

MODELING AND TESTING OF A NOVEL WEB BASED RECOMMENDATION SYSTEM

Sowmya K.Menon¹, Varghese Paul², M.Sudheep Elayidom³, Yashwant Sinha⁴

¹Assistant Professor, Department of Computer Applications, SNGIST N. Paravur, Kerala, India

²Varghese Paul Professor, Department of IT, CUSAT, Kochi, Kerala, India

³Associate Professor, Computer Science, CUSAT, Kochi, Kerala, India

⁴CUSAT, Kochi, Kerala, India Author

Abstract

In this era of social networking and its high demand, there is indeed a real need of an efficient system that recommends good websites of interest to each user. Here we are analyzing the user's browsing behavior habits. After making clusters by the system it will recommend the users the other web sites belonging to same cluster behavior. Here the browsing habits of a group of users are analyzed and the users are grouped to different clusters in the server such that the users in the same cluster have similar browsing habits to visit similar web sites. The core idea is that the system will recommend the frequent web sites visited by other users in the same cluster to a typical user of a cluster. The entire system is implemented over a cloud framework of many users and a cloud server based on the Google App Engine. In this paper, we focus more on the testing the effectiveness of the proposed recommendation system. To get the accuracy we are comparing the browsing behavior of the users in the database and the sites which is going to be recommended.

Keywords: Clustering, Google App Engine, Web Recommendation System.

-----***-----

1. INTRODUCTION

Web data mining is the application of data mining techniques to extract knowledge from web data. Web content mining is the process of extracting useful information from the contents of web documents. Web usage mining is the application of data mining techniques to discover interesting usage patterns from web data, in order to understand and better serve the needs of web based applications. Web Recommendation system is a system for predicting the user's browsing behavior. This recommendation system is an interactive software. Recommendation systems are a class of information filtering system that helps to rate the users preference to an item. Recommendation systems are nowadays very common and are used in many applications. Recommendation systems [4] are more effective and applicable to real life applications. The most popular applications of web recommendation systems are movies, music, news, websites etc.

Recommendation system produces a list of recommendations either through collaborative filtering or content based filtering. Collaborative filtering [1] builds a model from users past behaviour. Content based filtering uses a series of discrete characteristics of an item in order to recommend additional items with similar properties. The roots of web recommendation systems can be applied to information retrieval [4] also. This paper discusses the various testing phases of this web recommendation system in detail.

2. OBJECTIVES

The main objective is to model, implement and test a web based recommendation to predict interesting websites that suits a user's browsing interests. Here we collect the user's browsing patterns in the server and cluster it using the K-Means algorithm. After this process, we recommend the users those websites that are browsed by users in the same cluster.

3. PROPOSED ARCHITECTURE

The architecture used in this system is mainly the client server architecture. Users of same browsing behaviour are clustered and stored in the server. The clustering of similar users and the recommendation systems will take place in the server. Extensions are small software programs that can modify and enhance the functionality of the chrome browser. Written using web technologies such as HTML, JavaScript and CSS.

Google App Engine is a platform for developing and hosting web applications in Google-managed data centers. Google uses as a data storage a facility called big table. Big table is a distributed, persistent, multidimensional sorted map. big table is not a relational database. Google App Engine lets you run web applications on Google's infrastructure. App Engine applications are easy to build, easy to maintain, and easy to scale as your traffic and data storage needs grow. With App Engine, there are no servers to maintain: You just upload your application, and it's ready to serve your users.

