recommendation results. A concept graph has been used for ontology and it keeps on updating with every new session added. This model of ontology based conceptual graph compared with other flat models and experimental

results proved that it is better. It has been observed that context of the user behaviour varies from one visit of web page to another. Therefore a need of more adaptive and flexible mechanism for personalization has aroused. Authors have developed personalization architecture as a part of IKUM project [6] which utilizes web mining, personalization, content management to provide more suitable recommendations. A survey on how web usage mining is used for incorporating personalization has been done in [7]. Web Usage mining serves as a very important basis for personalization travelling through the road of data collection then pre processing it and finding the patterns and performing post processing of knowledge. Based on web usage mining a framework has been proposed in [8] which is specific to provide automatic recommendations in e-learning systems. In this framework recommendation is performed in two phases: first is modelling phase which is performed offline and second is recommendation phase that is performed in real time. In modelling phase pre processing of data is performed and several operations results into the generation of learner and content models then these models are used to provide the recommendations according to the student needs. Two metrics precision and recall have been used for evaluation of framework and it has been found that recommendation threshold reaches is better than previous strategies used. Another important factor that is considered to be of importance in personalization system is scalability. To achieve this, a framework based on association rule mining based on usage data has been proposed in [9]. To recommend the items a data structure have been used to store the frequent item set and then these frequent item sets serves as an input to recommendation engine. The output in the form of recommendations is generated by matching the current user activity with the discovered patterns. Experimental evaluation has been performed on real data sets and the proposed approach is found to be more effective than previously used collaborative filtering approaches. One of the drawbacks associated with the collaborative filtering approach has been their dependence on the user feedback for providing recommendations and also the data sparsity problem. To deal with both the above problems a cross domain triadic model has been proposed by authors [10]. The triadic factors that are considered include user-item-domain relationship. Triadic relationship helped in not losing the domain specific information and the experiment was performed on two real world datasets and the result found to be more valuable. Nowadays it has been found that web personalization has been gaining a great importance not only in research area but also in enterprises. A survey on how web mining is used for achieving personalization has been presented in [11]. The main steps include profiling of user, analysis of logs, acquisition of information, management of content and web site publishing. The detail of all these steps and tools used for providing personalization has been discussed by author in his study.

3. WEB PERSONALIZATION AND ITS APPROACHES

The thought for personalizing the web content arises to eliminate the problem of information overload on World Wide Web. It deals with delivering only that data which fulfils the needs of user. By doing so it helps in achieving the customer loyalty towards a particular web site. Personalization performs the function of providing a user friendly environment to customer and makes a strong and trustworthy relationship between customer and web sites [7]. Due to the importance of personalization a number of approaches have been proposed to achieve the personalization. With the perspective of architecture and algorithm web personalization systems falls into five different categories such as rule based systems, collaborative filtering systems, content based systems, hybrid personalization systems and web usage mining based systems. These approaches helps in achieving the web personalization and a brief overview of these are explained in this section.

2.1. Rule based Filtering systems

Rule based personalization system depends on the decision rules for recommending items to users. Decision rules can be generated manually or automatic. A list of questions is asked by the user based on demographic, psychographic and other details. In this rules are utilized and act as deciding factor for presenting personalized content to users. The variety of decision rules depends upon the type of organisation, sometimes they are domain dependent. System engineers are responsible for constructing domain specific rule base. The disadvantage of

