ISSN:2320-0790

An International Journal of Advanced Computer Technology

A Novel Approach for Web Personalization through Web Mining Techniques

[1]Doddegowda B J, [2]Dr. G T Raju

Asso. Prof, Prof. & Head
Dept. of CSE, AMC Engineering College, Bangalore

ABSTRACT: Data on World Wide Web has been growing in an exponential manner. This raises a severe concern over information over load challenges for the users. Retrieving the most relevant information from the web as per the user requirement has become hard because of the large collection of heterogeneous documents. It is time consuming for the users to go through the long list of odds and ends to choose their relevant one. One approach to overcome this is to personalize the information available on the Web according to user requirements. The information or services provided by a Web to the requirements of individual or cluster of users, by considering their navigational patterns is termed as Web Personalization. The objective of Web Personalization is to provide users with what they really want or need, without having to ask or search for it explicitly. This approach effectively improves the performance of Information Retrieval (IR) systems. This paper presents an extensive survey on the various approaches proposed by the researchers in Web Personalization and challenges with a focus on future work.

Keywords: Web Personalization, User Profile, Personalized Search, Ontology, Information Retrieval, Semantic Web

1. Introduction

The content on the Web in various domains is rapidly increasing and the need for identifying and retrieving the content exactly based on the needs of the users is more than required. Therefore, an ultimate need nowadays is that of *predicting the user needs* in order to improve the usability of a Web site. In brief, Web Personalization can be defined as any action that adapts the information or services provided by a web site to an individual user, or a set of users, based on knowledge acquired by their *navigational behavior*, recorded in the web site's logs. This information is often combined with the *content* and the *structure* of the web site as well as the *user's interests/preferences*. Using the above specified sources of information as input to pattern discovery techniques, the proposed system molds the provided content to the needs of each visitor of the website. The personalization process can result in the dynamic generation of suggestions, the creation of pages according to the needs of the user, highlighting of existing hyperlinks that are exactly required by the users. Most of the earlier research efforts in Web Personalization deal with Web Usage Mining [1].Pure usage-based personalization, however, presents certain shortcomings, such as when there is insufficient use of data available in order to extract patterns, or when the web site's content changes and new pages are added but are not yet included in the web logs. The users' visits usually aim at finding information concerning a particular subject, thus the underlying content semantics should be a dominant factor in the process of web personalization. There have been a number of research studies that integrate the web site's content in order to enhance the Web Personalization process [2]. Most of these efforts characterize web content by extracting features from the web pages. Usually these features are keywords subsequently used to retrieve similarly characterized content based on the requirements of the user. When Web Personalization approaches were embedded with Semantic Web, it yields more effective search response and user satisfaction.

1.1 Need for Web Personalization

Considering the amount of data and variety of users on the World Wide Web, key word based search results

may not serve the purpose of providing the relevant information to the user, as each users' intention is different and the same may not reflect in the key words they use. Because of the above reasons web personalization has attracted many researchers to look into and provide a mechanism to understand the user in a better way and provide most relevant information to the user. User may not have time to fill in the data (method Explicit)describing about his/her interests, likes, dislikes, background educational qualification etc. Many web mining researchers worked on the above challenge and provided a few techniques for automatic personalization, the best example till date was Amazon where user need not give his/her details, the system will fetch the relevant information to the users.

Today, internet has become a part and parcel of our lives and one cannot imagine a world without internet, every day millions of people use internet for various purposes mostly for information. And user is often not happy due the amount of information he has been provided with, as the user needs further filtering, which is very time consuming and expects the system to understand his/her thoughts. Understanding user is not as simple as it's said, and web personalization is one step towards the goal.

2. Related Work

Generally, personalization methodologies aredivided into two complementary processes which are:

- The user information collection, used to describe the user interests and
- The inference of the gathered data to predict the closest content to the user expectation.

