

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.017

*IJCSMC, Vol. 5, Issue. 10, October 2016, pg.104 – 113*

# Acquaintance of Semantic Web Personalization

Rinku Singh  
A.P.S.University, Rewa(M.P)  
486003, India

[Rinkusingh438@yahoo.com](mailto:Rinkusingh438@yahoo.com)

Dr. P.K. Rai, Head, Computer center  
A.P.S.University, Rewa(M.P)  
486003, India

[pkrapso@gmail.com](mailto:pkrapso@gmail.com)

Mukesh Shukla  
A.P.S.University, Rewa(M.P)  
486003, India

[needmukesh@gmail.com](mailto:needmukesh@gmail.com)

*Abstract: An improving without a pause, growth of information in World Wide Web (W3) is very difficult for the users to access the desired web pages from the website because day by day online information are growing or improving on the web. So, without taken any help of system the user may not spend more time to get the interested information from the website. There are much information on the web to search for the user's but problem is that which type of information they want and how they can get that information without wasting the time. In this paper, we have analyzed SWP which has emerged as one of the widely used techniques to search desired contents on the web in more effective way because of its ease in use and better performance. In recent year, we have seen lot of research focus on SWP and Data Mining.*

*Keywords: Semantic Web, Mining, Web Personalization.*

## 1. Introduction

Nowadays, the tremendous growth of the information available online. W3 has rapidly become the main source of information to the greater number of new users in many domains. This increases the number of inexperienced users. So, these factors raise the needfulness of analyzing and understanding the behavior and the interests of web users in order to personalize reaction to their requests. This may erect a friendly relationship between the web

and its users. W3 have brought us into a world of endless possibilities. It serves as a platform for exchanging various kinds of information such as research paper, educational content, multimedia contents, software and web logs etc. But how can users select a huge universe of information of widely varying quality? It is often necessary to make choices without sufficient personal experiences of the alternatives. Therefore, the ultimate need is that of understanding the behavior and the interest of web users in order to improve the usability personalize responses to their request. This may create a friendly relationship between the web and its users.

## 2. Preliminary

Currently, many research projects are working with Web Usage Mining(WUM) and Web personalization areas. Most of the efforts target on extracting useful patterns and rules using data mining techniques in order to understand the user's navigational behavior. In several cases, a recommendation engine can help the user to navigate through a site. All research efforts associate more than one of the previous methods in web personalization, namely, user profiling, Web usage mining techniques, web structure mining, content management and publishing mechanisms. Here, we have explored the most valuable research efforts in the Web mining and web personalization domain.

One of the earliest attempts to take advantage of the information that can be gained through exploring a visitor's navigation through a Web site resulted by [Lieberman 1995], a client-site agent that monitors the user's browsing behavior and searches for dormant interesting pages for recommendations. The agents looks ahead at the neighboring pages using a Best First Search(BFS) enhance by heuristics inferring user interest, as much as they are derived from the user's navigational behavior, and offers suggestions. A model that consist of two models proposed by [Yan et al.1996] such as offline module and online module. Offline module can perform cluster analysis on the web logs and an online module aiming at dynamic link generation.

Webwatcher is the most popular systems from the early days of WUM introduced by [Joachims et al.1997]. This can provide an idea to create a tour guide agent that provides navigation hints to us through collection of web, based on its knowledge of the user's interests, the location and relevance of various items in the location, as well as the way by profiling the user, acquiring information about user's interests. the user requests a page every day, this information is drive off through a proxy server in order to easily track the user session across the Web site and links trusted to be of interest for the user are highlighted. Its strategy for giving advice is learned from feedback from prior tours. A similar system is the Personal WebWatcher also introduced by [Mladenec 1999], which is structured to specialized for a particular user, modeling interests. It solely records the addresses of pages which has requested by the user and always highlights more interesting hyperlinks without involving the user in its learning techniques.

[Chen et al.1996] introduced the maximum forward reference concept in order to characterize user session for the mining of traversal methods. Their work is based on very effective paths and association rules discovery and a forwarded by reference is defined as the sequence of pages which has been requested by a user up to the last page before backtracking. The speed tracer project [Wu et al. 1998] is proposed by Chen et al. [1996]. Speed tracer uses the referrer page and the URL of the requested page like a traversal way and reconstructs the user traversal paths for session identification.