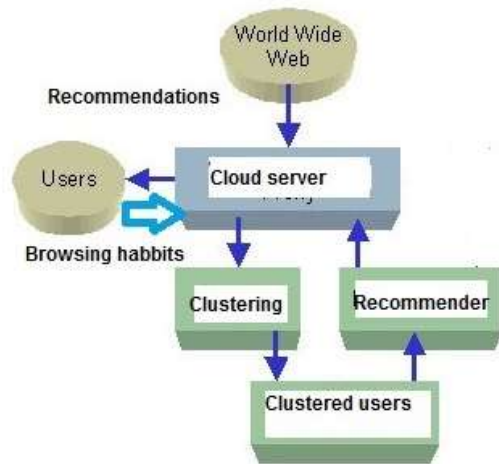


Fig -1: Proposed architecture

In the above figure, the proposed architecture for the recommendation system is shown. As the users browse websites, the client part of the software (installed as a plug-in in the user’s browser), collects the browsing information and communicates to the server on a periodic bases. The server is actually a cloud server based on the Google app engine architecture which does the actual clustering based on the k-means algorithm. The recommended web sites used by clustered users are returned by this recommendation system to the client plug-in as shown in the above architecture diagram. These recommendations are given to the user for further browsing. Google App Engine lets you run web applications on Google's infrastructure. App Engine applications are easy to build, easy to maintain, and easy to scale as your traffic and data storage needs grow. With App Engine, there are no servers to maintain: You just upload your application, and it's ready to serve your users.

This type of algorithm is quite different from hierarchical clustering because it is told in advance how many distinct clusters are being generated. The algorithm will determine the size of the clusters based on the structure of the data. K-means clustering assumes an initial value of *k* with randomly placed *centroids* (points in space that represent the center of the cluster), and assigns every item to the nearest one. After the assignment, the centroids are adjusted to the average location of all the points assigned to them, and the earlier assignments are redone. This process repeats until the assignments does not change further. Figure 2 shows a basic working of the K-Means algorithm.

On implementation, this system can be tested among a group of say, 100 users, by observing their browsing habits for a period of say, 1 month. They may install the plug-in from the web site and their browsing habits may be monitored for a period of two weeks. It should be proved that the users in the same cluster as returned by the software, are visiting the same web sites in next two weeks as suggested by the recommendation system.

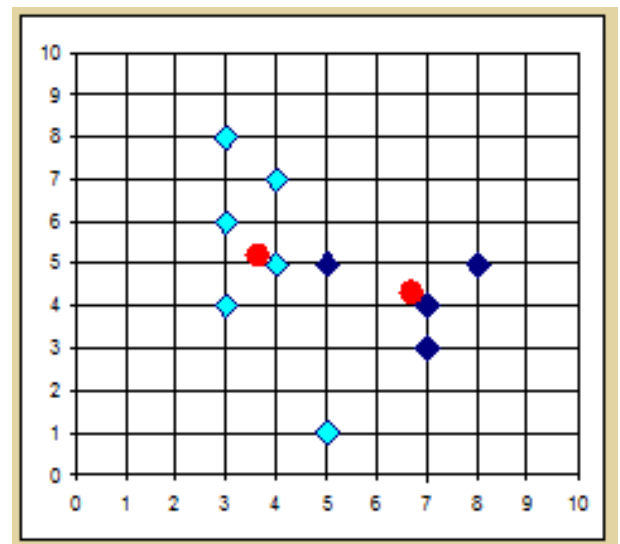
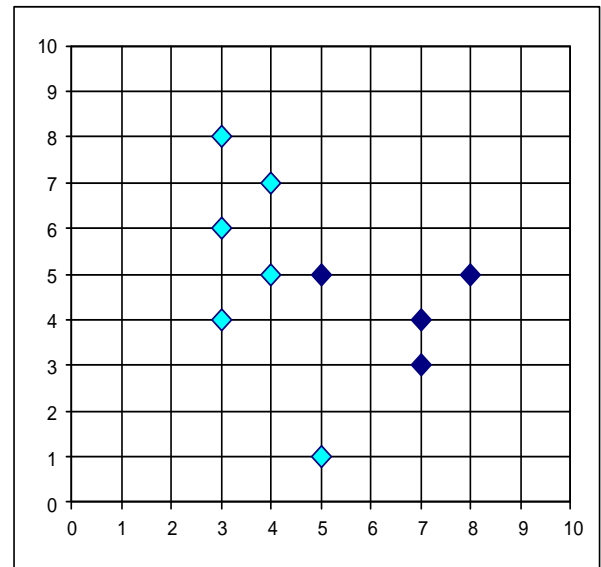


Fig-2: K-Means algorithm

4. TESTING OF WEB RECOMMENDATION SYSTEM

This configuration is made to install online and ask the user to install the chrome extension which is available online on chrome webstore. Anyone can download that plug in and install it. In order to use that plug in the user has to register through his or her email-id.

