



Analysis of user behavior to find interest priorities in big data log of web proxies

Ali Reza Honarvar¹, Ali Saem²

¹Department of Computer Engineering and Information Technology, Islamic Azad University, Safashahr Branch, Safashahr, Iran

²Department of Computer Engineering and Information Technology, Islamic Azad University, Safashahr Branch, Safashahr, Iran

Abstract: *The today's life is unimaginable without internet due to internet penetration level. Entertainment, communication, education, business, personal relationships and private rights are affected by this technology and a new scene has been emerged for people and firms. Internet users and their traces are valuable for companies and this indicate that when an advertisement is shown, their reference is information obtained from cookies; accordingly, this information is shown based on the users' interests. Statistical data obtained from virtual communities, personal interests, user groups and virtual character of users are valuable as their physical nature in real communities, because companies' income is related to correct identification of users and low error percentage of advertisement targeting. Also, we need some system that recommends the most suitable products and services due to increasing growth of internet and huge volume of information. Such systems are called recommender systems. These systems lead to more purchase and increased customer satisfaction with online shopping recommending best products and services. Majority of firms uses any method to collect user search as well as personal information of users. Some companies such as Google can also read the content of encrypted email line by line in order to identify interests and needs of users. In this study, we review recommender system and then introduce different type of recommender systems and evaluate them and at the end, we express the design of a recommender system using content analysis of user behavior in over internet in order to find interest priorities in big data log of web proxies.*

Keywords: *E-commerce, recommender systems, data mining, web mining*

I. Introduction

Web is developed during a chaotic and decentralized process and this trend leads top generation of an extensive volume of documents connected to each other with no logical organization and order. The boom increase in web use besides its capacities and capabilities has made it necessary to storage big data of million users and visitors. In fact, web is changed to a large set of structured and semi-structured data and web users insure loss due to overlapping data. Therefore, analysis of search behavior of web users and interests of users is an important factor. Examining behaviors of web users, a method to discover hidden knowledge in interaction between users and web, is one of important tools in scope of web search.

Web mining is associated with data mining technics in large storages if input data into web system [1]. This term was introduced by Etzioni in 1996 [2]. General process of web mining is divided into three different but linked groups by researchers based on input data used by them. These groups include web search structure, web content mining and web mining application. Cooley (1971) introduced a specific term about web search application or web mining that is defined as automatic browsing process of and definition of behavioral patterns available for users about website input data [3].

There are many researches in this field that study based on the existing information about the user behavior in interaction with web to mine this knowledge and use it in different scopes in web including personalization of web pages and pages recommendations [4, 5], determining the relation between documents [6] and self-organization of web. Accordingly, it is vital to understand these issues and understand user needs in order to deliver better services to websites owners [7].

This generation requires benefit information mining from big data related to website. This data is obtained from different contents of web documents such as text, graphic, data from web structure such as html and xml tags, data from log web such as address (URL) and IP, date or time of access to web pages or data belonged to users such as registration, customer specifications, etc.

Therefore, information obtained from users' activities in web space is precious to some extent that large companies tend to collect this data intensively. Internet user trace is highly valuable for companies. With replacing

advertisement among the main content of site, advertisement and target content is subsidiary and you, as a user, use advertisements instead of the main content; in this case, the right of distinguishing between advertisement [8] and main content of website is eliminated [9].

In recent year, importance of mining users' interest is a beneficial and effective recommendation because e-commerce is highly applied. In general, data mining is an interdisciplinary movement that encompasses some scopes such as database, machine learning, rapid calculations and visualization. Specifically, data mining is a branch of artificial intelligence that employs automatic processes to find information.

Hence, increase in number of web-based applications leads to collection of a large collection of data existing in log web server that its results is discovery of information by application of web mining technics in web mining application pattern, in which, the searches conducted by users are considered in order to find beneficial information. Commercial groups employ this information in order to increase the profit obtained from personalization of websites for their customers based on the increase in their satisfaction level.