[Zaiane et al.1998] adopt a different approach to combine the OLAP and data mining techniques and a multidimensional data cube to extract interactively implicit knowledge. The WebLogMiner system after filtering the data contained in the Weblog transforms into a relational database. The use of cube is proposed by [Huang et al.2001] that can explicitly search/identifies Web access sessions maintains the sequence of the session's components and uses multiple attributes to explore the visited Web pages . A model represented by [Borges and Levene 1999] is a set of user navigation sessions as a hypertext whose higher probability generated strings correspond to the user's preferred trails. The use of a client-side agent is proposed by [Shahabi et al.1997] that captures the client's behavior to create a profile.

Joshi et al.[2000] introduced the notion of uncertainty in Web usage mining searching a clusters of user session profiles using robust fuzzy algorithms. In this a user or a page can be assigned to more than one cluster. [Cooley et al.1999]&[Srivastava et al.2000] defined Web Usage Mining(WUM) as a three-phase process, consisting of preprocessing, pattern discovering, and pattern analysis. Their prototype system WebSift performs intelligent cleansing and preprocessing for recognizing users, server sessions, and collect cached page references through the use of the referrer field and also performs content and structure preprocessing are introduced by[Cooley et al. 1999]. Data mining techniques are applied by [Masseglia et al.1999] such as association rules and sequential pattern discovery on Web log files and then use to arrange the server hypertext organization dynamically. They regard two phase of web usage mining such as: preprocessing and web mining. The problem of incremental web usage mining is addressed recently by [masseglia et al.2000]. Using the WUM method they handle the problem of mining user patters when new transactions are added to the weblog file by considering only user patters obtained by an earlier mining.

[Bucher and Mulvenna1998] presented a Knowledge Discovery Process(KDP) in order to search marketing intelligence from Web data. They propose an environment that explores existing online analytical mining or Web usage mining access and incorporates marketing expertise. In a more recent work Buchner et al.[1999] introduced the data mining algorithm MiDAS for discovering sequential patterns from Web log files to perceive behavioral marketing intelligence. [Spilioulou et al.1999,2000] have designed MINT which is an another mining language for the operation of WUM is known as a sequence mining system for the specification, discovering, and visualization of interesting navigation pattern.

[Berendt 2000, 2001] has implemented STRATDYN, add on module that extends WUM capabilities by identifying the differences between navigation methods and exploiting the usage of site's semantics in the visualization of the result. [Coenen et al. 2000] introduced a

framework for self adaptive Web sites catching into account the site structure except for the site. This is based on the fact that the method used in Web usage mining produce recommendations including links that don't exist in the original site structure. [Perkowitz and Etzioni 1998,1999,2000] were the first who have defined the awareness of flexible Web sites as a sites that can semi-automatically promote their organization and presentation by learning from visitor access patterns. They proposed a system semi-automatically modifies a web site which is granting only non-destructive transformations. For this reason, nothing is removed or altered instead only index pages containing related links but currently unlinked pages are add to web site. In a more recent work, they have moved from the statistical cluster mining also page gather to index finder which focus statistical & logical information to synthesize index pages.

[Cingil et al.2000] described an architecture that provides a vaster view of personalization through the use of various World Wide Web Contents(W3C). They characterize how standards such as XML, RDF & P3P can be used to create application of personalization. The most advanced system is the Web Personalizer proposed by [Mobasher et al.1999,2000]. This can provide a framework for mining web log files to find knowledge for the categories of recommendations to active users which is totally based on their browsing similarities to previous users. This framework has been recently extended to associate the content profiles into the recommendation procedure as a way to upgrade the effectiveness of web personalization actions.

### **3. Semantic Web**

The Semantic Web Provides the way that how scientific data are collected, deposited and analyzed and it is a meaning of the data from different kinds of web resources to allow the machine to interpret and understand these enriched data to precisely answer and satisfy the web user's requests. It represents the extension of the W3 that gives users the facility to share their data beyond all the hidden boundary and the limitation of programs and websites using the meaning of the web. It is used to crack two specific problems like: a) the limitations of data access in the web. b) The delegation tasks by supporting access to data at web scale and enabling the delegation of certain classes of tasks.

