ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

Web Usage Mining: Personalization of Web Usage Data

Richa Soni[1], Gurpreet Kaur[2]

Student, Department of Computer Science & Engineering Chandigarh University (Gharuan, Mohali) India[1]

Assistant Professor, Department of Computer Science & Engineering Chandigarh University (Gharuan, Mohali) India[2]

Abstract: In this paper, we present a complete framework and findings in mining Web usage patterns from Web log files of a real Web site that has all the challenging aspects of real-life Web usage mining, including evolving user profiles and external data describing an ontology of the Web content. we present an approach for discovering and tracking evolving user profiles. We also describe how the discovered user profiles can be enriched with explicit information need that is inferred from search queries extracted from Web log data. Profiles are also enriched with other domain-specific information facets that give a panoramic view of the discovered mass usage modes. paper presents a knowledge discovery framework for the construction of Community Web Directories, a concept that we introduced in our recent work, applying personalization to Web directories. In this context, the Web directory is viewed as a thematic hierarchy and personalization is realized by constructing user community models on the basis of usage data. we enhance the clustering and probabilistic approaches presented in previous work and also present a new algorithm that combines these two approaches. The resulting community models take the form of Community Web Directories. The proposed personalization methodology is evaluated both on a specialized artificial and a general-purpose Web directory, indicating its potential value to the Web user. Web mining techniques seek to extract knowledge from Web data. This paper provides an overview of past and current work in the three main areas of Web mining research—content, structure, and usage as well as emerging work in Semantic Web mining. Statistical testing and reliability analysis can be used effectively to assure quality for Web applications. To support this strategy, we extract Web usage and failure information from existing Web logs. The usage information is used to build models for statistical Web testing. Optimizing components before optimizing the system as a whole can help large organizations deploy efficient, geographically redundant Web infrastructures.

Keywords: Clustering, Mining Evolving Clickstreams, Machine Learning, Personalization, Reliability Analysis, Statistical Testing, Semantic Web Mining, User Profiles, Usage Measurement, Web Usage Mining, Web Mining, World Wide Web.

I. INTRODUCTION

Customer Relationship Management (CRM) can use data from within and outside an organization to allow an understanding of its customers on an individual basis or on a group basis such as by forming customer profiles. An improved understanding of the customer's habits, needs, and interests can allow the business to profit by, for instance, "cross selling" or selling items related to the ones that the customer wants to purchase. Hence, reliable knowledge about the customers' preferences and needs forms the basis for effective CRM. Web usage mining techniques that can automatically extract frequent access patterns from the history of previous user click streams stored in Web log files. the Web has not achieved its goal of providing easy access to online information. As its size is increasing, the abundance of available information on the Web causes the frustrating phenomenon of "information overload" to Web users. Organization of the Web content into thematic hierarchies is an attempt to alleviate the problem. These hierarchies are known as Web Directories and correspond to listings of topics which are organized and overseen by humans. A Web directory, such as Yahoo www.yahoo.com) and the Open Directory Project (ODP) (dmoz.org), allows users to find Web sites related to the topic they are interested in, by starting with broad categories and gradually narrowing down, choosing the category most related to their interests. Web Personalization [1], i.e., the task of making Web-based information systems adaptive to the needs and interests of individual users, or groups of users, emerges as an important means to tackle information overload. However, in achieving personalization, we are confronted with the difficult task of acquiring and creating accurate and operational user models. we can overcome the deficiencies of Web directories and Web personalization by combining their strengths, providing a new tool to fight information overload. In particular, we focus on the construction of usable Web directories that model the interests of groups of users, known as user communities.