In the first case, user profiles can be used to enrich queries and to sort results at the user interface level [12]. Or, in other techniques, they are used to infer relationships like the social-based filtering [13] and the collaborative filtering [14]. For the second process, extraction of information on users' navigations from system log files can be used [15].Some information retrieval techniques are based on user contextual information extraction [16]. Information semantics are also used to enrich the personalization process, queries can be enriched by adding new properties from the available domain ontologies. The user modeling based on ontology can be coupled with dynamic update of user profile using results of information-filtering and Web usage mining techniques. Statistics collected through search engines show that spatial information is pervasive on the Web and that many queries contain spatial specifications, but it is more difficult to find relevant resources responding to query including a spatial component [17]. The spatial information personalization should consider spatial properties and relationships found in Web documents. Design of spatial Web applications requires at least three components: (1) a user model and associated user preference elicitation mechanisms and (2) a personalization engine combining spatial and semantic criteria and (3) a user interface enriched with spatial components [18]. The spatial Web personalization requires the representation of user features, particularly those relevant to the spatial domain. Semantic similarity and spatial proximity measures as well as relevance ranking functions on the behalf of the user is represented in [19].

Semantic similarity is the evaluation of semantic links existing between two concepts. [20] introduced a classification algorithm for measuring spatial proximity between two regions. Another aspect of spatial Web personalization techniques concerns interactive adaptive map generation and visualization. These techniques are concerned with Web maps adaptation according to user's needs [21].

The presented personalization approaches have contributed to the improvement of information systems use. However and despite their widespread use, these approaches have weaknesses and limitations. In fact, several approaches, like the collaborative ones, present the same recommendations for all users within the same cluster. Thus, they do not consider some specific users preferences when they represent a minority in a given group. Content based approaches facilitate items retrieval by proposing some alternatives and recommending similar items to the one that the user is visiting. However it focuses only on the user's actual and temporary needs and can't highlight the items that are related to the current query results. Other approaches try to determinate the interests of each user but they are limited by their items model that doesn't describe the differences between items properties. This lack of semantic description of the items decreases the quality of personalization since similarities and dissimilarities between items can't be measured accurately. In addition, in most personalization approaches, the spatial aspect is not taken into consideration, which requires an adaptation of those approaches to be relevant while applied to spatial information. The hybridization of existing approaches is presented as an alternative that would improve the quality of personalized systems [22].Dai and Mobasher [23] proposed a web personalization framework that characterizes the usage profiles of a collaborative filtering system using ontologies. The seprofiles are transformed to "domain-level" aggregate profiles by representing each page with a set of related ontology objects. In this work, the mapping of content features to ontology terms is assumed to be performed either manually, or using supervised learning methods. The defined ontology includes classes and their instances therefore the

aggregation is performed by grouping together different instances that belong to the same class. The recommendations generated by the proposed collaborative system are in turn derived by binary matching the current user visit expressed as ontology instances to the derived domain-level aggregate profiles, and no semantic relations beyond hyperonymy/hyponymy are employed. The idea of semantically enhancing the web logs using ontology concepts is independently described by Oberleet.al. [24]. This framework is based on a semantic web site built on an underlying ontology. This site contains both static and dynamic pages being generated out of the ontology. The authors present a general framework where data mining can then be performed on these semantic weblogs to extract knowledge about groups of users, users preferences, and rules. Since the proposed framework is built on a semantic web knowledge portal, the web contentis inherently semantically annotated exploiting the portal's inherent RDF annotations. The authors discuss how this framework can be extended using generalizations/specializations of the ontology terms, as well as for supporting the web personalization process, yet they mainly focus on web mining.

