Redundancy Removal in Multiserver Web Datamining for E-Business Applications

R.Sailaja, K.Sharmila, K.SelvaKumar

Abstract—The information on the web is growing day by day. With the huge amount of information and developments in the technology available online, the www is a fertile area which lead to the growth of e-business. As e-business is widely applied, web data mining technology is used for e- business to meet the queries and requirements of users. This paper proposes a normalized multi-server web mining technique which provides data to the user with the required result without redundancy. The proposed scheme reduces the complexity of the search result by focusing on normalization algorithm to remove repeated entries and page ranking technique for fast and effective services.

Index Terms— Normalization, Personalized Search, User Profile, Web Data Mining.

I. INTRODUCTION

 The World Wide Web is a popular and interactive medium to disseminate information today. The web as well all know stores large collection of data,

 Diverse and dynamic: It serves as a platform for a business oriented environment. Every day the web grows by roughly a million of electronic pages adding to the growth of relevant information. companies, organizations and individuals alike to gather information through web data mining to utilize that information in their best interest. The concept of personalized information services in existing application rely on mining information from single database extracted to have limited results and redundant occurrences. But for personalized business oriented mining it may not be sufficient for a user.

 Personalized search works in this context because it helps the user to find the respective result, but it also empowers e-business more importantly provides with the ability to remove redundancy. Due to these search situations we are currently drowning information and facing information in a repetitive manner. People either browse or use the search service according to the interest and requirement of the user. If a particular user checks for a commodity in single database there will be a number of search results. But in the case of more than one database for the same search the reoccurrence of the product or commodity can occur.   To overcome this situation for a personalized search in business applications

   Mrs. R. Sailaja, B. Sc, M.C.A., Research Scholor, School of Computing Sciences, Vels University, Chennai, India.
   Mrs. K.Sharmila, B.Sc, M.Phil, Assistant Professor, Department of B.C.A. &IT,Vels University, Chennai, India.
   Dr. K. Selva Kumar M.E. Ph.D., Assistant Professor, Department of Mathematics, University VOC College of Engineering, Anna University (Thoothukudi Campus), Thoothukudi, India.

we propose a redundant free search result from different databases. Actually in the proposed system the search issue can be solved using web mining technique by applying normalization directly to make a search personalized one. According to the personalized search, previous studies of Feng qiu, Junghoo Cho[2] proposed the concept of learning a users interest automatically based on her past click history, and how we can use the learned interest to personalize search results for future queries. Finally they developed an effective ranking mechanism that considers the learned interest of the user in generating the search result. Search engine users information needs are diverse. Even for the same query different users might express different preferences over the retrieved documents, resulting in distinct ranking requirements for search results.Hongning wang and Xiaodong[4], proposed a general ranking adaption model framework for personalized search. Using a given user independent ranking model trained offline and limited number of adaption queries from individual users, the framework quickly learns to apply a series of linear transformations.Extensive experimentation based on large set of search logs from a major commercial web search engine confirms the effectiveness of the proposed method compared to another ranking model adaption method. Ahu sieg,bamshad mobasher[6] proposed another efficient personalized search as an approach of user context as ontological profiles by assigning simplicity derived interest scores to existing concepts in a domain ontology. A spreading activation algorithm is used to maintain the interest scores based on the users on-going behavior. Researchers show that re-ranking search results based on the interest scores and the evidence in an ontological user profile is effective in presenting the most relevant result to the user.Research on personalizing retrieval has found that implicitly gathered information such as browser history, query history ,and desktop information, can be used to improve the ranking of search results for a given user. Short-term behavior from within the current search session has been used for tasks such as search result ranking or predicting future search interests. Although personalized search has been proposed for many years and many personalization strategies have been investigated, it is still unclear whether personalization is consistently effective on different queries for different users, andunder different search contexts.According to Zhicheng Dou,Ruihua Song,Ji-Rong Wen [16] personalized search has different effectiveness on different queries. Click entropy can be used as a simple measurement on whether the query should be personalized.   Experimental results also show that

click-based personalization strategies work well. Teevan etal found that the performance of the personalization algorithm they studied was improved as more data became available about the target user's interests. Long-term behavior has been used for personalizing search result ranking by building long-terms models of search interests including specifically using previous queries suggesting a pursuit of similar information needs.

In Sec.2,.we discussed about Personalized E-Business System,Sec3,.dicussion about Search Categories,Sec4–ArchitectureandSystemFlow,Sec5-Experimental Results and Sec.6-Performance Metrics

II. PERSONALIZED E-BUSINESS SYSTEM.

A. Concept

Personalized e-business site consists of various services which includes Commodity search, commodity recommendations, commodity customization. Research on Personalized search has recently got significant attention in addressing the challenge in the web search community. Personalized search something deals with the relevant specifications given to the user query. For an e-business site the personalized search can be given to the user on the basis of commodity search, commodity recommendation, and decision support service. Whatever the requirement specified by the user the search will be refined to that extent to avoid repetitions. The universal information services are giving gradually way to personalized information services that, according to the interest, status and requirements of a user so that the users are provided with the information on a targeted basis.