#### **3.1 Semantic Web Representation**

There are many techniques and models available to represent the semantic data such as the standard techniques recommended by W3C named XML,Resource Description Framework, and Web Ontology Language(WOL) which are briefly explained below.

- Extensible markup language  
This technique has been established as a generic technique to store, organize and retrieve data from the web by enabling users to create their own tags. It allows them to define their content easily.
- Resource description framework

This is a common language that enables the facility to store resources information that are available in the W3 using their own domain vocabularies. There are three types of elements contented in the RDF Resources, Literals, and Properties.

➤ Web ontology language

This is considered a more complex language with suitable machine interpret facility than RDF. It precisely identifies the resources nature and their relationships. The architecture of semantic web is based on the vision of Sir BernersLee is divided into seven layers -1) URI 2) XML,NS & XML schema 3) RDF schema 4) Ontology vocabulary 5) Logic 6) Proof 7) Trust.

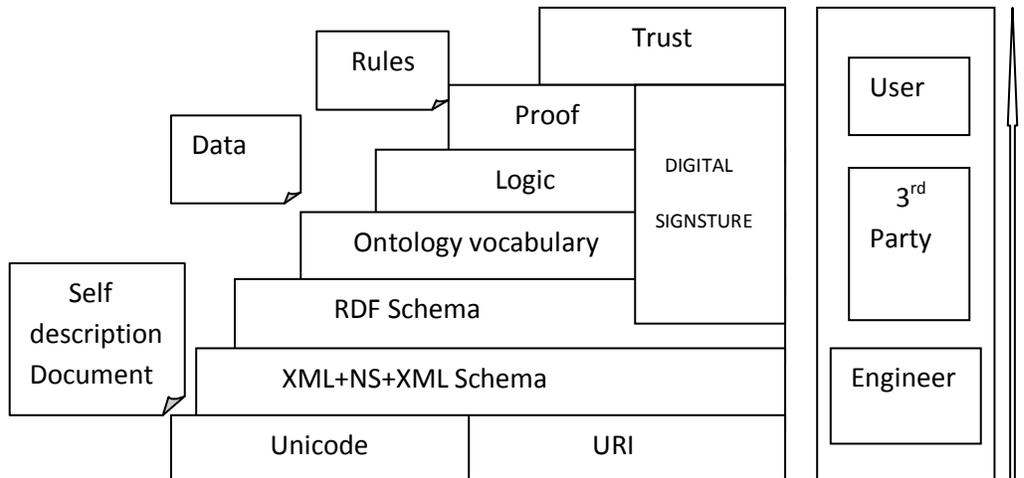


Figure. 1. The architecture of semantic web layers

### 3.2 Semantic web mining

Semantic Web Mining(SWM) has become one of the most important requirements for the web users. Web server stores various patterns of data such as: text, image, audio & video etc. But servers can't identify the contents of the data. The search techniques can be improved by semantic web mining and probabilistic analysis to get more accurate results. This technique can provide meaningful search of data resources by eliminating useless information with mining process. SWM is the outcome of two new and fast developing domains: such as, Semantic web and Data mining.

Semantic web is an enlargement of the active web in which information or knowledge is given well defined meaning, better enabling computers and people to work in co-operation[1]. Data

Mining is the nontrivial process of identifying valid, previously unknown, potentially useful patterns in data[2].

The aim of Semantic Web Mining is to discover and retrieve useful and interesting patterns from a large group of web data. This web data contain the various kind of information including web structure data, weblog data & user profiles data.