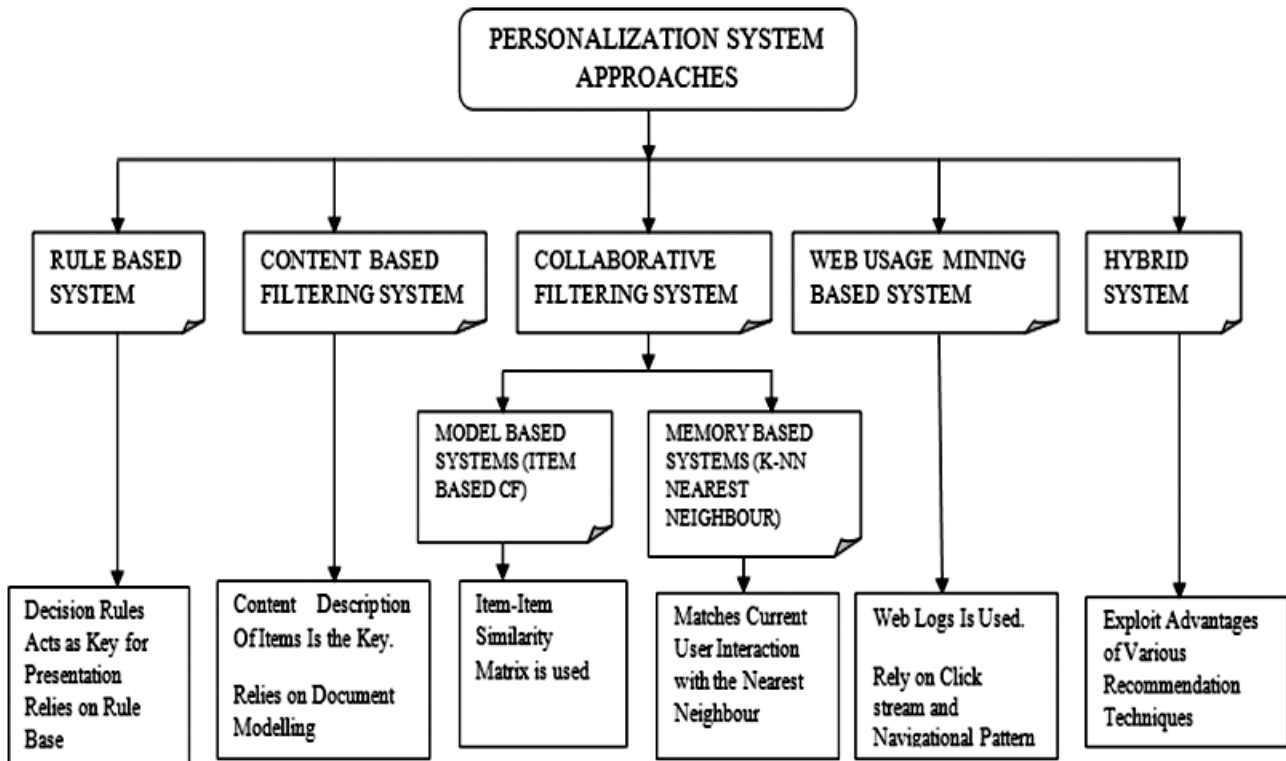


Figure 1: Approaches of Web Personalization

this type of systems is that as demographic details are considered for rule base construction, they are found to be less accurate. Another bottleneck is related to knowledge engineering and static user profiles [7].

2.2. Content based Filtering systems

Content based filtering is an information retrieval technique that performs the prediction about user interest according to his last time behavior on the web site [12]. This technique assumes that users past behavior will serve as a building block for his future interest. It makes the decision about user behavior current behavior by considering the content accessed by him in his previous interaction with the web site. This approach is usually applied on text based features extracted from web pages or description of the products. It depends upon user's expression of interest in any item. Document modeling techniques serves as a basis for this in these systems. It performs by matching the similarity between item based profile and content based user profile and results into recommendation of top n items that has highest similarity index [13]. The limitations with which these systems suffer with include the consideration of only textual features by the system while the web is unstructured and heterogeneous in nature. Other areas in which these systems are limited to are new user issues and overspecialized recommendations [14].

2.3. Collaborative Filtering systems

Collaborative filtering based recommendation systems are considered to be one of the most successful recommendation system [19]. These systems provide the recommendation on the basis of the interest of the neighbor of the customer having the similar likes about the products. It works by forming a database that contains the rating of customer about a long list of products [22]. It then computes the similarity between users and then recommends the product to a target user on the basis of interest of a similar user or neighbor. The advantage of this type of system is their freedom from considering the features of item. Collaborative filtering

systems are divided into two categories: memory based system and model based system. Memory based system depends upon measuring similarity between target user and historical user. A number of similarity measures are used to establish the relationship among target user and historical user. One of the main techniques used to achieve this is k nearest neighbor. It is suitable in the cases where prediction is performed on the dynamic data. Model based approach depends upon machine learning algorithms which are clustering, classification and dimensionality reduction. This approach works by constructing the model which simplifies the relationship between user and item and model is applied when prediction is required to be made [12]. Two main problems with which the collaborative approach suffers includes sparsity based problems and scalability problem. Many other approaches have been proposed to overcome these limitations.