The application of data-mining techniques to extract knowledge from Web data, in which at least one of structure or usage (Web log) data is used in the mining process (with or without other types of Web data). the World Wide Web is fertile ground for data mining principles, or *Web mining*. The Web mining field encompasses a wide array of issues, primarily aimed at deriving actionable knowledge from the Web, and includes researchers from information retrieval, database technologies, and artificial intelligence.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

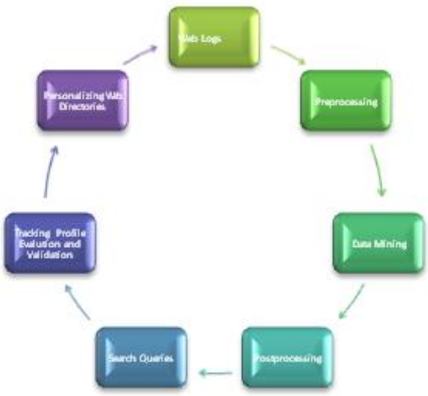

Figure. 1 Web Usage Mining Process with personalizing Web Directories.

WEB applications provide cross-platform universal access to Web resources for the massive user population. With the prevalence of the World Wide Web, quality assurance for the Web is becoming increasingly important.

The rest of the paper is organized as follows: In section 2, Overview of Web Uses Mining with characteristics and applications. in section 3, discuss emerging work in semantic Web Mining with Content, Structure and uses. in section 4, Personalizing Web directories with the Aid of Web Usage Data and finally, conclusion and future direction in section 5.

II. OVERVIEW OF WEB USES MINING WITH CHARACTERISTICS AND APPLICATIONS.

A. Overview of Web Uses Mining

Web usage mining, is traditionally performed in several stages [1], [3] to achieve its goals:

- Collection of Web data such as activities/clickstreams recorded in Web server logs,
- Preprocessing of Web data such as filtering crawlers requests, requests to graphics, and identifying unique sessions,
- Analysis of Web data, also known as Web Usage Mining [4], to discover interesting usage patterns or profiles, and
- Interpretation/evaluation of the discovered profiles. In this paper, we further added a fifth step after a repetitive application of steps 1-4 on multiple time periods, i.e.,
- Tracking the evolution of the discovered profiles.

Web usage mining can use various data mining or machine learning techniques to model and understand Web user activity. In [6], clustering was used to segment user sessions into clusters or profiles that can later form the basis for personalization. In [7], the notion of an adaptive Web site was proposed, where the user's access pattern can be used to automatically synthesize index pages. The work in [1] is based on using association rule discovery as the basis for modeling Web user activity, whereas the approach proposed in [8] used probabilistic grammars to model Web navigation patterns for the purpose of prediction. we present a complete framework and a summary of our experience in mining Web usage patterns with real-world challenges such as evolving access patterns, dynamic pages, and external data describing an ontology of the Web content and how it relates to the business actors (in the case of the studied Web site, the companies, contractors, consultants, etc., in corrosion). The Web site in this study is a portal that provides access to news, events, resources, company information (such as companies or contractors supplying related products and services), and a library of technical and regulatory documentation related to corrosion and surface treatment. The portal also offers a virtual meeting place between companies or organizations seeking information about other companies or organizations.

B. Characteristics And Applications.

Web applications possess various unique characteristics that make Web testing and quality assurance different from corresponding traditional techniques. Web applications can typically be characterized by the following:

- Massive user population. Web applications provide cross-platform universal access to Web resources for the massive user population. Although some traditional software systems, such as operating systems, also serve a massive user population, the systems are usually accessed locally, thus scattering the user population into subgroups of limited Size.
- Diverse usage environments. Web users employ different hardware equipments, network connections, operating systems, middleware and Web server support, and Web browsers, as compared to prespecified platforms for most traditional software.
- Document and information focus, as compared to the computational focus for most traditional software.

Although some computational capability has evolved in newer Web applications, document and information search and retrieval still remain the dominant usage for most Web users. Because of the user focus and the large size of the Web, one good candidate for effective Web quality assurance is statistical testing and related reliability analysis [10], [18]. These techniques can help us prioritize testing effort based on usage scenarios and frequencies for individual Web resources and navigation patterns to ensure the reliability of Web applications. more and more Web applications are increasingly becoming multilayered, with functionalities distributed across different layers and subsystems.