Acharyya and Ghosh [25] also propose a general personalization framework based on the conceptual modeling of the users' navigational behavior. The proposed methodology involves mapping each visited page to a topic or concept, imposing a tree hierarchy (taxonomy) on these topics, and then estimating the parameters of a semi- Markov process defined on this tree based on the observed user paths. In this Markov models based work, the semantic characterization of the context is performed manually. Moreover, no semantic similarity measure is exploited for enhancing the prediction process, except for generalizations/specializations of the ontology terms. Middleton et.al. [26] explore the use of ontologies in the user profiling process within collaborative filtering systems. This work focuses on recommending academic research papers to academic staff of a University. The authors represent the acquired user profiles using terms of a research paper ontology (is-a hierarchy). Research papers are also classified using ontological classes. In this hybrid recommender system which is based on collaborative and content-based recommendation techniques, the content is characterized with ontology terms, using document classifiers (therefore a manual labeling of the training set is needed) and the ontology is again used for making generalizations/specializations of the user profiles.Kearney and Anand [27] use an ontology to calculate the impact of different ontology concepts on the users' navigational behavior (selection of items). In this work, they suggest that these impact values can be used to more accurately determine distance between different users aswell as between user preferences and other items on the web site. This work focuses on the way these ontological profiles are created, rather than evaluating their impact in the recommendation process, which remains opens for future work.

3. Proposed Architecture for Web Personalization

The proposed architecture for Web Personalization in order to address some of the open issues is shown in Figure 1. The architecture uses *Web site's Content*, *Web logs* created by observing the user's navigational behavior and *User Profiles* created according to the user's preferences to analyze and extract the information needed for the user to find the pattern expected by the user. This analysis creates recommendations that are presented to the user. **Web usage mining** can be defined as automatic discovery of user navigational patterns. The goal of web usage mining has been to support decision making process of website owners to understand the user in a better way. However, these techniques can be used for personalization functions. Classifying the web content into semantic categories is done for predicting the pages for a user or group of users. Building User files will be done by gathering information specific to each user based on their interests/behavior and other demographic information. Web Personalization can be done to a group of specific interested users, based on the knowledge/patterns obtained from Web usage mining, Web content classification, and user profiles. Web Personalization can also include techniques such as use of cookies and machine learning strategies. Web Personalization may be viewed as a type of Recommender system, Clustering, Classification, or even Prediction of pages for a web user or group of Users. With personalization, the content of the web pages are modified to better fit for user needs. This may involve creating web pages, that are unique per user or using the desires of a user to determine what web documents to retrieve.

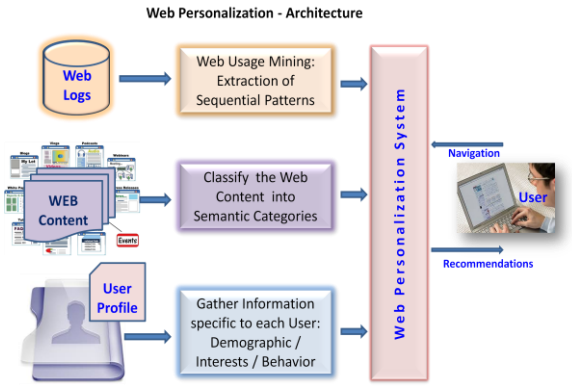

Figure 1: Proposed Architecture for Web Personalization through Web Mining Techniques

4. Data Sources

The proposed Web personalization system utilizes Web data in order to personalize a Web site. Web data are classified into four categories [Srivastava et al., 2000].

- ***Content data*** are presented to the end-user appropriately structured. They can be simple text, images, or structured data, such as information retrieved from databases.
- ***Structure data*** represent the way content is organized. They can be either data entities used within a Web page, such as HTML or XML tags, or data entities used to put a Web site together, such as hyperlinks connecting one page to another.
- ***Usage data*** represent a Web site's usage, such as a visitor's IP address, time and date of access, complete path (files or directories) accessed, referrers' address, and other attributes that can be included in a Web access log.
- ***User profile data*** provide information about the users of a Web site. A user profile contains demographic information such as name, age, country, marital status, education, interests etc. for each user of a Web site, as well as information about users' interests and preferences. Such information is acquired through registration forms or questionnaires, or can be inferred by analyzing Web usage logs.