2.4. Web Usage Mining based systems

The field of Web Mining is encouraged from the Data Mining. Web Mining makes use of various techniques of data mining on to the data present on websites of web [18]. In this web is scanned and exploited in search of useful patterns which can be used to further enhance the user experience on the web. Web Mining is categorized into three different parts: Web Content Mining, Web Structure Mining and Web Usage Mining [16]. Web Content Mining is defined as extraction of web pages on the basis of content present on the web pages. In this the data in various forms such as text, image, audio, video, hyperlinks etc are considered as raw information for extraction. Web Structure Mining considers the structure of the web site as the main source data. It discovers the information underlying link structure on the web. Web Usage Mining concentrates on the pattern in which the web is used by user [21]. This usage pattern is saved in the log files and various other sources such as server access logs, proxy server logs, cookies, browser logs etc of the web and then these log files are utilized as the raw material for usage based extraction [17]. The logs of web server contain all the information regarding usage of the web. Conventionally Web Usage Mining was established to fulfill the aim of achieving the decision making by market analyst who

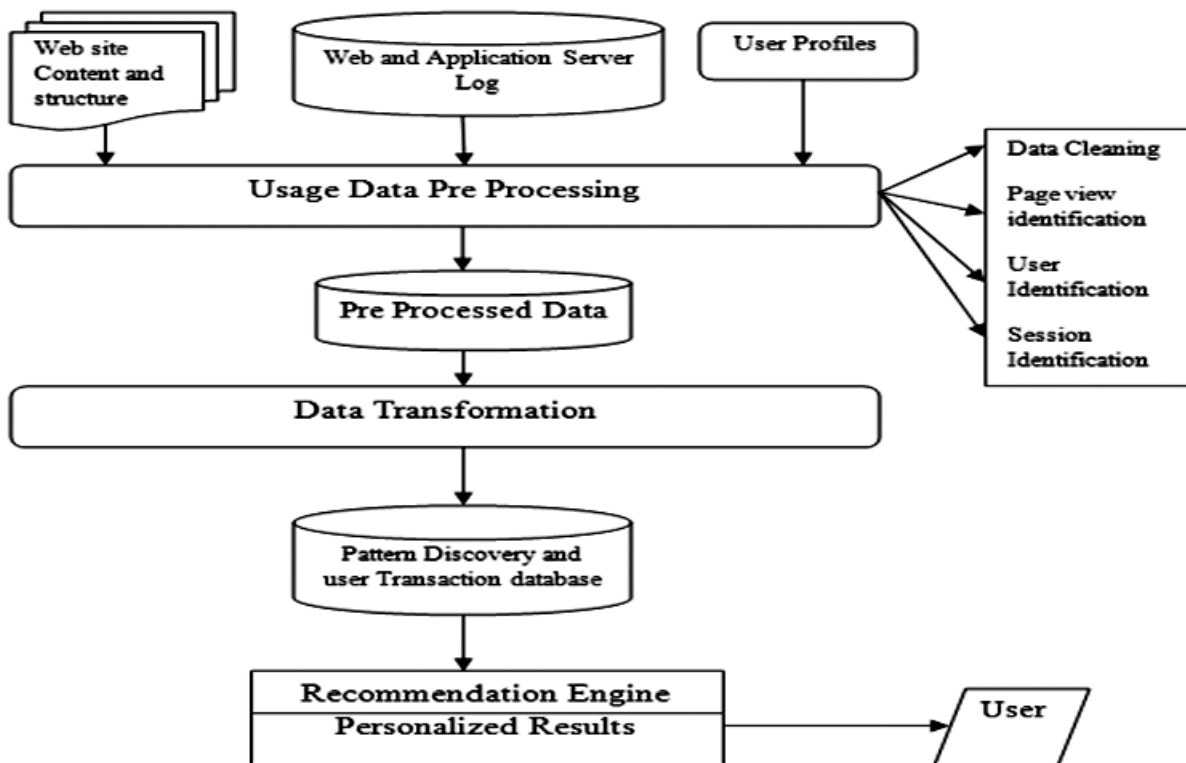


Figure2: Web Usage Mining based Personalization

strive for business intelligence or by web site administrator to reorganize their sites and improve the user experience of the site. Later it has been found that web usage mining can serve as a very important mechanism for personalization. It serves as a very relevant tool for achieving the personalization. Web usage mining based personalization [12] begins with the collection of web data from web logs and user profiles. Then the pre-processing phase begins that result into much cleaner and processed data. After that logs are analyzed and pattern discovery and evaluation is performed. Till this phase all the working is performed in offline mode. Finally the recommendation is performed online and in real time to mediate between the user and web. The overall process of web usage based personalization is shown in figure 2.