B. Search Categories

1. Establishment of user profiles.

2. Web data mining technique.

3. Search Normalization.

1) ESTABLISHMENT OF USER PROFILE

Establishment of user profiles is base of realizing the personalized search. In order to personalize the search, search combined with the individual. The totality of characteristics that distinguishes an individual. User profile data provide information about the users of a website usually name,age,country,marital status,education,interests and preferences .This type of information's can be acquired by means of registration forms or even his/her behaviour are traced when browsing a web site. Personalized search depends on a user profile that is unique to the individual.

2) WEB DATA MINING TECHNIQUE

Web data mining technology used as a key part in personalization. Web data mining can be used to analyse the user data and the interest in model that incorporate with user features making personalized search in the e-business site possible.

Web mining can be broadly classified into three types.

a. Web usage mining:

WUM considers general access patterns and customized access patterns. The purpose of web usage mining is to apply data mining techniques to the pre-processed web log data, in order to develop useful patterns. The overall process of Web personalization based on Web usage mining consists of three phases: data preparation and transformation, pattern discovery, and recommendation. Of these, only the latter phase is performed in real-time. The data preparation phase transforms raw Web log files into user profile or Web transaction data that can be processed by data mining tasks. This phase also includes data integration from multiple sources, such as backend databases, application servers, and site content. A variety of data mining techniques can be applied to this data in the pattern discovery phase, such as clustering, association rule mining, and sequential pattern discovery. The results of the mining phase are normalized into user models, suitable for the sequence to produce the next phase of the execution.

b. Web content mining:

WCM considers search results and web pages. Web content corresponds to the collection of web pages was designed to convey to the users. It may consist of text, images, audio, video or structured records such as tables and lists.

c. Web structure mining:

WSM considers links of all pages. This type of mining can be performed either at the (intra-page) document level or at the (inter-page) hyperlink level. This can used to retrieve useful information from the web.

3) Search Normalization

Normalization is an important activity in web mining. Web data can be retrieved in smoother way using effective normalization technique. Normalization also reduces lot of calculations in web mining activities. The proposed technique is based on content, structure and semantic similarity and web page redirection and forwarding similarity of the data collected. Normalization is the process by which searches are modified and standardised in a consistent manner. The goal of the normalization is to transform a search into a canonical search so it is possible to determine if syntactically two different searches may be equivalent. Search normalization is applied to the data e-business site in order to assign importance web pages and to reduce indexing of duplicate pages. Duplication of data is avoided by comparing parallel data removal technique.

III. ARCHITECTURE OF PERSONALIZED SEARCH IN AN E-BUSINESS SITE

The personalized e-business system makes personalized search when the user query is sent to the server databases, the query is executed in the server database which leads to fetch the data. Data is fetched by web usage mining is to apply data mining techniques to the pre-processed data in order to discover useful patterns. After discovering patterns from usage data normalization done to maintain a user database

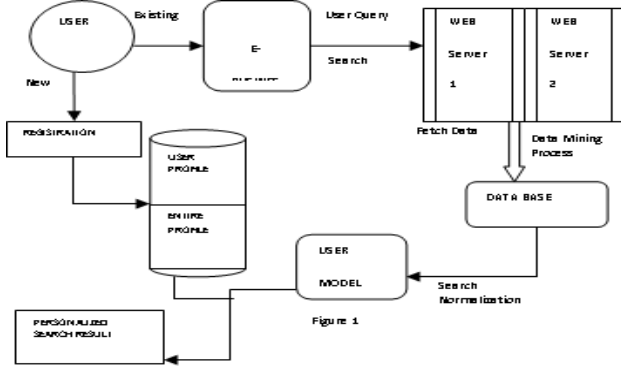which in turn to produce personalized search.

Fig.1. Architecture of personalised search in
e-business system

IV. SYSTEM FLOW

When a new user registers him/her in an e-business website platform, a user profile is created according to which the platform will be presented to the user.

A user logs on to e-business website platform. Then the user agent enables this function module on the basis of user usage information.

According to the user profile, the personalized requirements of the users are gathered and the data is organized and the exact use of the data to the user is found.

The user data is pre-processed including cleaning, conversion, integration and transformed and loaded to the pre-processing data resource bank.

By selecting suitable data mining method along with the user agent so that the user model can be built.

Based on the data mining results, the data is integrated with the expertise and area rules, normalized and the user is offered with the personalized e-business service through an efficient e-business platform.

V. EXPERIMENTAL RESULTS

We conducted a series of experiments with the search query given for an e-business envoirnment.The site contains six personalized services in which we focuses on commodity search, and commodity recommendations. For both the services, the ultimate goal is to refine the search into a broader area. By enhancing the search and recommendations to a greater extent by avoiding the reoccurrences of the search item and the recommendations whatever we are offering to the user which is based upon age group, purchase pattern, and number of visitors viewed.

Fig.2. Search based on user query with repeated entries.
The above figure 2, illustrates about the user query given, the

search produced with repeated entries.