C. Web Failures

Quality assurance and testing for Web applications focus on the prevention of Web failures or the reduction of chances for such failures. We define Web failures as the inability to correctly deliver information or documents required by Web users. we can consider the following failure sources associated with different Web-layers:

- Host or network failures. Hardware or systems failures at the destination host or home host, as well as network failures, may lead to Web failures. These failures are mostly linked to middleware and Web server layers in Fig. 1. However, such failures are no

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

different from the regular system or network failures, and can be analyzed by existing techniques. Therefore, these failure sources are not the focus of our study.

- Browser failures. Browser failures are linked to problems at the highest layer in Fig. 1 on the client side. These failures can be treated the same way as software product failures, thus, existing techniques for software testing and reliability analysis [4], [10] can be used to assess, predict, and improve browser reliability. However, for Web applications, there is a strong emphasis on compatibility between different Web browsers or different browser versions and the information source. This issue will be considered later when we discuss browser rendering problems.
- Source or content failures. Web failures can also be caused by the information source itself at the server side, associated with the lowest layer in Fig. 1. In our strategy for Web testing, we will primarily deal with this kind of Web failures.

III. DISCUSS EMERGING WORK IN SEMANTIC WEB MINING WITH CONTENT, STRUCTURE AND USES.

As a large and dynamic information source that is structurally complex and ever growing, the World Wide Web is fertile ground for data mining principles, or *Web mining*.

Researchers have identified three broad categories of Web mining:2,3

- *Web content mining* is the application of data mining techniques to content published on the Internet, usually as HTML (semistructured), plaintext (unstructured), or XML (structured) documents.
- *Web structure mining* operates on the Web's hyperlink structure. This graph structure can provide information about a page's ranking4 or authoritativeness5 and enhance search results through filtering.
- *Web usage mining analyzes* results of user interactions with a Web server, including Web logs, clickstreams, and database transactions at a Web site or a group of related sites. Web usage mining introduces privacy concerns and is currently the topic of extensive debate.

A. Web Content and Structure Mining

Some researchers combine content and structure mining to leverage the techniques' strengths. Although not all researchers agree to such a classification, we list research in these two areas together.

Web as a Database

Early work in the area of Web databases focused on the Web's layered view, as suggested by Osmar Zaiane and colleagues.9 Placing a layer of abstraction containing some semantic information on top of the semistructured Web lets users query the Web as they would a database.

Hubs and Authorities

Hyperlink-induced topic search (HITS) is an iterative algorithm for mining the Web graph to identify topic *hubs* and *authorities*. "Authorities" are highly ranked pages for a given topic; "hubs" are pages with links to authorities. The algorithm takes as input search results returned by traditional text indexing techniques, and filters these results to identify hubs and authorities. The number and weight of hubs pointing to a page determine the page's authority. The algorithm assigns weight to a hub based on the authoritativeness of the pages it points to.

Document Classification

Classification's roots are in machine learning, pattern recognition, and text analysis. The basic idea is to classify pages using supervised or unsupervised methods. In simple terms, supervised learning uses preclassified training data, which is not required in unsupervised learning. Classification is useful in such areas as topic aggregation and Web community identification. Early work in document classification applied text-mining techniques to Web data directly. (Text mining is a subcategory of Web content mining that does not use Web structure.) Later research showed that harnessing the Web graph structure and semistructured content in the form of HTML tags improved results.

Identifying Web Communities

Many communities are well organized on the Web, with *webrings* (interlinks between Web sites with a ring structure) or information portals linking them together. The community core represents those Web sites that are part of the same community without links between themselves. Trawling is the process of identifying such subgraphs from the Web graph.

Clever: Ranking by Content

Basic hub and authority approaches do not consider a link's semantics for page ranking. The Clever15 system addresses this problem by considering query terms occurring in or near the anchor text (a certain window) in an HTML page as a hint to link semantics, and thus leverages content-mining techniques for structure analysis. Clever gives greater weight to links that are similar to the search query.