2.5. Hybrid based systems

Several approaches that has been used for personalization consists of their own pros and cons. So more than one approach is taken and their advantages are exploited in the case of hybrid approach for personalization. Most of the time a combination of collaborative filtering and content based approach is taken into consideration [23]. There are seven categories in which hybridization techniques are divided [15]. First is weighted in which the weight of the prediction made by each approach that has been taken is calculated and then final prediction is made by considering all the weights. Second is switched in which more emphasis is given to one technique at a time for making the predictions. Third is known as mixed which makes separate recommendation of the item on the basis of all the techniques used. Fourth is feature combination in which a single algorithm for recommendation is developed by utilizing all the approaches those have been considered. Fifth technique is cascade in which the outcome of one technique is fine tune by another technique. Feature augmentation is the sixth technique where result of one approach is serves as input to another. Last is Meta level in which an overall model is generated by one recommendation technique and than that model is used by another recommendation technique.

3. FUNCTIONS OF WEB PERSONALIZATION

Broadly there are four generic categories in which the functionality of personalization can be classified into [7]. The range of functions performed can begin from simply providing user salutation to more complicated delivery of personalized content. Overall the task performed by personalization can be divided into four classes i.e. memorization, guidance customization and task performance support and a hierarchy of classes and their examples is shown in the fig3.

Memorization- This class of function deals with simply remembering the user information by storing in its memory. The information can be user browsing history or its name and when user returns to the site this information is utilized for depicting the user interests. Examples of memorization includes first is user salutation in which web site welcomes the user by displaying his name along with a welcome note, second is bookmarking in which web site bookmarks all the pages that have been accessed by user in the past and display the bookmarked page to him in present and third is personalized access rights for distinguishing among authorised user and general user.

Guidance- This is the second class in the hierarchy of function and this helps the user by recommending his interest pages quickly. It provides the direction to the path of reaching to needed information. The examples include Recommendation of hyperlinks and user tutoring which is provided by adding explanatory stuff to the web site.

Customization- This is one of the most important functions which are performed by personalization. This works by transforming the web site in the form of content, structure and usage according to the will of the user. There are five examples of customization first is personalized layout which changes the way how a web site looks by making transformation in its colour, layout or locale information. Second example includes customization of content in which the length or description of content present on the web site depends upon user interest in the