#### 4. Semantic Web Mining Taxonomy

SWM is divided into three categories according to the various sources of data analyzed as shown in Fig.2

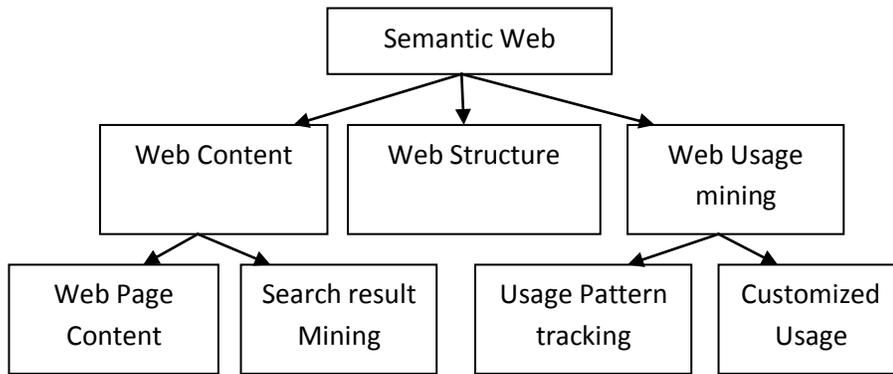


Figure 2. Semantic Web mining Taxonomy.

##### 4.1 Web Content Mining

WCM is more than selecting relevant documents on the web. It focuses on the discovery of knowledge from the content of web pages and it is related to extract the information and knowledge discovery from analyzing a collection of web documents. It is the effort for organizing the semi-structured web data into structured collection of resources are best to do more efficient querying mechanisms and more efficient information collection or extraction.

##### 4.2 Web Usage Mining

WUM focus on the discovery of knowledge from user navigation data when visiting a website. The target data are requested from users recorded in special files stored in the website's servers called log files. This is a finding pattern of navigational behavior from users visiting a website. Navigational behavior can be valuable when searching answers to questions such as

how efficient is our website in delivering information? How the users perceive the structure of the website? Etc. Answer to these questions may come from the analysis of the data from log files stored in web server.

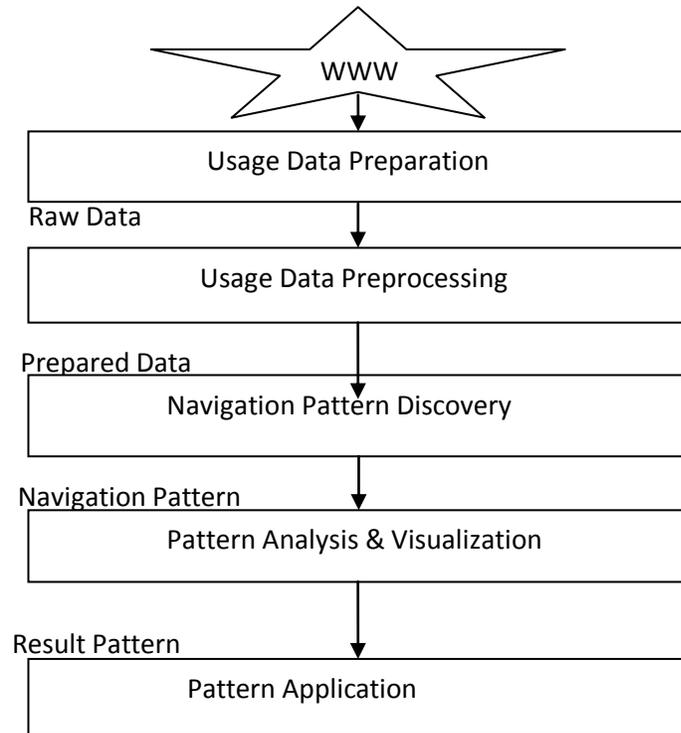


Figure. 3. Web Usage Mining Architecture.

### 4.3 Web Structure Mining

WSM closely related to analyzing hyperlinks and link structure on the web for information retrieval and knowledge discovery. It can be used by search engines to rank the relevancy between websites classifying the according to their similarity and relationship. Personalization and recommendation systems based on hyperlinks. WSM is used for identifying the power of web pages that are pointed to by a large set of other web pages that make them candidates of good sources of information. It is also used for discovering community network by extracting knowledge from similar links.