The chrome extension record the browsing behavior of each user and store that value in the database and the server side. We make the cluster using the k-means clustering algorithm on the basis of the browsing behavior. After making clustering by the system, it recommend the user their browsing behavior belonging to the same cluster.

We store 100 sites of different categories in the database for making different clusters which are listed in the category.pdf.

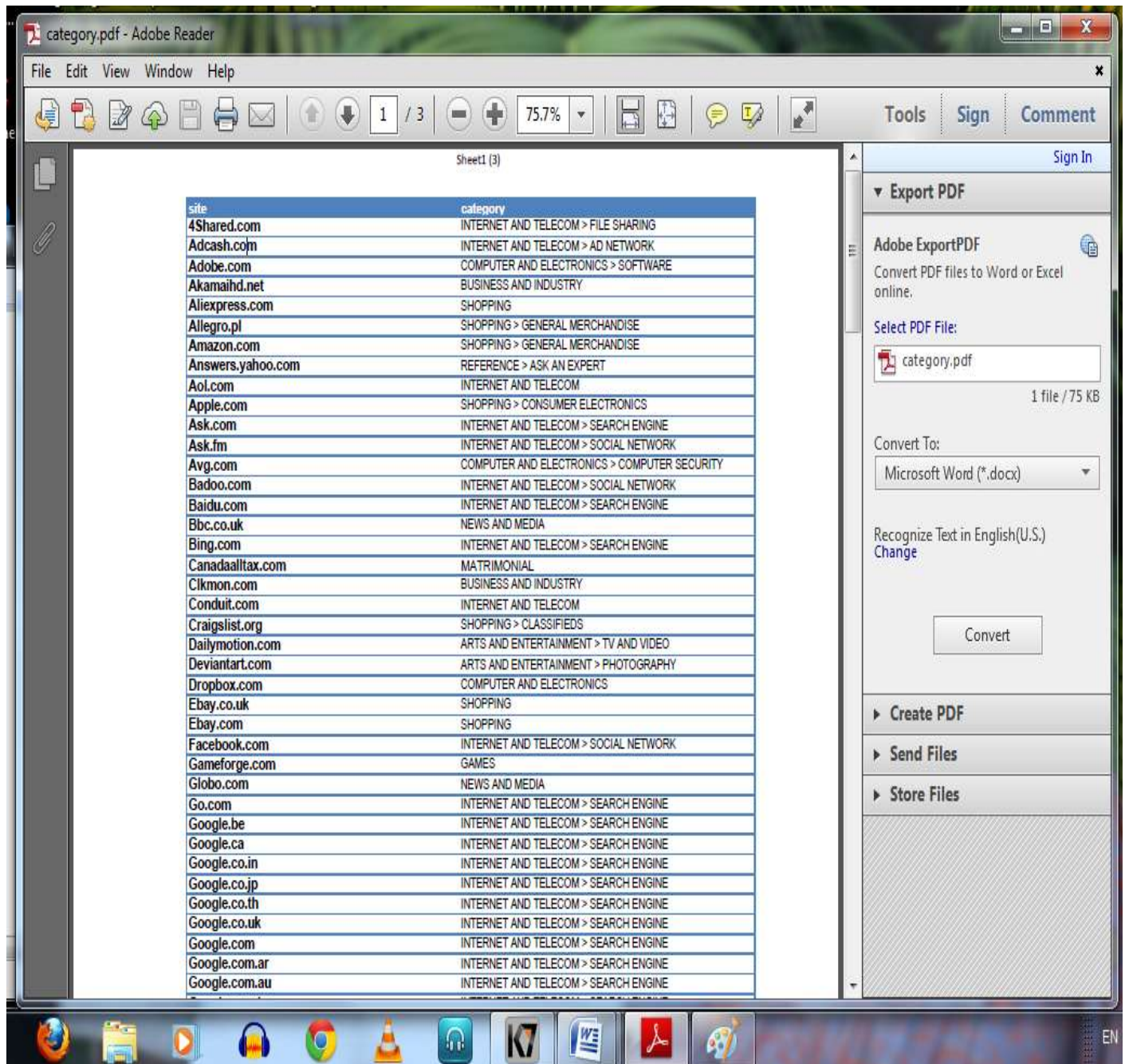


Fig-3: category.pdf

In recommendations, the utility of a website depends on the ratings. The number of times a particular website is being visited by the user is being recommended by the user.