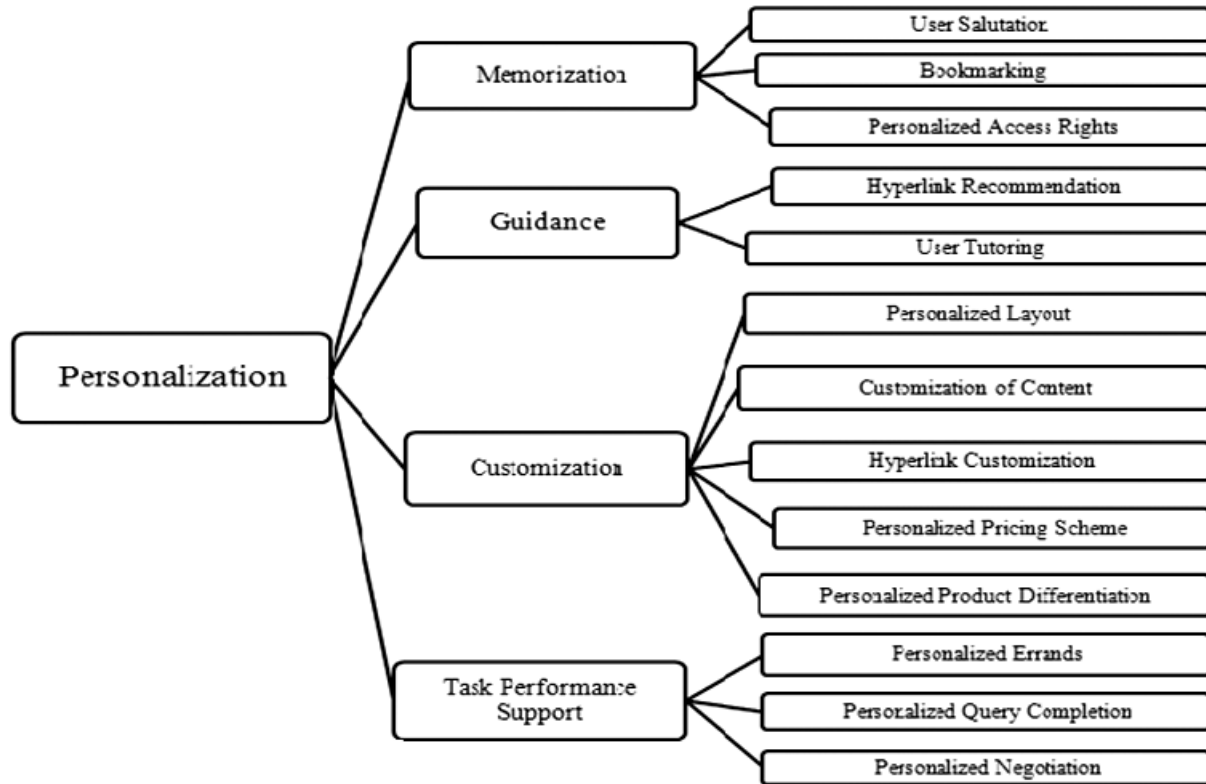


Figure 3: Functions of Web Personalization

content. Third is customization of hyperlinks in which hyperlinks are added or moved according to user which also affects the overall structure of the web site. Fourth example is related to personalize pricing scheme in which your loyalty with a web site can win you the discount from the company and fifth example deals with modifying a product as a specialized solution for user and this is known as personalized product differentiation.

Task Performance Support- In this category of function personalization is so modernized that sometimes it can work on user behalf also in performing some routine task like sending an email etc. Its example includes personalized errands which deal with doing simple task such as downloading an item to more complex task. Personalized query completion helps in improving the information retrieval by completing the query presented by user and last example which is personalize negotiation requires a high degree of sophistication to participate in place of user in auctions.

4. CONCLUSION

Web personalization is an advanced domain which deals with tailoring the content according to user's requirement. In this survey a brief description of various approaches of reaching to personalization is presented which can always be enhanced and improved. The survey aims to provide the knowledge about the personalization how it is achieved and what are all the functionalities performed by it. The personalization goal is to improve the information retrieval and provide the user friendly environment according to the needs of particular user. It considers every customer as an individual entity and provides the content to him specific to his needs. Web personalization has gained a great momentum nowadays in both research field as well as industries. A lot of advancement has happened but still many topics are open for more enhanced versions such as modelling of user context, incorporations of technologies like user profiling, web content acquisition and management has not yet been proposed. Also how semantic can enhance the users experience needs to be explored further.

