

WEB PERSONALIZATION APPROACHES: A SURVEY

K. SRIDEVI¹ and Dr. R. UMARANI²

Assistant Professor, Department of Computer Science, Nehru Memorial College, Puthanampatti, Trichy District, Tamilnadu, India¹

Associate Professor, Department of Computer Science, Sri Sarada College for Women, Salem, Tamilnadu, India²

ABSTRACT: Retrieving the most relevant information for the Web becomes difficult because of the huge amount of documents available in various formats. It is mandatory for the users to go through the long list of snippets and to choose their relevant one, which is a time consuming process. User satisfaction is secondary in this aspect. One approach to satisfy the requirements of the user is to personalize the information available on the Web, called Web Personalization. Web Personalization is the process that adapts information or services provided by a Web to the needs of each specific or set of users, taking the facts of the knowledge gained from the users. Web Personalization can be the solution to the information overload problem, as its objective is to provide users with what they really want or need, without having to ask or search for it explicitly. It is a multi discipline area for putting together data and producing personalized output for individual users or groups of users. This approach helps the researchers to improve the efficiency of Information Retrieval (IR) systems. By considering all the benefits of the Web Personalization, this paper presents elaborately the various approaches used by researchers to achieve Web Personalization in Web Mining.

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

I. INTRODUCTION

Web Personalization

The content on the Web in various fields 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 system molds the provided content to the needs of each visitor of the web site. 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.

Web Personalization Architecture

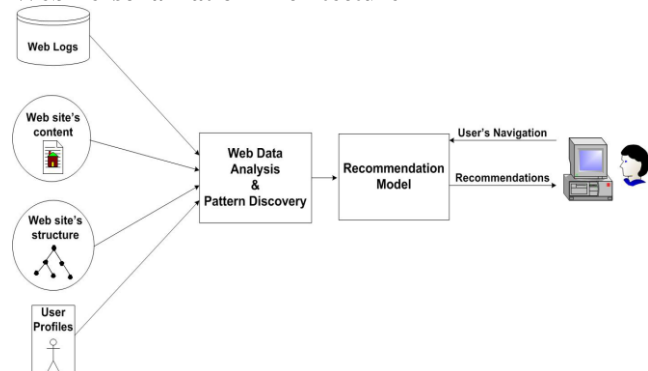


Figure 1: Web Personalization System Process



The above architecture uses Web site's structure, Web logs created by observing the user's navigational behavior and User Profiles created according to the user's preferences along with Web site's content to analyze and extract the information needed for the user to find the pattern expected by the user. This analysis creates a recommendation model which is presented to the user.

II. WEB PERSONALIZATION APPROACHES

Web Mining is a mining of Web data on the World Wide Web. Web Mining does the process on personalizing these Web data. The Web data may be of the following.

- Content of the Web pages (actual Web Content)
- Inter page Structure
- Usage data includes how the web pages are accessed by users
- User profile includes information collected about users (Cookies/Session data)

With personalization the content of the web pages are modified to better fit for user needs. This may involve actually creating web pages, that are unique per user or using the desires of a user to determine what web documents to retrieve. Personalization can be done to a group of specific interested customers, based on the user visits to a websites. Personalization also includes techniques such as use of cookies, use of databases, and machine learning strategies. Personalization can be viewed as a type of Clustering, Classification, or even Prediction [3].

Web Personalization and User Profile

As it has been observed that there is an explosive growth in the information available on the Web gathering useful information from the web has become a challenging issue for users. The Web users expect more intelligent systems to gather the useful information from the huge size of Web to meet their information needs. 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.

A user profile is a collection of personal data associated to a specific user. A profile refers therefore to the explicit digital representation of a person's identity. Thus the user profile can be used to store the description of the characteristics of person. A user profile can also be

described as the computer representation of a user model. 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. For example, 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 bases and to produce effective performance. The commonly used knowledge bases 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.

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.

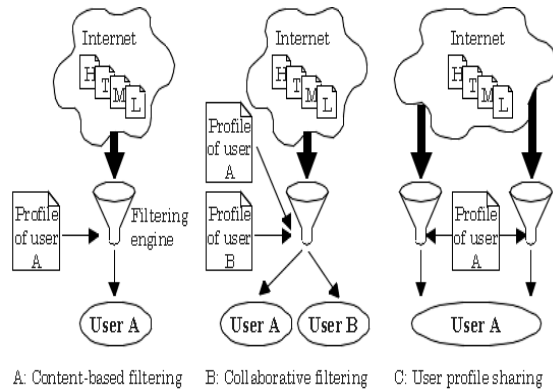


Figure 2: Different uses of Profiles

A user profile can also be shared between different personalized applications that require the same user profile's content as in 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.

Management of User Profiles:

Several architectures are used for personalized services on the Web and they differ mainly in the locations of the management and storage functions.