5. Web Personalization and User Profile

A user profile is a collection of personal data associated to a specific user. A profile refers to the explicit digital representation of a person's characteristics. User profiles can also be described as the computer representation of a user model. The user profiles are created for user background knowledge description[4][5][6]. User profiles represent the concept models possessed by users when gathering web information. A concept model is implicitly possessed by users and is generated from their background knowledge. This knowledge is used to gather relevant information about a user's preference and choices. User profiles are categorized into three groups: *Interviewing, semi-interviewing, and non-interviewing.*

Interviewing user profiles are considered to be perfect user profiles. They are acquired by using manual techniques such as questionnaires, interviewing users, etc. In these methods each is recommended to read each document and give a positive or negative judgment to the document against a given topic. *Semi-interviewing* user profiles are acquired by semi automated techniques with limited user involvement. For example, these techniques usually provide users with a list of categories and ask users for interesting or non interesting categories. *Non interviewing* techniques do not involve users at all, but discover user interests instead. They acquire user profiles by observing user activity and behavior and discovering user background knowledge.

The *interviewing, semi-interviewing, and non interviewing* user profiles can also be viewed as *manual*, *semiautomatic*, and *automatic* profiles, respectively.

There are many models that have been developed for representing user profiles. These models provide knowledge from either a global or local knowledge base. The *global analysis* uses existing global knowledge base and to produce effective performance. The commonly used knowledge base include generic ontology such as Word net, Thesauruses, Digital Libraries. The *local analysis* observes user behavior in user profiles. The user background knowledge can be better discovered and represented if global and local analysis is integrated. Local analysis is used for analyzing the user behavior in user profiles. It can be better improved by using ontological user profiles.

5.1 Techniques using User Profiles

The most common way to use a profile is to store information that enables personalization on an individual basis as represented in Figure 2A. This is called Content based Filtering which, applied to a textual document, evaluates the document's relevance by matching the keywords contained in a user profile with the keywords extracted from the text [7]. On the Web, to prevent the user profiles transmitting through the network, user profiles are stored at the server. Social or collaborative filtering [8] is another effective way to take advantage of user profiles. This method collects the user profiles of a group of people and generates recommendations based on the similarities of the profiles as given in the Figure 2B. To implement collaborative filtering, the profiles of all users must be compared and therefore the best storage location is also to centralize them at the server. A user profile can also be shared between different personalized applications that require the same user profile's content as in Figure 2C. This collaboration enables both applications to gain a much more knowledge about the user's interest. Because all the personalized Web applications (on different servers) need to have access to the complete set of profiles for a specific user, it is required to store user profile at the browser.

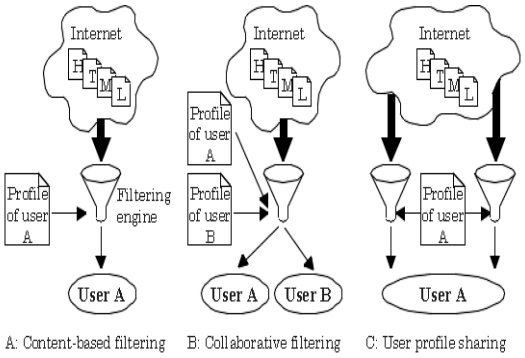

Figure 2: Different Uses of Profiles

6. Semantic based Personalized Search

Personalization aims to find a subset of Web data thatmatches the interested profile of a user or a group of users.This can be achieved by recommending Web pages orWebsites to the users, or by filtering Web pages that are ofinterest to the users [9]. For example, this can done byanalyzing the historical data recording user accesses toWeb data, and mining the topics relevant to a user byclustering previously accessed Web pages based oncontent similarities. When a new Web page is found to besimilar to one of the clusters, it can be routed to the user.Personalized search takes advantage of Semantic Webstandards (RDF and OWL) to represent the content and theuser profiles. Semantic based Personalization of Web dataaccess can be effectively used for improving the precisionand recall in search, particularly by re-ranking the searchresults based on the learner's past activities. The core partof Semantic approach on Web Personalization is the use ofOntology. As Web pages are annotated with ontologyentity labels, the Web pages accessed by a user can lead tomore effective content recommendation. More detailed survey on semantic web personalization has been presented in [32].