B. Web usage Mining

Web usage mining has several applications in e-business, including personalization, traffic analysis, and targeted advertising. The main areas of research in this domain are Web log data preprocessing and identification of useful patterns from this preprocessed data using mining techniques. Most data used for mining is collected from Web servers, clients, proxy servers, or server databases, all of which generate noisy data. Because Web mining is sensitive to noise, data cleaning methods are necessary.

Adaptive Web Sites

Personalization is one of the most widely researched areas in Web usage mining. An early effort in this direction was the adaptive Web site challenge posed by Oren Etzioni and colleagues. 21 Adaptive sites automatically change their organization and presentation according to the preferences of the user accessing them. Other contemporary research seeks to build agent-based systems that give user recommendations. All these approaches primarily use association rules and clustering mechanisms on log data and Web pages.

Robust Fuzzy Clustering

Anupam Joshi and colleagues23 use fuzzy techniques for Web page clustering and usage mining, and they use the mined knowledge to create adaptive Web sites.24 They argue that given the inherent ambiguity and complexity of the underlying data, clustering results should not be clearly demarcated sets but rather fuzzy sets—that is, overlapping clusters. For instance, a user can belong to multiple user interest groups because at different times he or she accesses the Web for different information or merchandise. Insisting that each user fit only a single group is clearly inconsistent with this reality. Moreover, given the noise expected in the data despite cleaning attempts, the clustering process must be robust in the statistical sense. Raghu Krishnapuram and colleagues discuss fuzzy clustering and its application to Web-log analysis and present a fast linear clustering algorithm that can handle significant data noise.24 They use this algorithm to cluster Web access logs and use the traversal patterns identified for specific groups to automatically adapt the Web site to those groups.

Association Rules

Early systems used collaborative filtering for user recommendation and personalization. Bamshad Mobasher and colleagues25 used association-rule mining based on frequent item sets and introduced a data structure to store the item sets. They split Web logs into user sessions and then mined these sessions using their suggested association rule algorithm. They argue that other techniques based on association rules for usage data do not satisfy the real-time constraints of recommender systems because they consider all association rules prior to making a recommendation. Ming-Syan Chen and colleagues26 proposed a somewhat similar approach that uses a different frequent item set counting algorithm.

Recommender Systems

J. Ben Schafer and colleagues27 note that recommender systems have enhanced e-business by

- converting browsers to buyers,
- increasing cross-sell by identifying related products, and
- building loyalty.

These systems primarily use association rule mining for pattern detection. In an e-business scenario, a recommender system uses customers' Web baskets (shopping carts) as data sources.

Web Site Evaluation

Myra Spiliopoulou28 suggests applying Web usage mining to Web site evaluation to determine needed modifications—primarily to the site's design of page content and link structure between pages. Such evaluation is one of the earliest steps in Web usage analysis conducted by Web sites and is necessary for repeat visitors. Evaluation is important because all subsequent Web usage mining techniques are effective only in the presence of large amounts of data created by repeat visitors.

Hamlet: To Buy or Not to Buy

Etzioni and colleagues29 applied Web mining to airline ticket purchasing. Airlines use sophisticated techniques to manage yield, varying ticket prices according to time and capacity. Etzioni's approach mined airline prices available on the Web and price changes over time to produce recommendations regarding the best time to buy tickets. Many more innovative areas are yet to be explored.

Privacy Issues

Double Click's (www.doubleclick.com) online advertising is an instance of tracking user behavior across multiple sites. If a user's transactions at every Web site are identified through uniquely identifiable information collected by Web logs, they could create a far more complete profile of the user's shopping habits. However, Web sites and popular Web browsers offer limited support for such tools. Web mining research should accommodate this preference set and enforce it across organizations and databases. Lorrie Cranor surveys possible research direction in this area.