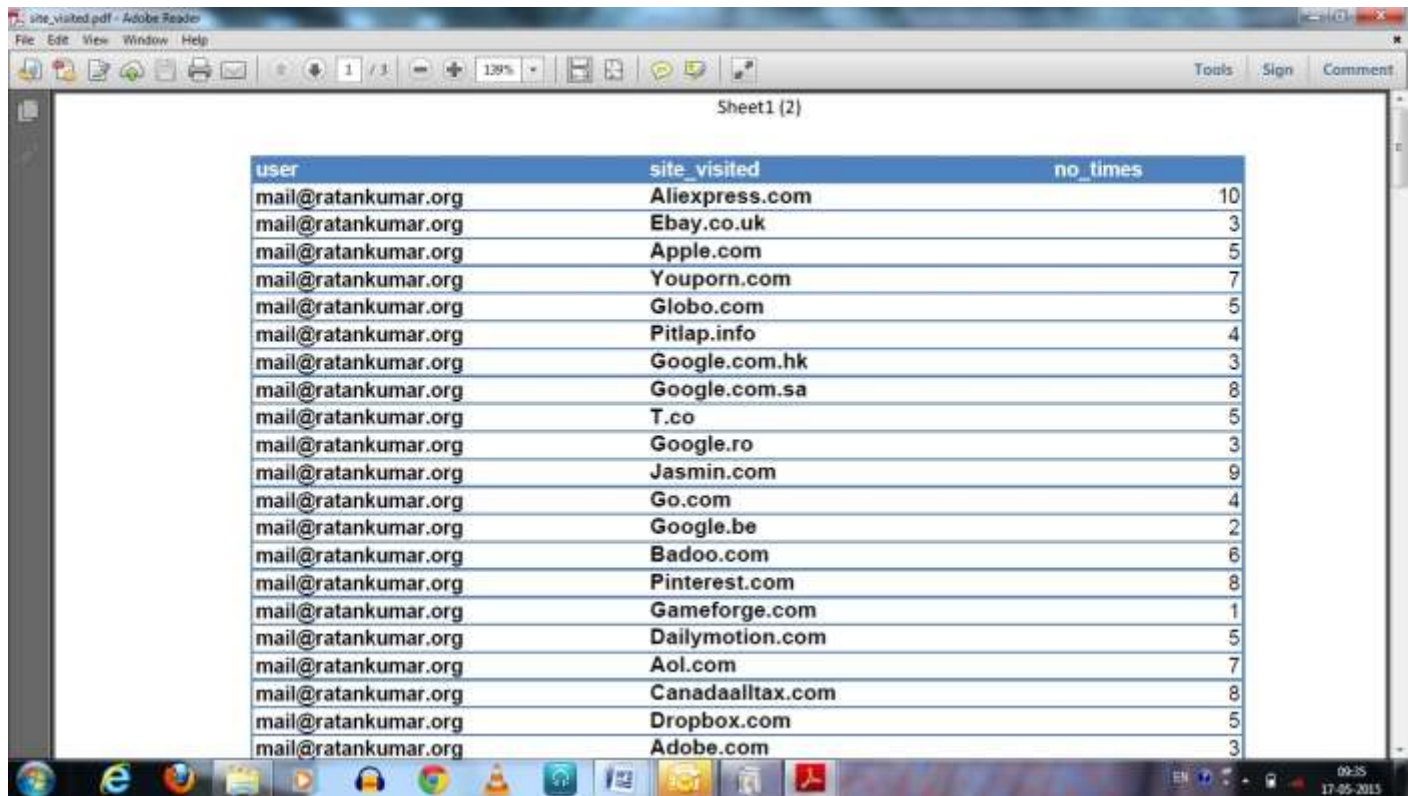
We conduct the test in three phases: -

In first phase we record the browsing behavior of 100 users for 15 days. Each user is supposed to have visited 20-21 sites of different categories which are listed in site_visited.pdf.

In the second phase on the server side of the system cluster the browsing behavior. Now we are using only two clusters.

As the number of visitors increases, we can increase the value of 'k' accordingly.

In the third phase for testing, we compare the websites that are actually visited by users during the next 15 days, against the sites that are recommended by our system. We could infer that the users were browsing the sites, actually recommended by the system itself with an accuracy of 85%.



user	site_visited	no_times
mail@ratankumar.org	Aliexpress.com	10
mail@ratankumar.org	Ebay.co.uk	3
mail@ratankumar.org	Apple.com	5
mail@ratankumar.org	Youporn.com	7
mail@ratankumar.org	Globo.com	5
mail@ratankumar.org	Pitlap.info	4
mail@ratankumar.org	Google.com.hk	3
mail@ratankumar.org	Google.com.sa	8
mail@ratankumar.org	T.co	5
mail@ratankumar.org	Google.ro	3
mail@ratankumar.org	Jasmin.com	9
mail@ratankumar.org	Go.com	4
mail@ratankumar.org	Google.be	2
mail@ratankumar.org	Badoo.com	6
mail@ratankumar.org	Pinterest.com	8
mail@ratankumar.org	Gameforge.com	1
mail@ratankumar.org	Dailymotion.com	5
mail@ratankumar.org	Aol.com	7
mail@ratankumar.org	Canadaalltax.com	8
mail@ratankumar.org	Dropbox.com	5
mail@ratankumar.org	Adobe.com	3

Fig-4: site_visited.pdf

FUTURE SCOPE

Web Sites like Amazon.com, Pandora radio, Netflix uses the concepts of recommendation systems. It has been in recent use, to help users to get recommended according to their likes. Analytics have successfully proliferated into applications to support customer recommendations, customer value and churn management, campaign optimization, and fraud detection.