The limitations in the existing system falls into six personalized services for an e- business site which in turn gives the services for the user on the basis of single database retrieval .The user provided with information retrieved mainly from one database without any emerging features for the future requirement. In our case the difference between the search and the services mainly exists in the database view of the project. We conclude here the search occurs from the multiple databases to produce the search result for the user.
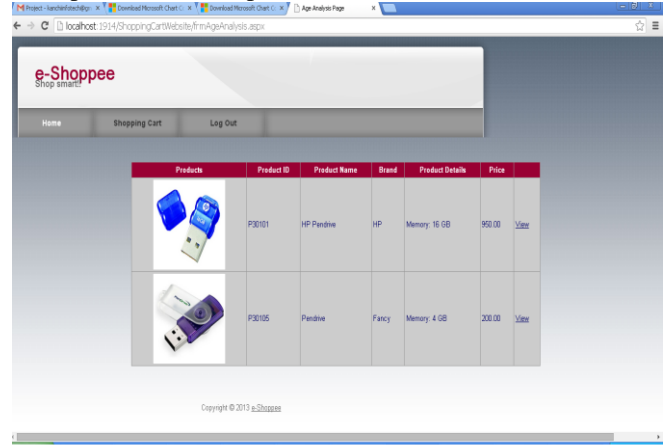

Fig.3.Search result without repeated entries.

The above figure.3 shows how the user query normalized to produce an enhanced search

VI. PERFORMANCE METRICS

For an e-business site there will be emerging techniques and technologies implied. The main drawback in the existing system relies on providing six personalized services without any user profile developing from single database. But in the proposed architecture we suggest with establishment of user profile. We fetch the data from multiple servers by web data mining techniques mainly web usage mining and match the data for parallel data removal technique so that a normalized user database can be built..We then combine these profiles with user database to produce comprehensive search entries without redundancy

VII. CONCLUSION

In this paper we achieved redundant free personalized search result for multiserver e-business application. The internet based e-business system is growing rapidly and the search based on the user profiles should be given more attention. The personalized search provides for e-business system has room for further improvement and research, giving more importance to the privacy of user data (i.e., the search data is not shared between the users) and also providing the users with optimized data which are really useful for the users.

ACKNOWLEDGEMENT

R<small>EFERENCES</small>

[1]   Yang Lin and Mao Yurong,Personalization: Customize Your
       Network Service", Software Engineer, v.7, 2003.

[2]   Feng Qiu and Junghoo Cho Automatic identification of user
       interest for personalized search (WWW '06).

[3]   ZhichengDou,RuihuaSong,andJi
       Rong Wen. A large-scale evaluation and analysis of personalized search
       strategies.(WWW '07).

[4]   .Hongning Wang, Xiaodong He,Ming-Wei Chang, Yang Song, Ryen W.
       White, and Wei Chu. 2013. Personalized ranking model adaptation for
       web search..

[5]   .Milad Shokouhi, Ryen W. White, Paul Bennett, and Filip
       Radlinski. Fighting search engine amnesia: reranking repeated results..

[6]   Learning Ontology-Based User Profiles: A Semantic Approach to
       Personalized Web Search-Ahu Sieg, Bamshad Mobasher, Robin
       Burke-2007

[7]   James C and Charles L., "Personalized Information Environment: An
       Architecture for Customizable Access to Distributed Digital Libraries",
       Library Magazine, v.6, 1999.

[8]   Scafer J.B., Konstan J.A., Riedll J.: "E-Business Recommendation
       Applications" [J]. Data Mining and Knowledge Discovery, vol. 5, no. 1/2,
       115-153, 2001

[9]   Zhao Hongxia and Zhao Yumei, "Web Data Mining Applications in
       E-Business", Financial Computer of Huanan, v.11, 2004.

[10] Tang Chunsheng and Jin Yihui, "Proxy Mechanism Based Internet
       Information Automatic Extraction", Computer Engineering and
       Applications, v.10, 2001: 38-41.

[11] Servastava J., Cooley R., Deshpande M., Tan P.-N..: "Web Usage
       Mining: Discovery and Applications of Usage Patterns from Web Data
       [C]., SIGKDD Explorations,2000.

[12] Rosario Sotomayor, Joe Carthy and   John Dunnion, The Design and
       Implementation of an Intelligent Online Recommender System
       [C],Mixed-Initiative Problem – Solving Assistants, 2005 Fall
       Symposium Series,AAAI, Virginia , USA, 3rd – 6th November 2005,
       AAAI Press.

[13]   Beale Russell Supporting serendipity. Using ambient intelligence to
       augment user exploration for data mining and web browsing [J].
       International Journal of Human sComputer Studies, 2007,
       65(5):421-433.

[14]   Kwork KurLan, Grunfeld Laszlo, Deng Peter ,Employing web minging
       and data fusion to improve weak and hoc retrieval[J].Information
       Processing and Management, 2007, 43(2):406-419

[15] Brian Pinkerton, Web Crawler Finding What People Want [D].University
       of Washington, 2000.

[16]   A Large-scale Evaluation and Analysis of Personalized
        Search Strategies and analysis