7. Conclusion

Although the World Wide Web is the largest source ofelectronic information, it lacks with effective methods forretrieving, filtering, and displaying the information that isexactly needed by each user. With the advent of theInternet, there is a dramatic growth of data available on the World Wide Web. Hence the task of retrieving the onlyrequired information keeps becoming more and moredifficult and time consuming. To reduce informationoverload and create customer loyalty, WebPersonalization, a significant tool that provides the userswith important competitive advantages is required. IT Personalized Information Retrieval approach that is mainlybased on the end user modeling increases user satisfaction.Also personalizing web search results has been proved asto greatly improve the search experience. This paper reviews the various research activities carried out to improve the performance of personalization process and also the Information Retrieval system performance. The purpose of this survey is to describe the state-of-the-art of personalized recommender systems and various techniques employed for the same. We have presented a comprehensive description about various web personalization systems in the recent years. Though a lot of research has been done in this field, yet a system that effectively integrates various diverse requirements of the users has not yet been proposed. Future work in web personalization includes the in depth study on fusion of User profiles, web content data and web mining techniques for effective web personalization. This survey also identified a few areas to be explored like learning techniques including the vector space model, Genetic algorithms, and the probabilistic model or clustering in the field of web personalization. Integrating the systems like social networking and blogs and other popular websites is also a potential area to explore. Owing to the spread of mobile devices in the current era, web personalization needs to be explored on the mobile arena as well.

8. REFERENCES

[1]. M. Albanese, A. Picariello, C. Sansone, L. Sansone, "*A Web Personalization System based on Web Usage Mining Techniques"*, in Proc. of WWW2004, May 2004, New York, USA.

[2]. 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 (ECWeb2000), Greenwich, UK, September 2000.

[3]. Jiawei Han AndMichelineKamber "*Data Mining: Concepts and Techniques*", 2nd ed., Morgan Kaufmann Publishers, March 2006. ISBN 1-55860-901-6.

[4]. S. Gauch, J. Chaffee, and A. Pretschner, "*Ontology-Based Personalized Search and Browsing*" Web Intelligence and Agent Systems, vol. 1, nos. 3/4, pp. 219-234, 2003.

[5]. Y. Li and N. Zhong, "*Web Mining Model and Its Applications for Information Gathering*" Knowledge-Based Systems, vol. 17, pp. 207-217, 2004.

[6]. Y. Li and N. Zhong, "*Mining Ontology for Automatically Acquiring Web User Information Needs*" IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 554-568, Apr. 2006.

[7]. Morita M., Shinoda, Y., *"Information Filtering Based on User Behaviour Analysis and Best Match Retrieval"*, in Proceedings of the 17th International ACM-SIGIR Conference on

Research and Development in Information Retrieval, 1994, pp. 272-281.

[8]. Shardanand U., Pattie M., "*Social Information Filtering: Algorithms for Automating "Word of mouth"*, in Proceedings of the Human Factors in Computing System, Denver, May 1995, pp. 210-217.

[9]. Xiaohui Tao, Yuefeng Li, and NingZhong, "Senior Member, IEEE, "*A Personalized Ontology Model for Web Information Gathering*", IEEE Transactions On Knowledge and Data Engineering, VOL. 23, NO. 4, APRIL 2011.

[10]. BhaganagareRavishankar, DharmadhikariDipa, *"Web Personalization Using Ontology: A Survey"*, IOSR Journal of Computer Engineering (IOSRJCE) ISSN : 2278-0661 Volume 1, Issue 3 (May-June 2012), PP 37-45 *www.iosrjournals.org.*

[11]. Xiaohui Tao, Yuefeng Li, and NingZhong, Senior Member, IEEE. "*A Personalized Ontology Model for Web Information Gathering*", IEEEtransactions on Knowledge and Data Engineering, Vol. 23, No. 4, April 2011.

[12]. Koutrika, G., Ioannidis, Y.: "*A Unified User Profile Framework for Query Disambiguation and Personalization*", in "Workshop on New Technologies for Personalized Information Access", held in conjunction with the 10th International Conference on User Modeling, pp. 44–53 (2005).

[13]. Mladenic D., "*Text-learning and Related Intelligent Agents: a Survey*", IEEE Intelligent Systems, 14(4):44–54 (1999).

