

Analysis and Design of Web Personalization Systems for E-Commerce

P. S. Lokhande¹, B. B. Meshram²

¹Asst. Professor, Computer Engineering, AIKTC, University of Mumbai, Navi Mumbai, India

²Professor, Computer Engineering, VJTI, Matunga, Mumbai, India

Email: pslokhande@gmail.com, bbmeshram@vjti.org

Abstract- Web personalization facilitate customizing a web site as per users preferences by using the various data mining techniques, taking advantage of the knowledge acquired from the analysis of the user's navigational usage data. The web personalization reduces the searching time of the user and serves the purchasing interest of the user. In this paper, we have proposed the Analysis and design of the E-Commerce Systems. On the basis of the proposed Software specification Requirements ,for analysis of the systems, we proposed the class diagram for Preprocessing Systems, Knowledge Discovery Systems and Recommendation Subsystems and then provided workflow of the systems by proposing the required functional model of the systems. In the design of the systems we proposed the data structure design and Dynamic Web Systems architecture and implemented the systems. Lastly we have shown the results of the navigational behavior of the user during particular transaction.

Keywords: E-commerce, web personalization, web mining, web navigation, User experience, Data Mining, Log mining.

I. INTRODUCTION

The World Wide Web (www) is a popular and interactive medium to disseminate information today. The ability to track user's web surfing behavior down to individual mouse clicks has brought the e-commerce and customer closer than ever before. It is now possible for e-commerce vendor to personalize the product message for individual customers at a big scale which is referred as Web Page Personalization. There are many techniques to serve the purpose of the personalization methodology. To engage visitors to a web site at a very early stage (i.e. before registration or authentication with login credentials) personalization tools must rely primarily on click stream data captured in the form of web server logs. The problem with the www is the user's behavior is dynamic in nature, which changes depending on the user's interest and many constraints. The size of the data to be handled is huge in capacity. To deal with such problem the data mining techniques is one solution, which is applied on the WWW database and called as web mining.

1.1 Need for Data mining

The web contains the rich and dynamic collection of hyperlink information and web page access and usage information,

providing the rich sources of data mining .So one of the popular applications of data mining is World Wide Web Mining. The web poses great challenges for effective recourse and knowledge discovery. The challenges are due the characteristics of the web sites which are mentioned below:-

We can structure web application development into four different dimensions

- A) Application related characteristics (product itself)
- B) Usage related characteristics (use of product)
- C) Development related characteristics (development of product)
- D) Evolution related characteristics (evolution of product)

1.2 Web personalization

Web site personalization can be defined as the process of customizing the content and structure of a Web site to the specific and individual needs of each user taking advantage of the user's navigational behavior.

Principal elements of Web personalization include

- (a) The categorization and pre-processing of Web data
- (b) The extraction of correlations between and across different kinds of such data and
- (c)The determination of the actions that should be recommended by such a personalization system

The ways that are employed in order to analyze the collected data include content-based filtering, collaborative filtering [1], rule-based filtering and Web usage mining. The site is personalized through the highlighting of existing hyperlinks, the dynamic insertion of new hyperlinks that seem to be of interest for the current user.

II. PREVIOUS WORK

With web page personalization, content to be displayed to a potential customer are chosen based on specific knowledge related to that customer. The goal here is to attract a current customer to purchase something he or she may not have thought about purchasing. Personalization is almost the opposite of targeting. Personalization include such techniques as use of cookies, use of databases and more complex data mining and machine learning strategies. Web usage mining focuses on techniques that could predict user behavior while the user interacts with the web. One of the most successful and widely

used technologies for building the personalization system is record collaborative filtering (CF). Given a target user's record of activity, CF based technique, such as k-nearest neighbor (KNN) approach compare the record with the historical record of the other users in order to find top k neighbor who have similar test or interest. But these techniques suffer through some of its basic limitation, which reduces the efficiency, scalability etc. Using data mining technique is one of the solutions to it.

The goal of the personalization system is to provide the user most likely links of the web page depending upon user's current navigated links. The system will use the data mining techniques like association rule and sequence pattern matching on the web data and transaction data to find out the likely accessed pattern. Existing systems used by many companies, as well as approaches based on collaborative filtering rely heavily on getting human input e.g. user profile for determining the personalization action.