REFERENCES

- [1] V. Krishnaraju, S. Mathew, V. Sugumaraman, “Web Personalization for User Acceptance of Technology: An Empirical Investigation of E-government Services”, *Information Systems Frontiers* pp. 1-17, 2016.
- [2] M.Amin and R. Nayak, “Theoretical Model of User Acceptance: In the View of Measuring Success in Web Personalization,” *Human-Computer Interaction*, Springer Berlin Heidelberg, pp. 255-264, 2010.
- [3] Y. Wang and A. Kobsa, “Respecting Users’ Individual Privacy Constraints in Web Personalization,” 11th International Conference, UM 2007, Corfu, Greece, June 25-29, pp. 157-166, 2007
- [4] M. Dudev, S. Elbassuoni and J. Luxenburgur, “Personalizing the Search for Knowledge”, *PersDb*, pp 1-8, 2008.
- [5] H. Zhang, Y. Song and H. Song, “Construction of Ontology-based User Model for Web Personalization”, In *International Conference on User Modeling* . Springer Berlin Heidelberg, pp. 67-76, 2007.
- [6] M. Eirinaki, J. Vlachakis and S. Anand “Ikum: An Integrated Web Personalization Platform based on Content Structures and User Behavior.” *Intelligent Techniques for Web Personalization*. Springer Berlin Heidelberg, pp.272-288, 2005.
- [7] D. Pierrakos, G. Paliouras and C. Papatheodoru “Web Usage Mining as a Tool for Personalization: A survey. *User Modeling and User-adapted Interaction* 13(4), pp 311-372, 2003.
- [8] M. Khribi, M. Jemni and O. Nasroui “Automatic Recommendations for E-learning Personalization based on Web Usage Mining Techniques and Information Retrieval. In *Eighth IEEE International Conference on Advanced Learning Technologies* pp. 241-245, 2008.
- [9] B. Mobasher , H Dai, T Luo and M. Nakagawa “Effective Personalization Based on Association Rule Discovery from Web Usage Data” In *Proceedings of the 3rd International Workshop on Web Information and Data Management ACM*, pp. 9-15, 2001.
- [10] L Hu, J Cao, L Cao, Z Gu and C Zhu “Personalized Recommendation via Cross-domain Triadic Factorization.” In *Proceedings of the 22nd international conference on World Wide Web ACM*, pp. 595-606., 2013.
- [11] M. Eirinaki and M. Vazirgianis “Web Mining for Web Personalization.” *ACM Transactions on Internet Technology (TOIT)* 3, no. 1 pp. 1-27, 2003.
- [12] B. Mobasher, “Data Mining for Web Personalization.” In *The adaptive web*, Springer Berlin Heidelberg, pp. 90-135, 2007
- [13] B. Mobasher, H Dai, T Luo, Y Sun and J Zhu “Integrating Web Usage and Content Mining for More Effective Personalization.” In *International Conference on Electronic Commerce and Web Technologies*. Springer Berlin Heidelberg, , pp. 165-176, 2000
- [14] J Xu, K. Waharmann and S Li “The development, Status and Trends of Recommender Systems: A Comprehensive and Critical Literature Review.” In *Proceedings of International Conference Mathematics and Computers in Science and Industry*, pp. 117-122, 2014.
- [15] G. Lekakos and GM Giaglis. “A Hybrid Approach for Improving Predictive Accuracy of Collaborative Filtering Algorithms.” *User Modeling and User-Adapted Interaction* 17, no. 1-2 pp. 5-40, 2007.
- [16] J. Srivastava, R. Cooley and M. Deshpande “Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data.” *Acm Sigkdd Explorations Newsletter* 1, no. 2, pp 12-23, 2000.
- [17] K. Chaudhary and S. Gupta ”Prioritizing Web Links based on Web Usage and Content Data.” In *Issues and Challenges in Intelligent Computing Techniques (ICICT)*, International Conference on IEEE, pp. 546-551, 2014
- [18] K. Chaudhary and S. Gupta “Web Usage Mining Tools & Techniques: A Survey.” *International Journal of Scientific & Engineering Research* 4, no. 6 , 2013.
- [19] J Lu, D Wu, M Mao, W Wang and G Zhang. “Recommender System Application Developments: a Survey.” *Decision Support Systems Elsevier* 74 pp. 12-32 , 2015.
- [20] AL Montgomery, and MD Smith, “Prospects for Personalization on the Internet.” *Journal of Interactive Marketing* 23 Elsevier, no. 2 pp. 130-137 , 2009.
- [21] C.J. Carmona, S. Ramírez-Gallego, F. Torres , E. Bernal , M.J. del Jesus, S. García “Web Usage Mining to improve the design of an E-Commerce Website: OrOliveSur. com.” *Expert Systems with Applications* 39 Elsevier, pp . 12 11243-11249, 2012

- [22] M Salehi, IN Kamalabadi and MBG Ghouschi "Personalized Recommendation of Learning Material using sequential pattern mining and attribute based Collaborative Filtering." *Education and Information Technologies* 19 Springer , no. 4 pp. 713-735, 2014.
- [23] W Chen, Z Niu, X Zhao and Y Li "A hybrid Recommendation Algorithm adapted in e-learning environments." *World Wide Web* 17, no Springer. pp 2 271-284, 2014.