[14]. Goldberg, K., Roeder, T., Gupta, D., Perkins, C.: Eigentaste, "*IT Constant Time Collaborative Filtering Algorithm. Information Retrieval*"**,** Journal, 4(2):133–151, (2001).

[15]. 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 Databases", LNCS, vol. 3202, pp. 552-554, Springer (2004).

[16]. Jones, G.J.F., Brown, P.J., "*Context-Aware Retrieval for UbiquitousComputing Environments"*, Invited paper in Mobile and Ubiquitous Information Access, LNCS, vol. 2954, pp. 227–243. Springer (2004).

[17]. Yang, Y., Aufaure, M.A., Claramunt, C., "*Towards a DL-Based Semantic User Model for Web Personalization"*, in "Third International Conference on Autonomic and Autonomous Systems", pp. 61-61. IEEE Computer Society (2007).

[18]. Kuhn, W., Handling Data Spatially, "*Spatializing User Interfaces*", in "7th International Symposium on Spatial Data Handling", Advances in GIS Research II, vol. 2, pp.13B.1-13B.23, IGU (1996).

[19]. Resnik, P., "*Semantic Similarity in a Taxonomy: an Information-Based Measure and its Application to Problems of Ambiguity in Natural Language*", Journal of Artificial Intelligence Research, 11:95–130 (1999).

[20]. Larson, R.R., Frontiera, P., "*Spatial Ranking Methods for Geographic Information Retrieval in Digital Libraries*", in Heery, R., Lyon, L., (eds.) ECDL. LNCS, vol. 3232, pp. 45–56. Springer (2004).

[21]. Maceachren, A.M., Kraak, M.J., "*Research Challenges in Geovisualization, Cartography and Geographic*", Information Science, 28(1):3–12 (2001).

[22]. Burke, R., "*Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction*", 12(4):331– 370 (2002).

[23]. H. Dai, B. Mobasher, "*Using Ontologies to Discover Domain-Level Web Usage Profiles"*, in Proc. of the 2nd Workshop on Semantic Web Mining, Helsinki, Finland, 2002.

[24]. D.Oberle, B.Berendt, A.Hotho, J.Gonzalez, "*Conceptual User Tracking"*, in Proc. of the 1st Atlantic Web Intelligence Conf. (AWIC), 2003.

[25]. S. Acharyya, J. Ghosh, "*Context-Sensitive Modeling of Web Surfing Behaviour Using Concept Trees"*, in Proc. of the 5th WEBKDD Workshop, Washington, August 2003.

[26]. S.E. Middleton, N.R. Shadbolt, D.C. De Roure, "*Ontological User Profiling in Recommender Systems"*, ACM Transactions on Information Systems (TOIS), Jan. 2004/ Vol.22, No. 1, 54-88.

[27]. P. Kearney, S. S. Anand, "*Employing a Domain Ontology to gain insights into user behaviour"*, in Proc. of the 3rd Workshop on Intelligent Techniques for Web Personalization (ITWP 2005), Endiburgh, Scotland, August 2005.

[28]. ChhaviRana, "Trends in Web Mining for Personalization", IJCST Vol. 3, Issue 1, Jan. - March 2012 ,

[29]. M. Eirinaki, M. Vazirgiannis, "Web Mining for Web Personalization", ACM Transactions on Internet Technology, February 2003

[30]. A Survey On Ontology Based Web Personalization, KiranJammalamadaka, I V SrinivasIJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308, Volume: 02 Issue: 10 | Oct-2013,

[33]. ISSN 2224-896X (Online) Vol 2, No.6, 2012

[31]. Web Personalization Approaches: A Survey, K. Sridevi1 And Dr. R. Umarani, International Journal of Advanced Research in Computer and Communication EngineeringVol. 2, Issue 3, March 2013

[32]. Semantic Web Personalization: A Survey, Ayesha Ameen 1* Khaleel Ur Rahman Khan 2 B.Padmaja Rani, Information and Knowledge Management ISSN 2224-5758 (Paper)