II. Literature review

In research [13], Tao et al. have introduced an integrated intention-based algorithm for web transaction mining that can process all data collections with different types of data simultaneously through an efficient method of intentional browsing and can convert the number of intentional browsing data to linguistic understandable items using fuzzy set concept. The main algorithm considers Web Transaction Mining (WTM) of a product on a webpage that is indicated with B[li] that means a webpage B with item li. The objective of this algorithm, which is focused on the purchased items (IWTMp), is to indicate the point that average level of user interest in an item can be shown with a specific set of Intentional Browsing Data (IBD) and this algorithm can improve predicting power of the main transaction data mining association rules. On the other hand, the not-purchased items and not considered before (IWTMnp) items are used to search webpages without any purchase and with intentional browsing data, that data mining transaction of main algorithm was not possible without intentional browsing data. In ref [14], Rana Forsati et al. presented a hybrid algorithm that uses information of user browsing and link between pages in order to propose pages to users. The introduced criterion for calculation of weight (page rank) of observed by users would employ "page visit duration" and "page visit frequency" that indicate interest rate and importance level of pages among users. In this section, the presented method combines linked information of pages and users usage. In this algorithm, pages are ranked based on a new criterion that indicates users' interests and importance of the page properly. Since, the accuracy of the first propose page is high in the provided algorithms and increase in number of proposed pages decreases accuracy considerably [10, 15, 16], the first page is proposed based on the usage data of users and the other pages will be classified based on the data of site structure and pages classification rules. This method can improve the quality of recommended page partially. The evidence for this idea is based on the implicit information of pages link because designers of webpages provide a link between pages with same titles and contents. On the other hand, use of site structure for new pages or pages with low visit frequencies makes it possible to be present in recommended pages and solves the problem in recommendations of new pages in dynamic sites.

In ref [11], Dipha Dixit et al conducted a study to provide two chained architectures to record direct understanding of users like a list of recommendations. It is considerable that this list is made of pages visited by user and the pages visited by other users that work on the same scope. Online navigation performance is developing; hence, intellectual information mining is a hard issue in this case. An approach of web mining usage is such designed that can work on web server logs that include user tracking. Implementation of algorithm and architecture will prove that accuracy in recording direct understanding of user is improved to what extent.

In research [12], Bommepally et al. discussed implementation and design of proxy analyzers that consist of three main components including input analyzer, data base loader, and data analysis.

1) Input analyzer: the purpose of this module is to extract benefit information from inputs. Inputs usually consist of IP address or host name, request time, user password, demanded address and demand status. 2) Database loading: this module list the information obtained from input analyzers. This module consists of four constituent components. The first component tracks the information path such as username, password, or IP address. The second component tracks information domain path. Third components tracks access time path. 3) Data analyzer: data analyzer is a module that interacts with user (such as network manager).

In ref [17], Ali Haroon Abadi et al. presented a developed version of PageRank algorithm in which, interest rate of users in webpages and Ant Colony Algorithm are used. A coefficient is added to PageRank algorithm in proposed version of this algorithm. Results obtained from simulation show that ranks of proposed version are more real and more number of distinguished ranks is generated. PageRank algorithm is a combination of web mining usage and web mining structure.

In ref [18], Kiatcomojong et al. proposed a design for classification of network traffic in three normal, average, and high rates of traffic. The proposed system consists of three main steps including data process, determining out of range data, and inputs classification. The initial input into system includes raw proxies inputs obtained from web proxy server. In preprocess, initial proxy inputs are processes and purified properly. When determining the

distinct part, the purified inputs are processed to find the distinct part and all natural inputs are filtered. Finally, inputs of distinct part (out of range) are classified to average, high and burst rate for each file category.

III. Research methodology

As we know, user performance consists of some logs and we found some logs of an institution that these logs saved as text format and included some parts such as searched date, access time, connection and start time, IP address of source, username, network card specification, and destination link.

Now, we should recover pages visited by user in order to analyze user performance and then distinguish Language Detection (distinguishing Persian Language), and extract the pattern from webpages at next step.

Keyword Matching is one of methods. Accordingly, it is possible to match keywords existing in each webpage with our keyword domain in order to determine category of pages.

Machine Learning algorithm is the other method in which, language of pages should be recognized then some web features should be illustrated for webpage, keywords are used at the next step and finally, Classification methods are used.

IV. Results

This part of study includes illustration and description of results obtained from implemented system. First, a brief review on different parts of system is presented in this part.