## 5. Web Personalization

WP is the process of customizing the content and structure of a web site to the specific requirement of each user taking advantage of user's navigation behavior. Recommendation systems support web site personalization by indicating user's behavior and recommending similar items by inviting users to rate objects and state their choice and interests so that, recommendations can give service to them based on other users rates with similar preferences. Collaborative Based Filtering(CBF) is the most common method for web personalization from server log files and has considerable attention from researchers for constructing user models that represent the behavior of users. Web personalization reorganizes web sites or pages based on the web experience to fit individual user needs. It is a large area that contains adaptive web sites and recommender systems as special cases. WP system uses a subset of web log and session clustering process to derive usage profiles, which is used to generate recommendations.

### 5.1 Computing of Web Personalization

- A. Personalization Mechanisms(PM) – PM is based on filter the data such as CBF. in this, the user model includes information about the contents of items of interest.
- B. Collaborative Based Filtering(CBF) – The system constructs rating profiles of its users in the social or collaborative filtering and locates other users with similar rating profiles and returns items that the similar users rated highly.
- C. Collection of Web Data(CWD) – There are two way to collect the data such as Implicit and Explicit. Implicit data includes past click streams as recorded in web server logs or via cookies/session tracking schedule. Explicit or specific data usually comes from registration forms and rating questionnaires.
- D. Preprocessing of Web Data(PWD) – Preprocessing may include cleaning data of inconsistencies, filtering out irrelevant information according to the goal of analysis.
- E. Analysis of Web Data(AWD) – This is known as Web usage mining and it can frequently results in automatic user profiling.
- F. Final Recommendation Phase(FRP) – The recommendation process involves generating dynamic web content on the server.

## 6. Conclusion

Web Personalization is becoming more and more important with the explosive growth of information source available on the W3. It has come a long way over the last decade. In this paper, we have looked at some fundamental aspects of semantic web and we have outlined the different taxonomy of Semantic Web Mining namely WCM,WSM,WUM. We have tried to summarize the significance of introducing the Web Mining Aspects in the area of web personalization also we have acquaintance some useful process of web personalization that are being used on the web. In the future, we will focus on further possibilities to get the online

contents semantically by web personalization which could help users to search the desired information very easily.

## REFERENCES

1. [1995]. Lieberman,H.Letizia,"An agent that assists web browsing", In proceedings of the Fourteenth international joint conference on artificial intelligence(Montreal).
2. [1996]. Chen,M.S,Park,J.S,and Yu,P.S,"Data mining for path traversal patterns in a web environment", In proceedings of the Sixteenth international conference on Distributed Computing Systems(May),385-392.
3. [1996]. Yan,T.W.,Jacobsen,M.,Garcia Mollina,H.,& Dayal,U,"From user access patterns to dynamic hypertext linking", In Fifth International World Wide Web Conference(WWW5)(Paris).
4. [1997]. Joachims, Freitag,D,&Mitchell,T,"Web Watcher: A tour guide for the world wide web", In proceeding of IJCA197(August).
5. [1997]. Shahabi,C,Zarkesh,A.M,Adibi,J, & Shah,V,"Knowledge discovery for users web page navigation", In Workshop on Research Issues in Data Engineering(Birmingham,UK).
6. [1998]. Wu,K.L,Yu,P.S,& Ballman,A,"Speed Tracer: A web usage mining and analysis tool",IBM Syst.J.37,1.
7. [1998]. Zaiane,O.R,Xin,M,& Han,J,"Discovery web access patterns and trends by applying OLAP and data mining technology on web logs", In proceedings of Advances in Digital Libraries Conference(ADL 98)(Santa Barbara,CA,April).
8. [1998]. Buchner,A & Mulvenna,M.D,"Discovering Internet marketing intelligence through online analytical Web Usage Mining",SIGMOD Rec.27,4,54-61.
9. [1998]. Perkowit,M,&Etzioni,O,"Adaptive web sites: Automatically synthesizing web pages", In proceedings of the Fifteenth National Conference on Artificial Intelligence(Madison,WI,July).
10. [1999]. Borges,J., & Levene,M.,"Data Mining of user navigation patterns", In Web Usage Analysis and User Profiling, Lecture Notes in Computer Science, vol.1836,Springer-verlag New York,92-111.
11. [1999].Buchner,A.G,Baumgarten,M.,&S.S,Mulvenna,M.D,&Hughes,J.G,"Navigation Pattern Discovery from Internet Data", In proceedings of the Web Usage Analysis and User Profiling Workshop(WEBKDD'99),Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(San Diego,August),25-30.
12. [1999]. Cooley,R., Mobasher,B & Srivastava,J,"Data preparation for mining world wide web browsing patters", Knowl.Info.Sys.,(Feb).
13. [1999]. Cooley,R.,Tan,P.N,& Srivastava,J,"WebSift: The web site information filter system", In proceedings of the Web Usage Analysis and User Profiling Workshop(KEBKDD'99),Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(Boston,August).
14. [1999]. Masegla,F.,Poncelet,P.,&Teisseire,M,"Using data mining techniques on web access logs to dynamically improve hypertext structure", In ACM SigWeb Lett.,8,3(Oct.)13-19.