C. Semantic Web Mining

In the Semantic Web, adding semantics to a Web resource is accomplished through explicit annotation (based on an ontology). Humans cannot be expected to annotate Web resources; it is simply not scalable. Hence, we need to automate the annotation process through ontology learning, mapping, merging, and instance learning. Web content-mining techniques can accomplish this. For instance, we can use topic classification to automatically annotate Web pages with information about topics in an ontology. Annotations of this kind enable new possibilities for Web mining. Ontologies can help improve clustering results through feature selection and aggregation (for instance, identifying that two different URLs both point to the same airfare search engine). Web-as-database approach, which is limited to simple metadata (topics, author, creation date, and so on). Moreover, these assertions will be in a language with explicit semantics, making it machine interpretable. Web-as-database approach, which is limited to simple metadata (topics, author, creation date, and so on). Moreover, these assertions will be in a language with explicit semantics, making it machine interpretable. Similarly, aggregating data in a central site and then mining it is rarely scalable, hence the need for distributed mining techniques. Finally, researchers will need to leverage the semantic information the Semantic Web provides. Exposing content semantics and the link explicitly can help in many tasks, including mining the hidden Web—that is, data stored in databases and not accessible through search engines.

IV. PERSONALIZING WEB DIRECTORIES WITH THE AID OF WEB USAGE DATA

We enhance the clustering and probabilistic approaches presented in previous work and also present a new algorithm that combines these two approaches. The

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

resulting community models take the form of Community Web Directories. The proposed personalization methodology is evaluated both on a specialized artificial and a general-purpose Web directory, indicating its potential value to the Web user.

A. Discovery of Web Directories from Web Usage Data.

User communities

User communities are formed using data collected from Web proxies as users browse the Web. The goal is to identify interesting behavioral patterns in the collected usage data and construct community Web directories based on those patterns. The process of getting from the data to the community Web directories is summarized below:

- Usage Data Preparation comprises the collection and cleaning of the usage data, as well as the identification of user sessions.
- Web Directory Initialization provides the characterization of the Web pages included in the usage data, according to the categories of a Web directory. We compare two different approaches for the characterization of the Web pages. The first approach organizes Web pages into an artificial Web directory using hierarchical document clustering. The second approach classifies them onto an existing Web directory, like ODP.
- Community Web Directory Discovery is the main process of discovering the user models from data, using machine learning techniques and exploiting these models to build the community Web directories. The construction of community Web directories is a fullyautomated process, resulting in operational personalization knowledge, in the form of user models.

B. Usage Data Preparation

The usage data that form the basis for the construction of community Web directories are collected in the access log files of ISP cache proxy servers. These data record the navigation of the subscribers through the Web, and hence, they are usually diverse and voluminous. The outgoing traffic is much higher than the usual incoming traffic of a Web site and the Web pages more diverse semantically. The task of usage data preparation, detailed in [8], [9], is to assemble these data into a consistent, integrated, and concise view. The next stage is the identification of individual user sessions. The fact that we are focusing on the discovery of behavioral patterns in the data, rather than individual users, allowed us to overcome the lack of user registration data or other means of user identification, such as cookies, and led us to exploit a simple kind of user session. A user session is defined as a sequence of log entries, i.e., accesses to Web pages by the same IP address, where the time interval between two subsequent entries does not exceed a certain time threshold.

C. Community Web Directory Discovery

The mapping and the associations between Web pages, user sessions, and Web directory categories, we employ unsupervised learning to discover patterns of interest in the thematic user sessions. In our recent work, we employed two methods for the discovery of community Web directories. In [8], we presented an extension of the cluster mining algorithm, named Community Directory Miner (CDM), while in [9], we presented an approach based on the discovery of latent semantics using PLSA. These algorithms are used to extract a subset of the categories of the initial Web directory that correspond to the community models, i.e., usage patterns that occur in data and represent the browsing preferences of community members. Each community model _ is subsequently exploited to construct the community Web directory. The general process of discovering community Web directories can be seen as a construction of the subgraph G' of the Web directory G which corresponds to the community Web directory.