This work has a future scope in finding out the browsing behavior of a specific group of users such as customers of electronic products, books, dress materials etc and giving recommendations according to it. The recommendation is based on those users who are having the same browsing behavior. Thus a user can get some websites of their own interest. Recommendation systems[12] can be applicable in real time examples such as movies, ratings of movies etc.

The clustering algorithm that is being in this work is the K-Means clustering algorithm. Since the field of data mining algorithms is emerging with tremendous pace, some new generation fast clustering algorithms may be developed and tested for the future clustering purpose. Also instead of the Google App Engine/ Big table frame work in this work, Hadoop /Map reduce frame works may also be used and tested as a future work.

One of fastest growing areas of research in the area of recommender systems is mobile recommender systems. The market of mobile phones is increasing day by day. With the increase of internet-accessing smart phones, it is now possible to offer personalized, context-sensitive

recommendations [2, 4]. This is particularly a difficult area of research, as mobile data is more complex than conventional recommender systems, since it is heterogeneous, noisy, requiring spatial and temporal auto-correlation. One example of a mobile recommender system is one that offers potentially profitable driving routes for taxi drivers in a busy city with frequent traffic jams. Another example of the application of recommendation system is movies. Movies are rated and applied the recommendation systems in it. The recommendation system facilitates easy driving routes for taxi drivers. The system uses machine learning techniques and reasoning process in order to adapt dynamically the mobile recommender system to the evolution of the user's interest.