Recently number of approaches has been developed with specific aspects of WUM for the purpose of automatically discovering the user profiles. For example Perkowiz and Etzioni [2] proposed the idea of optimizing the structure of Web site based co-occurrence patterns of pages within usage data for the site. Schechter et al [3] have developed the techniques for using the path profiles of users to predict future HTTP request, which can be used for network and proxy caching. Bucher and Mulvenna [4] have applied data mining techniques to extract usage patterns from web logs, for the purpose of deriving the market intelligence. Yan et.al [5] have proposed clustering of user sessions to predict user behavior. Bamshad Mobsaher, Robert Cooley, and Jaideep srivastatva [6] has proposed the application of association rule and clustering on the transaction data related to the web site.

III. PROPOSED E-COMMERCE SYSTEMS

The growth of e-commerce is happening drastically due to its ease of online payments. The ability to track user's browsing behavior down to individual mouse clicks has brought the vendor and end customer closer than ever before. Due to Web Page Personalization, it is now possible for vendor to personalize the product message for individual customers at a massive scale.

There are various techniques to serve the purpose of the personalization task. To engage visitors to a web site at a very early stage (i.e. before registration or authentication) personalization tools must rely primarily on click stream data captured in web server logs [1]. The lack of explicit user ratings as well as the sparse nature and the large volume of data lead serious challenges. The problem with the www is the user's dynamic behavior, which changes depending on the user's interest and many constraints. The size of the data to be handled is in terms of terabytes. To deal with such problem the data mining techniques is one solution, which is applied on the WWW database.

3.1 Statement of The Problem

Use of data mining technique on the web data to serve the personalization task is one of the best choices to solve the problem. The main task in Web Personalization System is to provide the most likely visiting link in the future for particular user depending on various users current navigated browsing pattern. The job is to design the Web Personalization System for an e-commerce site.

In our project, the website is of the superstore shop which sells everything from food to electronic goods. The sellers want the site to be modified:-

To increase the cross sells.

To provide the different advertising policy to different user.

To provide the ease of surfing the site to the user.

To get the group of user of particular interest.

This will help the seller to get potential customer and new market policy in future. Here the shop is maintaining the data for the purchased items and the customers of the shop. It is also maintaining server log data. The task is to mine both these data and get the association rule from these and provide the recommendation to most likely visiting link on the site.

3.2 Analysis of Dynamic E-Commerce Website

3.2.1 Class Diagram for systems

It describes the structure of objects in the system –their identity, their relationship to other objects, their attributes and their operations. The objects model is represented graphically with the help of object diagram, class diagram.

3.2.1.1 Object Model for Preprocessing System

The preprocessing system , collects and cleans the data. The data is again stored back in the various database. The preprocessing system reads the web usage log files which are stored at the web server and uses string tokenizer class to separate the fields of the text file so that it can be stored in the database. The non required field will be removed from the data and only useful information like IPAddress, session no. , Page visited, Date time hour etc. will be separated and stored in the database. This preprocessing is required for the data mining algorithm to which these tables will be given as input.

The class diagram shown below is for the preprocessing module of the system. The classes like user identification, sitemapper will collect data for preprocessing. Sitemapper will collect the data for visited links of HTML pages. User identification will use the cookies to identify the new user and previous users which will help for the personalization process.

3.2.1.2. Object Model for Data Mining Engine

The diagram shown below is the class diagram for the mining engine. Hastree class uses the Hash tree to speed up the mining process. The class has two nodes that is ListNode and the has node. ListNode is used to maintain the list of items in a transaction. Hasnode is used to maintain the key for the

particular transaction. DBReader and DBWriter are used to provide the read and write function on the file respectively. Apriori rule is used find whether there is association between the rule or the rule satisfy the minimum support criteria.