V. CONCLUSION AND FUTURE DIRECTIONS

This paper advocates the concept of a community Web directory, as a Web directory that specializes to the needs and interests of particular user communities. Furthermore, it presents the complete methodology for the construction of such directories with the aid of machine learning methods. User community models take the form of thematic hierarchies and are constructed by employing clustering and probabilistic learning approaches. We applied our methodology to the ODP directory, as well as to an artificial Web directory, which was generated by clustering Web pages that appear in the access log of a Web proxy. For the discovery of the community models, we introduced a new criterion that combines the a priori thematic informativeness of the Web directory categories with the level of interest observed in the usage data. The proposed methodology addresses these issues by reducing the dimensionality of the problem, through the classification of individual Web pages into the categories of the directory. we have developed an approach for statistical Web testing and reliability analysis supported by automated information extraction from existing Web logs. There are many issues we need to address in the future, including: capturing various missing Web usage information for UMM construction, usage time measurement for better reliability modeling, long-term quality tracking and reliability modeling, UMM state-related reliability information calculation, and risk identification for focused reliability improvement. The acquired results lead us to the conclusion that although we have obtained good performance by all methods, the use of PLSA for the personalization of Web directories appears to be the most promising. It helps identifying latent information in the users' choices and derives high-quality community directories that provide significant benefits to their users. However, we have only approximated the gain of the end user and have not taken into account the cost of "losses" that could be encountered in the case that the users do not find what they are looking for in the personalized directory. This issue requires the evaluation of community Web directories in user studies which are in our immediate plans for future work. The proposed methodology provides a promising research

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

direction, where many new issues arise. An analysis regarding the parameters of the community models, such as PLSA, is required.

REFERENCES

[1] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic Personalization Based on Web Usage Mining," Comm. ACM, vol. 43, no. 8, pp. 142-151, 2000.

[2] J. Srivastava, R. Cooley, M. Deshpande, and P.T. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," SIGKDD Explorations, vol. 1, no. 2, pp. 12-23, 2000.

[3] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C.D. Spyropoulos, "Web Usage Mining as a Tool for Personalization: A Survey," User Modeling and User-Adapted Interaction, vol. 13, no. 4, pp. 311-372, 2003.

[4] G. Paliouras, C. Papatheodorou, V. Karkaletsis, and C.D. Spyropoulos, "Discovering User Communities on the Internet Using Unsupervised Machine Learning Techniques," Interacting with Computers J., vol. 14, no. 6, pp. 761-791, 2002.

[5] G. Xu, Y. Zhang, and Y. Xun, "Modeling User Behaviour for Web Recommendation Using lda Model," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence and Intelligent Agent Technology, pp. 529-532, 2008.

[6] W. Chu and S.-T.P. Park, "Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models," Proc. 18th Int'l Conf. World Wide Web (WWW), pp. 691-700, 2009. PIERRAKOS AND PALIOURAS: PERSONALIZING WEB DIRECTORIES WITH THE AID OF WEB USAGE DATA 1343

[7] The Adaptive Web, Methods and Strategies of Web Personalization, P. Brusilovsky, A. Kobsa, and W. Neijdl, eds. Springer, 2007.

[8] O. Etzioni, "The World Wide Web: Quagmire or Gold Mine?," *Comm. ACM*, vol. 39, no.11, 1996, pp. 65–68.

[9] J. Srivastava, P. Desikan, and V. Kumar, "Web Mining: Accomplishments and Future Directions," *Proc. US Nat'l Science Foundation Workshop on Next-Generation Data Mining* (NGDM), Nat'l Science Foundation, 2002.

[10] R. Kosala and H. Blockeel, "Web Mining Research: A Survey," *ACM SIGKDD Explorations*, vol. 2, no. 1, 2000, pp. 1–15.

[11] L. Page et al., *The PageRank Citation Ranking: Bring Order to the Web*, tech. report, Stanford Digital Library Technologies, 1999-0120, Jan. 1998.