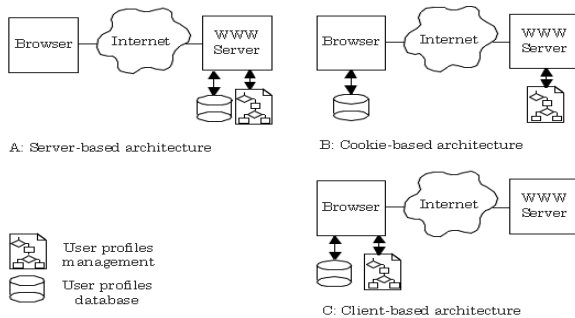


Figure 3: Common Architectures for managing User Profiles.

The most common architecture is the server-based architecture, in which the user profiles are both stored and managed at the server as in figure 3A. Since the profiles of all users are centralized, the server needs to identify the user in order to extract the right user profile. This is done by using an authentication mechanism. This architecture is efficient in that the user profiles do not transit through the network. The centralization of all the user profiles enables the use of both content-based and collaborative filtering but prevents user profiles from being shared between

applications on different servers. With this architecture, the service provider has to supply both hardware and software for the management and storage of the users' profiles. For a worldwide service, those profiles may represent a large amount of data.

The second architecture stores the user profiles on the client side and manages them on server side as given in the figure 3B. This architecture enables the use of content-based filtering and profile sharing but not of collaborative filtering. The browser must provide a mechanism for permanently storing data on the user's computer, and this is a sensitive issue because most browsers, for security purpose do not allow a Web application (for example a Java applet) to permanently store any information on the terminal. The "Cookie" mechanism introduced by Netscape is an exception to this rule. By setting a cookie, an application can get data permanently stored by the browser and automatically sent back when the user accesses the application again. The main advantage of this second architecture is the distributed nature of the storage, which frees the service provider from supplying software and disk space for the database, but the transmission of the user profile between its storage location (client) and the management location (server) increases the response delay.

The third architecture manages and stores the user profiles on the client side as represented in figure 3C. In this, the personalization is done by the browser, and the architecture is therefore not a client-server architecture anymore (at least with respect to the personalization). This architecture enables the use of content-based filtering and user profile sharing but not of collaborative filtering.

Although all these architectures enable the use of content-based filtering, none of them can at the same time support collaborative filtering and the sharing of user profiles among different applications. Furthermore, no standard such as the *Common Gateway Interface* (CGI) [9] has been defined for the management and storage of the user profiles on the server side. Each personalized Web application that uses the server-based architecture has to interface individually with the database that contains the user profiles.

Semantic based Personalized Search

Personalization aims to find a subset of Web data that matches the interest profile of a user or a group of users. This can be achieved by recommending Web pages or Websites to the users, or by filtering Web pages that are of interest to the users [10]. For example, this can be done by analyzing the historical data recording user accesses to



Web data, and mining the topics relevant to a user by clustering previously accessed Web pages based on content similarities. When a new Web page is found to be similar to one of the clusters, it can be routed to the user.

Personalized search takes advantage of Semantic Web standards (RDF and OWL) to represent the content and the user profiles. Semantic based Personalization of Web data access can be effectively used for improving the precision and recall in search, particularly by re-ranking the search results based on the learner's past activities. The core part of Semantic approach on Web Personalization is the use of Ontology. As Web pages are annotated with ontology entity labels, the Web pages accessed by a user can lead to more effective content recommendation.

Web Personalization and Ontology

An Ontology:

Ontology [11] describes a standardized representation of knowledge as a set of concepts within the domain, and the relationship between those concepts. Ontology is also used to represent user profiles in personalized web information gathering. Thus ontologies [12] are the structural frameworks for organizing information. In computer science and information science, ontology formally represents knowledge as a set of concepts within a domain, and the relationships between those concepts. It can be used to reason about the entities within that domain and may be used to describe the domain. It is worth mentioning that with the improvement of user profiles, the development of ontologies is very fast.

The Need of Ontology Model:

An Ontology [13] is the study of the nature of being, existence, as well as the basic categories of being and their relations. As a model for knowledge description and formalization, ontologies are widely used to represent user profiles.

Reasons for developing an ontology:

- To explicit the knowledge contained within software applications, and within enterprises and business procedures for a particular domain.
- To reuse of the domain knowledge
- To separate the domain knowledge from the current databases

Advantages of Ontology model in User Profile:

An Ontology model discovered user background knowledge from user local instance repositories, rather than documents read and judged by users.

Compared to the web data used by the web model, the Ontology model were controlled and contained less uncertainties.

Large number of uncertainties were eliminated when user background knowledge was discovered. As a result, the user profiles acquired by the Ontology model performed better than the web model.

Personalized Ontology:

Personalized ontologies [14] are a conceptualization model that formally describes and specifies user background knowledge. Web users might have different expectations for the same search query [15]. For example, for the topic "Apple", an IT person may demand different information from normal users. An IT person expects "Apple" as system but normal users consider this as fruit. Sometimes even the same user may have different expectations for the same search query if applied in a different situation. Based on this observation, an assumption is formed that web users have a personal concept model for their information needs. A user's concept model may change according to different information needs.