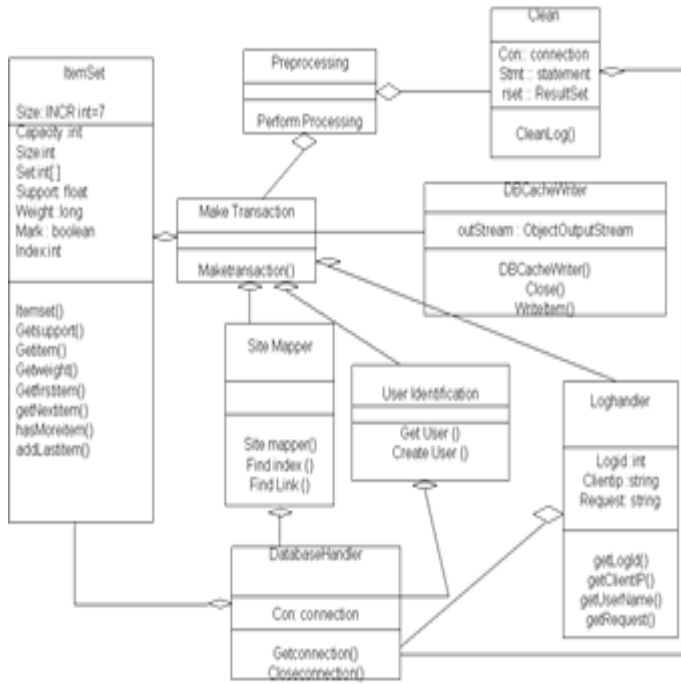


Fig. 1: Class Diagram for Preprocessing System

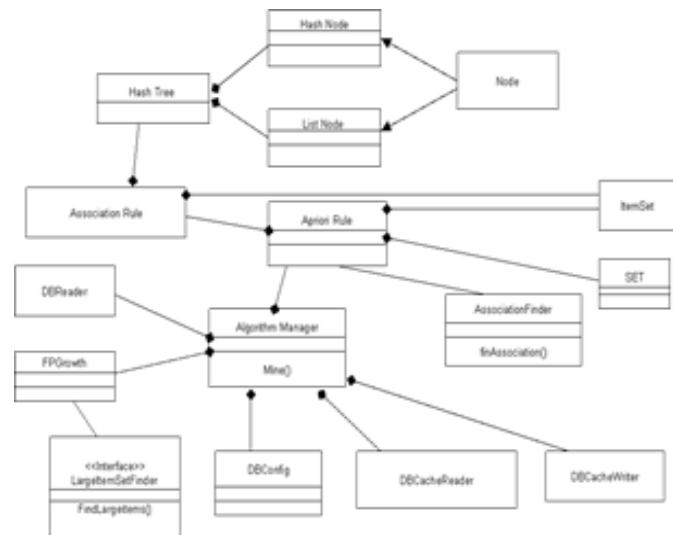


Fig. 2: Class Diagram for Knowledge Discovery module

3.2.1.3 Object Model For Recommendation Module

In the diagram shown for recommendation module ,homepage is main class which starts the application. If login process completes successfully , system will create personal page for user depending on the output of classification algorithm. Also Homepage will check the cookies ,if already present it will

greet the user accordingly. Otherwise new cookie will get created.

The class diagram shows the basic activities and classes involved in the shopping process on the superstore web site. The classes linkreco.java will recommend the links to the user. Login.java will create login session for each user and personalize web site according to interest of each user and output of classification algorithm output.

Figure shows the class diagram for the shopping package in which creation of session, handling client session and destroying session are the important activities. Also order and payment are also important activities.



Fig. 3 Class Diagram for Recommendations Module

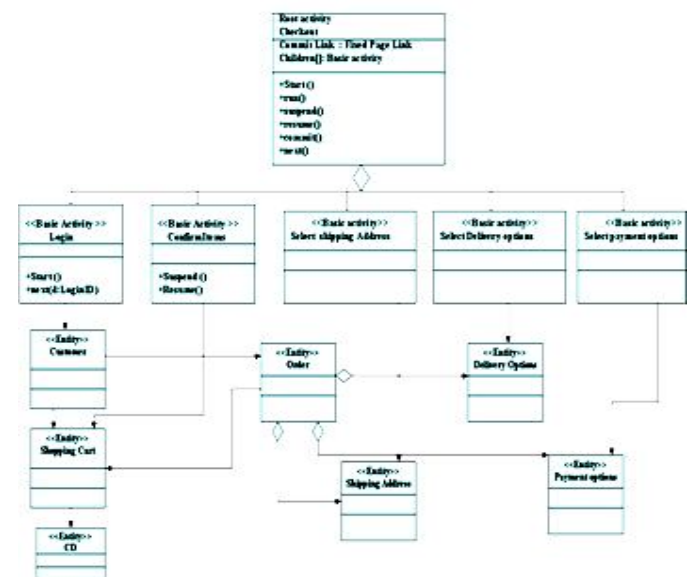


Fig. 4 Class diagram for Shopping package

3.3 Functional Model for web personalization System: Data flow diagrams

3.3.1 Preprocessing System

The initial step in the system will be carrying the preprocessing activity on the data stored at the server log. This preprocessing activity is broadly classified in some small activities like cleaning, identifying the user and user transaction.