ACKNOWLEDGEMENT

We would like to thank the students of Cochin University namely, Mr. Yashwant sinha, Mr.Ratan Kumar who have contributed in the implementation and documentation part of this work.

CONCLUSION

In this work, testing the effectiveness of a web recommendation system is proposed. Testing is done in three phases. In first phase we record the browsing behavior of 100 users for 15 days. In second phase, on the server side of the system we cluster the browsing behavior of users having similar tastes. The value of 'k' depends upon the number of users. In the third phase we are actually comparing the websites that are being visited by users during the next 15 days. We came to the conclusion that

the users were browsing the sites actually recommended by the system with an accuracy of 85%. This type of approaches also helps to build the users confidence in recommendations. This type of recommendations shows certain websites that are more prone to be used by users. Nowadays web usage mining has become an essential tool for getting user friendly web services. Web usage mining is mainly used by e-commerce sites to increase their profits. This web mining is used in search engines to improve their quality of searching. In this paper we are testing web recommendation systems rather than e-commerce applications.

REFERENCES

- [1]. Jia Li and Osmar R. Ziane, "Combining Usage, Content, and Structure Data to Improve Web Site Recommendation", *In Proceedings of the 5th International Conference, EC-Web*, Spain, 2004, pp 305-315. Reference 1
- [2]. Harita Mehta, Shveta Kundra Bhatia, Punam Bedi and V. S. Dixit, (November 2011). "Collaborative Personalized Web Recommender System using Entropy based Similarity Measure", *IJCSI International Journal of Computer Science Issues*, 8(6), pp. 231-240.
- [3]. Toby Segaran, *Programming Collective Intelligence*, 1st ed, O'Reilly, 2007.
- [4]. G. Adomavicius and A. Tuzhilin, (2005), "Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions", *IEEE Trans. on Data and Knowledge Engineering*, 17(6), pp. 734-749.
- [5]. I.J Witten, E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques," Elsevier, 2011.
- [6]. F.Y. Tani, D.M. Farid and M.Z. Rahman (January 2012). Ensemble of Decision Tree Classifiers for Mining Web Data Streams. *International Journal of Applied Information Systems* 1(2), pp. 30-36.
- [7]. Ungar L.H. and Foster D.P, "Clustering Methods for Collaborative Filtering", In the Proceedings of the Workshop on Recommender Systems, AAAI Press, 1998.
- [8]. Badrul M. Sarwar, George Karypis, Joseph A. Konstan, and John Riedl, "Analysis of recommendation algorithms for e-commerce", In *Proc of ACM Conference on Electronic Commerce*, Minneapolis, MN, USA, 2000, pp. 158-167.
- [9]. Upendra Shardanand and Patti Maes, "Social information filtering: Algorithms for automating "word of mouth"", In *Proc of ACM CHI'95 Conference on Human Factors in Computing Systems*, Denver, Colorado, USA, 1995, pp. 210-217.
- [10]. W. Wong and A. Fu. Incremental document clustering for web page classification, 2000. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.21.2379&rep=rep1&type=pdf>
- [11]. Arbee L.P, Chen Yi-Hung Wu, Yong-Chuan Chen, "Enabling personalized recommendation on the web based on user interests and behaviours", In *proc. of 11th International Workshop on research Issues in Data Engineering*, Heidelberg, 2001, pp.17-24.
- [12]. Januz Sobocki, "Implementations of web based recommender systems using hybrid methods", In *International journal of computer science and applications*, Vol.3 issue 3, pp 52-64.
- [13]. C. Lin, S. Alvarez, and C. Ruiz, "Collaborative recommendation via adaptive association rule mining", 2000.
- [14]. Xiaobin Fu, Jay Budzik, and Kristian J. Hammond. Mining navigation history for recommendation. In *Intelligent User Interfaces*, pages 106-112, 2000.
- [15]. Wong and A. Fu. Incremental document clustering for web page classification, 2000.