III. WEB PERSONALIZATION AND RELATED WORKS

Lot of research had been conducted in Personalized Ontology. Generally, personalization methodologies are divided into two complementary processes which are (1) the user information collection, used to describe the user interests and (2) 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 [16]. Or, in other techniques, they are used to infer relationships like the social-based filtering [17] and the collaborative filtering [18]. For the second process, extraction of information on users' navigations from system log files can be used [19]. Some information retrieval techniques are based on user contextual information extraction [20]. 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 [21]. 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 [22]. 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 [23]. Semantic similarity is the evaluation of semantic links existing between two concepts. [24] 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 [25].

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. These limitations explain the importance given to hybrid approaches. The hybridization of existing approaches is presented as an alternative that would improve the quality of personalized systems [26].

Dai and Mobasher [27] proposed a web personalization framework that characterizes the usage profiles of a collaborative filtering system using ontologies. These profiles 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 Oberle et.al. [28]. 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 web logs 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 content is 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 [29] 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. [30] 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 [31] 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 as well as between user preferences and other items on the web site, two basic operations carried out in content and collaborative filtering based recommendations. The similarity measure they employ is very similar to the Wu & Palmer similarity measure presented here. 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.

IV. CONCLUSION

Although the World Wide Web is the largest source of electronic information, it lacks with effective methods for retrieving, filtering, and displaying the information that is exactly needed by each user. With the advent of the Internet, there is a dramatic growth of data available on the World Wide Web. Hence the task of retrieving the only required information keeps becoming more and more difficult and time consuming. To reduce information overload and create customer loyalty, Web Personalization, a significant tool that provides the users with important competitive advantages is required. A Personalized Information Retrieval approach that is mainly based on the end user modeling increases user satisfaction. Also personalizing web search results has been proved as to 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.

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 And Micheline Kamber "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] "CGI ocumentation", <http://hoohoo.ncsa.uiuc.edu/cgi/>
- [10] Xiaohui Tao, Yuefeng Li, and Ning Zhong, "Senior Member, IEEE, "A Personalized Ontology Model for Web Information Gathering", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 4, APRIL 2011.
- [11] http://en.wikipedia.org/wiki/Ontology_%28information_science%29
- [12] <http://www.unicist.org/what-is-an-ontology.pdf>
- [13] <http://en.wikipedia.org/wiki/Ontology>
- [14] Bhaganagare Ravishankar, Dharmadhikari Dipa, "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.
- [15] Xiaohui Tao, Yuefeng Li, and Ning Zhong, Senior Member, IEEE. "A Personalized Ontology Model for Web Information Gathering", IEEE transactions on Knowledge and Data Engineering, Vol. 23, No. 4, April 2011.
- [16] 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).
- [17] Mladenic D., "Text-learning and Related Intelligent Agents: a Survey", IEEE Intelligent Systems, 14(4):44-54 (1999).
- [18] Goldberg, K., Roeder, T., Gupta, D., Perkins, C.: Eigentaste, "A Constant Time Collaborative Filtering Algorithm. Information Retrieval", Journal, 4(2):133-151, (2001).
- [19] Paulakis, S., Lampos, C., Eirinaki, M., Vazirgiannis, M.: Sewep, "A 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).
- [20] Jones, G.J.F., Brown, P.J., "Context-Aware Retrieval for Ubiquitous Computing Environments", Invited paper in Mobile and Ubiquitous Information Access, LNCS, vol. 2954, pp. 227-243. Springer (2004).
- [21] 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).
- [22] 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).
- [23] 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).
- [24] 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).
- [25] Maceachren, A.M., Kraak, M.J., "Research Challenges in Geovisualization, Cartography and Geographic", Information Science, 28(1):3-12 (2001).
- [26] Burke, R., "Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction", 12(4):331-370 (2002).
- [27] 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.



- [28] D.Oberle, B.Berendt, A.Hotho, J.Gonzalez, "Conceptual User Tracking", in Proc. of the 1st Atlantic Web Intelligence Conf. (AWIC), 2003.
- [29] S. Acharyya, J. Ghosh, "Context-Sensitive Modeling of Web Surfing Behaviour Using Concept Trees", in Proc. of the 5th WEBKDD Workshop, Washington, August 2003.
- [30] 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.
- [31] 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), Endinburgh, Scotland, August 2005.

BIOGRAPHY

K. Sridevi received her M.Phil. degree from Bharathidasan University, Trichy in Dec 2004. She is presently working as Assistant Professor in Computer Science Department, Nehru Memorial College, Puthanampatti, Trichy District, Tamil Nadu, India. She is currently guiding M.C.A., M.Phil. research scholars of various universities. Her area of research is Data and Web Mining.

Dr. R. Uma Rani received her Ph.D., Degree from Periyar University, Salem in the year 2006. She is a rank holder in M.C.A. from NIT, Trichy. She has published around 60 papers in reputed journals and National and International Conferences. She has received the best paper award from VIT, Vellore, Tamil Nadu in an International Conference. She has done one MRP funded by UGC. She has acted as Resource Person in various National and International Conferences. She is currently guiding 15 Ph.D. scholars. She has guided 30 M.Phil. scholars and currently guiding 4 M.Phil. Scholars. Here areas of interest include Information Security, Data Mining, Fuzzy Logic and Mobile Computing.