15. [1999]. Mladenic,D,"Machine learning used by personal web watcher", In proceeding of ACAI-99 Workshop on Machine Learning and Intelligence Agents(Chania,Greece,July).
16. [1999]. Mobasher,B,Cooley,R & Srivastava,J,"Creating adaptive web sites through usage based clustering of URLs", In proceeding of the IEEE Knowledge and Data Engineering Exchange Workshop(KDEX'99)(Nov.).
17. [1999]. Perkowitz,M,&EtzionimO,"Adaptive web sites; Conceptual cluster mining", In proceedings of the Sixteenth International Joint Conference on Artificial Intelligence(IJCAI99)(Stockholm).
18. [1999]. Spiliopoulou,M.,Faulstich,I.,& Wilkler,K,"A data miner analyzing the navigational behavior of web user", In proceedings of the Workshop on Machine Learning in User Modelling of the ACA199(Chania,Greece,July).
19. [2000]. Berendt,B,"Web Usage mining, site semantics, and the support of navigation", In proceedings of the Workshop (WEBKDD) Web Mining for e-commerce challenges and opportunities, Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(Boston,August).
20. [2000]. Cingil,I.,Dogac,A.,& Azgin,"A broader approach to personalization", Commun.ACM,43,8(Aug.),136-141.
21. [2000]. Coenen,F.,Swinnen,G.,Vanhoof,K.,&Wets,G,"A framework for self adaptive websites: Tactical versus strategic changes", In proceedings of WEBKDD Web Mining for e-commerce challenges and opportunities",Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Boston,August).
22. [2000]. Joshi,A, Joshi, K & Krishnapuram,R,"On mining Web Access Logs", In proceedings of the ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery,63-69.
23. [2000]. Masseglia,F.,Poncelet,P., & Teisseire,M,"Web Usage Mining: How to efficiently manage new transactions and new customers", In proceedings of the Fourth European Conference on Principles of Data Mining and Knowledge Discovery(PKDD'00)(Lyon,France,Sept.).
24. [2000]. Mobasher,B,Cooley,R, & Srivastava,J,"Automatic personalization based on web usage mining", Commun.ACM,43,8(August),142-151.
25. [2000]. Perkowitz,M,&Etzioni,O,"Towards adaptive web sites: Conceptual framework and case study", In Artificial Intelligence,118,1-2,245-275.
26. [2000]. Spiliopoulou,M,"Web Usage Mining for web site evaluation",Commun.ACM 43,8(August),127-134.
27. [2000]. Srivastava,J, Cooley,R.,Deshpande,M & Tan, P.N,"Web Usage Mining : Discovery and applications of usage patterns from Web Data", SIGKDD Explorations I,2(Jan.),12-23.
28. [2001]. Berendt,B,"Understanding Web Usage at different levels of abstraction: Coarsening ad visualizing sequences". In proceedings of the Workshop (WEBKDD) Mining Log Data across all customer touchpoints, Seventh ACM SIGKDD International Conferences on Knowledge Discovery and Data Mining(San Francisco,August).
29. [2001]. Huang,Z,NG,J,Cheung,D.W,Ng,M.K,& Ching,W.K,"A cube model for web access sessions and cluster analysis", In proceedings of the Mining Logs Data Across All Customer TouchPoints Workshop(WEBKDD'01)Seventh ACM SIGKDD International Conference on Knowledge Discovery and data mining(San Francisco,August).