Context Level DFD of the System

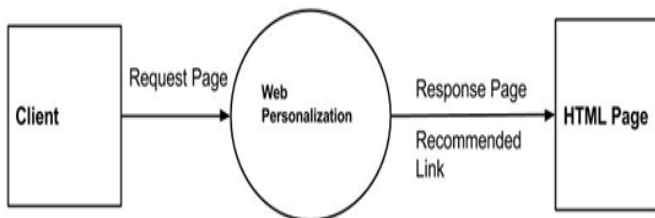


Fig. 5 : Level 0 DFD for the system

Functional Model for the preprocessing system is shown below.

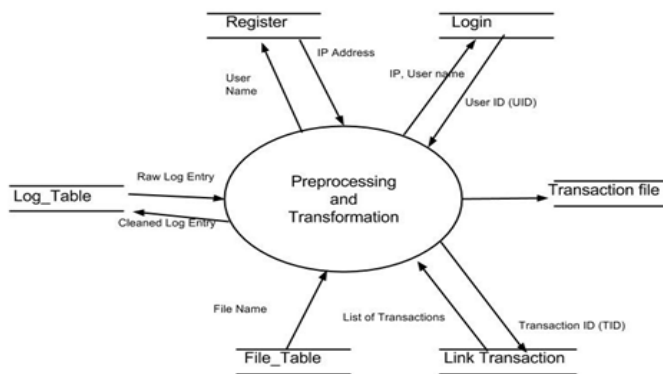


Fig. 6: Context level DFD for the preprocessing module

3.3.2 Data Mining System

The input to the system is from various database files such as transaction records, registration details or web access log files. The result from the mining on these databases is association rule or Mining Log or Classification of registered data.

3.3.3 Data Mining Module

The complete Data mining step is divided in three parts depending upon which algorithms we are going to use for mining. There are three methods implemented Apriori, ID3 and Access log Mining.

3.3.5 Recommendation System

Recommendation System is responsible for handling the recommendation link to the web page. Also the system will decide upon the links based on the outputs from the data

mining algorithms i.e. association rules, classification or sequence pattern.. Before selecting the link to be displayed, module will check which product user is currently viewing or buying and accordingly displays the recommendation link for other products.

3.3.6 Level 1 DFD for Recommendation Module

Level1DFD is as shown in the figure which shows how the data flows between different functions. Here it will first check the login. If it is proper, then the greeting to the user will be based on the classification. After once user start browsing the site, personal page for each user will be created by using the knowledge generated from the association rule and sequence pattern.

3.4 Design of Dynamic E-Commerce Website :

System Design : Design of the system contains system file Details, Component Diagram, Deployment Diagram.

3.4.1 Data Structure Design

For the Web personalization system various databases are maintained which are as follows:

- 1. Login table :** Here login information about the user is stored
 Login (Name, Emailid, pswd, voice)
 The fields of the table are as follows:
 Emailid: this field stores the emailed of the user which should be unique per user.
 Pswd : this field stores the password of the user.
 Voice: this field stores the address of the welcome voice file created for the particular user.
- 2. Advertisements :** Advertisements to be shown to user are stored in this table.
 Advertisement(addid,Title,Pagelink,Image,Descriptipon)
where
 Addid: is the field for the identification of advertisement number.
 Title: This field have values like sport, music, reading etc. to identify their category.
 Pagelink : this field stores the address of the html page to which program should go after clicking the link.
 Image : this field have address of the image file to be displayed to the user for advertisement.
 Description: is the text to be displayed alongwith image for the product advertisements.
- 3. Register :** Here the registration information provided by user is stored. After registration form is completed data will directly go to the REGISER table.
Register (Fname, Lname, Emailid, pswd, Sport, Shopping, Sport, Music, Reading, Shopping, Gender, Age, Profession)
 These are all field which stores the information collected from the registration form.

FNAME: first name of the user is stored in this field.

LNAME: Last name of the user is stored in this field.

EMAILID : Email id of the user is stored in this field.

PSWD : password entered by the user is stored in this field.