[12] J. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *Proc. 9th Ann. ACM–SIAM Symp. Discrete Algorithms*, ACM Press, 1998, pp. 668–677.

[13] F. Sebastini, *Machine Learning in Automated Text Categorization*, tech. report B4-31, Istituto di Elaborazione dell'Informatione, Consiglio Nazionale delle Ricerche, Pisa, 1999.

[14] S. Chakrabarti, "Data Mining for Hypertext: A Tutorial Survey," *ACM SIGKDD Explorations*, vol. 1, no. 2, pp. 1–11, 2000.

[15] J. Fürnkranz, "Web Structure Mining: Exploiting the Graph Structure of the World Wide Web," *Österreichische Gesellschaft für Artificial Intelligence* (ÖGAI), vol. 21, no. 2, 2002, pp. 17–26.

[16] O.R. Zaïane and J. Han, "Resource and Knowledge Discovery in Global Information Systems: A Preliminary Design and Experiment," *Proc. 1st Int'l Conf. Knowledge Discovery and Data Mining* (KDD), AAAI Press**,** 1995, pp. 331–336.

[17] L.V.S. Lakshmanan, F. Sadri, and I.N. Subramanian, "A Declarative Language for Querying and Restructuring the Web," Proc. *6th IEEE Int'l Workshop Research Issues in Data Eng., Interoperability of Nontraditional Database Systems* (RIDE-NDS), IEEE CS Press, 1996.

[18] A. Mendelzon, G. Michaila, and T. Milo, "Querying the World Wide Web," *Proc. 1st Int'l Conf. Parallel and Distributed Information System*, IEEE CS Press, 1996, pp. 80–91**.**

[19] T. Bemers-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific Am.*, vol. 279, no. 5, 2001, pp. 34–43.

[20] T. Hofmann, "Learning What People (Don't) Want," Proc. 12th European Conf. in Machine Learning, pp. 214-225, 2001.

[21] X. Jin, Y. Zhou, and B. Mobasher, "Web Usage Mining Based on Probabilistic Latent Semantic Analysis," Proc. ACM SIGKDD, pp. 197-205, Aug. 2004.

[22] X. Jin, Y. Zhou, and B. Mobasher, "Task-Oriented Web User Modeling for Recommendation," Proc. 10th Int'l Conf. User Modeling, L. Ardissono, P. Brna, and A. Mitrovic, eds., pp. 109-118, 2005.

[23] D. Chen, D. Wang, and F. Yu, "A PLSA-Based Approach for Building User Profile and Implementing Personalized Recommendation," Proc. Joint Ninth Asia-Pacific Web Conf. (APWeb '07) and Eighth Int'l Conf. Web-Age Information Management (WAIM '07),pp. 606-613, 2007.

[24] B. Mehta and N. Wolfgang, "Unsupervised Strategies for Shilling Detection and Robust Collaborative Filtering," User Modeling and User-Adapted Interaction, vol. 19, nos. 1/2, pp. 65-97, 2009.

[25] P.A. Chirita, W. Nejdl, R. Paiu, and C. Kohlschu¨ tter, "Using odp Metadata to Personalize Search," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 178-185, 2005.

[26] A. Sieg, B. Mobasher, and R. Burke, "Ontological User Profiles for Representing Context in Web Search," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence and Intelligent Agent Technology—Workshops, pp. 91-94, 2007.

[27] Z. Ma, G. Pant, and O.R.L. Sheng, "Interest-Based Personalized Search," ACM Trans. Information Systems, vol. 25, no. 1, article no. 5, Feb. 2007.

[28] T. Oishi, K. Yoshiaki, M. Tsunenori, H. Ryuzo, F. Hiroshi, and M. Koshimura, "Personalized Search Using odp-Based User Profiles Created from User Bookmark," Proc. 10th Pacific Rim Int'l Conf. Artificial Intelligence, pp. 839-848, 2008.