SPORT : if user selects sport as his interest in the form, 'Sport' value will be entered in the table.

MUSIC: if user selects music as his interest in the form, 'Music' value will be entered in the table.

READING: if user selects Reading as his interest in the form, 'Reading' value will be entered in the table.

SHOPPING: if user selects shopping as his interest in the form, 'shopping' value will be entered in the table.

GENDER : Gender value entered by the user will be stored in the field.

AGE: Age value selected from the drop down box from the registration form will be stored.

PROFESSION: profession value selected from the registration form will be stored.

4. **Product :** In this table product details are stored

The field description is as below:

Product(Productid, SD, LD, Cost, Productname, Productpage)

PRODUCTID: each product will have unique id & will be stored in this field.

SD: Product's short description or title is stored in this table.

LD: Product's detailed or Long description is stored in this table.

COST: cost of the product is stored in this table.

PRODUCTNAME: name of the product is stored in this field.

PRODUCTPAGE: If any separate html link is there for the particular product, then that address is stored in this field.

5. **Classification:** In this table output from ID3 algorithm is stored. The registered users will be classified.

Classification (Emailid, Fname, Class,)

The classification table have following fields.

EMAILID : Email id of the user is stored in this field.

FNAME: FNAME: first name of the user is stored in this field.

CLASS: The class of the user found from the classification algorithm i.e. ID3 is stored in this field. The class will help to personalize the web page of user faster.

6. **Access Log:** Log data mined from the web server is stored in this table.

Log data after preprocessing is stored in this table.

Accesslog (IPaddress, Datetime, GMT, Method, Pagelink, Protocol, Port, ID)

IPADDRESS: Ipaddress of the user who visited the web site is stored in this field.

DATETIME: date and time when user accessed the site are stored here.

GMT: GMT value from the log for particular record is stored in this field.

METHOD: Method of accessing particular page link GET or POST is stored here.

PAGELINK: The accessed page by the user is stored here.

PROTOCOL: protocol used for accessing the web page of the site is stored here

PORT: port number used to access the web page is stored.

ID : Identification number of the user session is store.

7. **Mining Log :** Here the second iteration output from the access log is stored.

The fields of the table are as follows

Mininglog (IPaddress, Date, Hour, MIN, Seconds, Pagelink)

IPADDRESS: Ipaddress of the user who visited the web site is stored in this field.

DATE: date when user accessed the web page is stored here.

HOUR: At what hour user accessed the web page is stored here.

MIN: minute value when user accessed the web page is stored here.

SECONDS : second value when user accessed the web page is stored here.

PAGELINK: The accessed page by the user is stored here.

8. **Transaction:** Transaction performed on the web site will be stored in the Transaction table.

Transaction (TID, Television, Refridgerator, Computer, Laptop, Livingroomset, Diningtable, Computer Table, Bed)

Transactions on the Web site are stored in this table. The descriptions of the fields are as follows.

TID: transaction identification number is stored here.

TELEVISION: If TV is purchased during shopping this field have value "1", otherwise "0".

REFRIDGERATOR: If fridge is purchased during shopping this field have value "1", otherwise "0".

COMPUTER: If computer is purchased during shopping this field have value "1", otherwise "0".

LAPTOP: If laptop is purchased during shopping this field have value "1", otherwise "0".

LIVINGROOMSET: If living room set is purchased during shopping this field have value "1", otherwise "0".

DININGTABLE: If dining table is purchased during shopping this field have value "1", otherwise "0".

COMPUTER TABLE: If computer table is purchased during shopping this field have value "1", otherwise "0".

BED: If bed is purchased during shopping this field have value "1", otherwise "0".

9. **TRP:** Output from the apriori algorithm with support value is stored in the table. The field's are storing the association rules and support value will be stored in the support field. The item1 to Item8 fields get filled as we find out association rule's between them . Item1 field is for TV, Item2 for Fridge and so on.

10. **PageLinkID:** This table is used to store the pagelinks and their path.

Pagelink (Pagelink, ID)

This table is helpful for giving the identification to the pages for sequence pattern matching.

PAGELINK: This stores the value of the page address which is there in access log.

ID: The ID which we want to allot to the page.

3.5 Proposed Software Architecture.

A component diagram models the pieces of software, embedded controllers etc. that will make up a system. A component diagram has a higher level of abstraction than a class diagram—usually a component is implemented by one or more classes. Component can encompass a large portion of a system. The component diagram for the proposed system is as shown below fig. no. 3.15. Homepage.java is the main program which creates the default user interface to the web site. The Data mining module is the separate module which is run by the administrator to update the information for personalization like new association rules, classes of new user’s and sequence pattern’s. InitApriori.java, ID3mod.java and Updatelog.java are the programs for mining the data from the database Transaction, Register and AccessLog respectively. The output of the mining algorithm will be stored in the TRP(transaction processing output), Classification and Mininglog database tables.

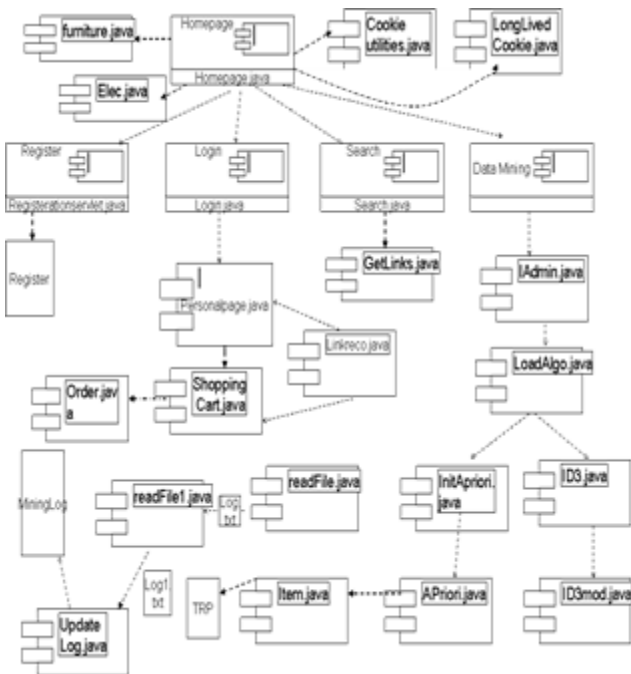


Fig. 7: Component diagram for web personalization system.

IV.RESULT

We have implemented the system by using analysis and design of Web Personalization , however due to space constraint the most important screenshots are given below, which deals with the transactions of the user and the recommendation link to the user for doing the next transaction of the interest of the user. The Few screen shots are as shown below.

This is homepage or entry point to the project web personalization. The web site is for online shopping of a superstore. The top menu shows the navigational schema. The middle of the page shows the current offers on the site.

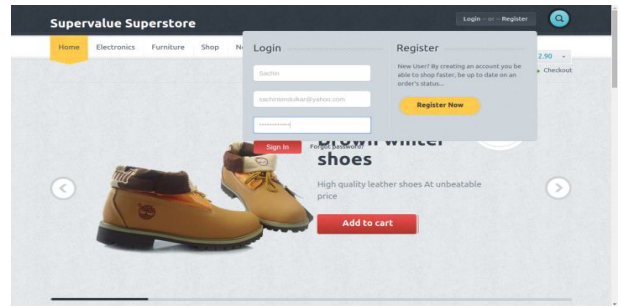


Fig. 8: Superstore’s Homepage

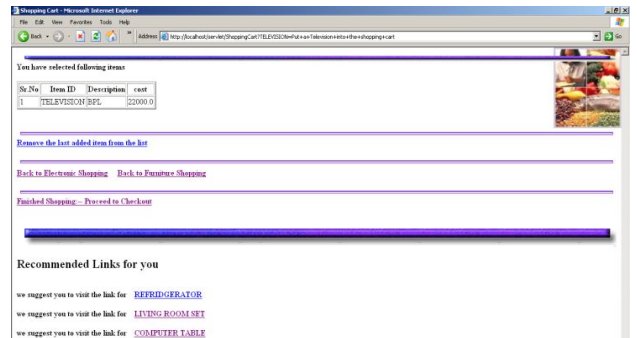


Fig. 9 Shopping Web Page

Figure shows the web page which will be generated once an item is added to the shopping cart. The recommendation link’s can be shown on the page based on the association rules. Association rules tell us that the product currently purchased and the recommended are associated. So the recommendation links will always differ depending on which product has been currently selected.

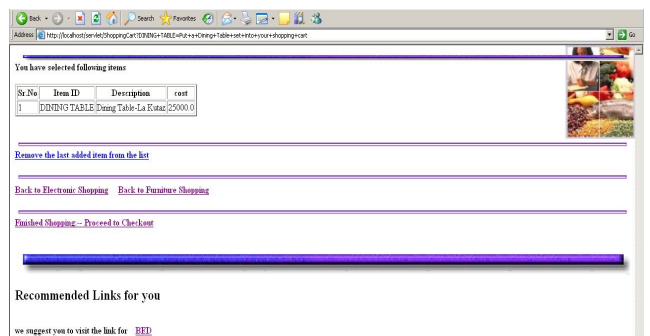


Fig. 10 Shopping Web Page

Here you can see as Dining table is selected, the recommendation links for the selection will be only Bed which is derived from the output of the Apriori association rule mining algorithm.

After product selection for shopping of user is completed, he will click on the checkout.html. Then his final selection of shopping, total price will be shown to user. The system will ask user to enter his/her credit card system

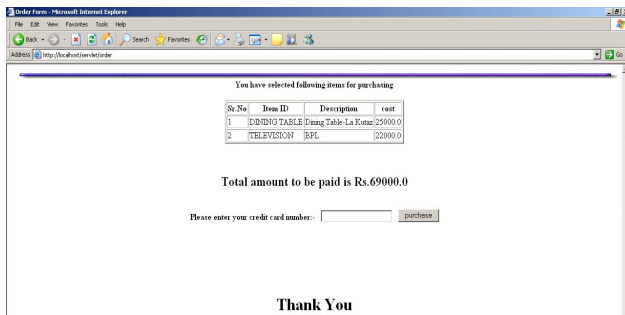


Fig. 11: Order Web Page

V. CONCLUSION

Web personalization technique is growing in E-commerce industry due to its feature to keep customer attached with online E-commerce store. We implemented Preprocessing Systems, Knowledge Discovery Systems and Recommendation Subsystems. In the proposed web personalization the process of customizing and displaying the content to Web site visitor according to individual needs proved to be a successful mantra. Showing personalized content can be achieved by taking advantage of the user's navigational behavior, as it can be revealed through the processing of the Web usage logs, as well as the user's characteristics and interests. The main element of a Web personalization system is the usage miner. Log analysis and Web usage mining is the procedure where the information stored in the Web server logs is processed by applying Knowledge Discovery and data mining techniques, such as clustering, association rules discovery, classification and sequential pattern discovery, in order to reveal useful patterns that can be further analyzed. There are various Data mining algorithms available but we have proposed to use ID 3 Algorithm in this project to get the personalized recommendations. Personalized recommendation without breaching the privacy of user is given to the user. This paper shows design work for building the web personalization system for e-commerce.

REFERENCES

- [1] Zitao Liu, Gyanit Singh, Nish Parikh, Neel Sundaresan, "A Large Scale Query Logs Analysis for Assessing Personalization Opportunities in E-commerce Sites", WSCD '2014 New York, New York USA, ACM-2014
- [2] Perkowitz, Mike, and Oren Etzioni. "Towards adaptive web sites: Conceptual framework and case study." *Artificial intelligence* 118.1 (2000): 245-275.
- [3] Schechter, Stuart, Murali Krishnan, and Michael D. Smith. "Using path profiles to predict HTTP requests." *Computer Networks and ISDN Systems* 30.1 (1998): 457-467.
- [4] Büchner, A. G., & Mulvenna, M. D. (1998). Discovering internet marketing intelligence through online analytical web usage mining. *ACM Sigmod Record*, 27(4), 54-61.
- [5] Yan, T., Jacobsen, M., Garcia-Molina, H., Dayal, U., From user access patterns to dynamic hypertext linking. In *Proceedings of the 5th International WWW Conference*, Paris, 1996
- [6] Mobasher, Bamshad, Robert Cooley, and Jaideep Srivastava. "Automatic personalization based on web usage mining." *Communications of the ACM* 43.8 (2000): 142-151.

WEB REFERENCES

- [YAH] Yahoo! <http://www.yahoo.com>
- [APP] Apple Computer <http://www.apple.com>
- [DEL] Dell <http://www.dell.com>
- [EXC] Excite <http://www.excite.com>
- [AMA] amazon.com <http://www.amazon.com>
- [MSN] Microsoft Network <http://www.msn.com>
- [EBA] Ebay www.ebay.in
- [SNA] Snapdeal www.snapdeal.com