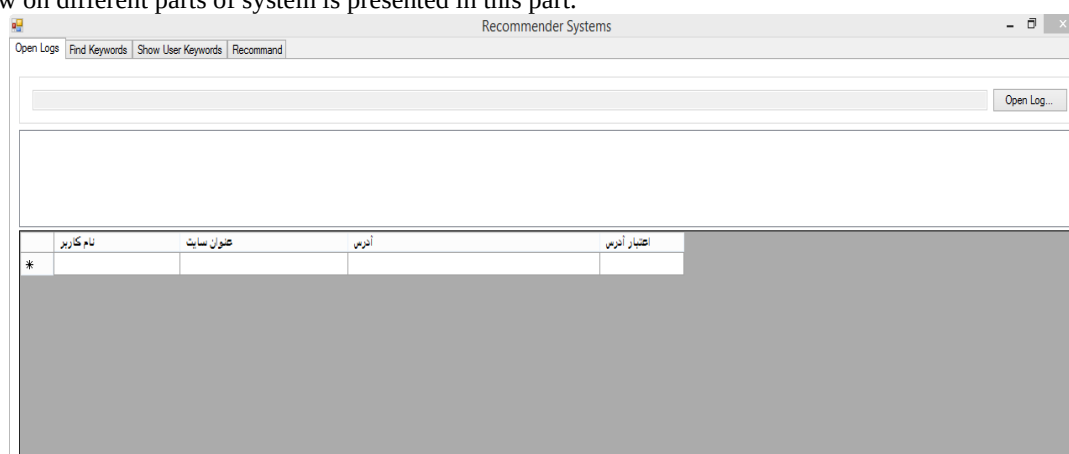


Figure 1: General Schema of recommender system

Format of stored logs in system are as username, date, IP (Internet Protocol) Address, access time, and search links that are illustrated in figure.

```
[09/May/2016 13:35:14] 172.27.22.65 jafarima@int.ir "???" http://seirogasht.com/fileweb.html
[09/May/2016 13:35:15] 172.27.102.114 tajvarpoura@int.ir "بترین ها | Bartarinha.IR | بترین ها" http://www.bartarinha.ir/fa/news/326714/%D8%B9%D9%84%
[09/May/2016 13:35:16] 172.17.8.205 amoozesh "TSETMC: : شرکت مدیریت فناوری بورس تهران" http://www.tsetmc.com/Loader.aspx?ParTree=15
[09/May/2016 13:35:16] 172.27.104.87 kouhpeymam@int.ir "MSN.com - Hotmail, Outlook, Skype, Bing, Latest News, Photos & Videos" http://www.msn.com/en-us?devi
[09/May/2016 13:35:16] 172.27.103.55 bagheria@int.ir "خراب جاری" http://172.16.0.167/117/Dashboard/NoActions.aspx
[09/May/2016 13:35:16] 172.17.8.16 hamediz@int.ir "http://enamad.ir/trustseal/symbol.aspx
[09/May/2016 13:35:17] 172.17.8.224 hamzeheil@int.ir "مينا نوروژی فرمدهوالم در خرو جراب فروشی کتم" http://tabnakbato.ir/fa/news/36192/%D9%85%D8%8C%D9%86%D8%A7-%D9%86%
[09/May/2016 13:35:18] 172.17.8.224 hamzeheil@int.ir "http://www.server.adro.co/pixel.aspx?id=1723&param=
[09/May/2016 13:35:18] 172.17.8.224 hamzeheil@int.ir "بررسی کد ملی" http://azmoon.railjob.ir/RegistrationNationalCode
[09/May/2016 13:35:19] 172.27.102.95 hoseini@int.ir "http://172.16.1.2/FormM/WebForm501.aspx?Id=WebForm501_46164,6784351
[09/May/2016 13:35:19] 172.17.8.224 hamzeheil@int.ir "http://www.server.adro.co/pixel.aspx?id=1723&param=
[09/May/2016 13:35:19] 172.17.8.224 hamzeheil@int.ir "http://www.server.adro.co/pixel.aspx?id=1723&param=
[09/May/2016 13:35:19] 172.17.8.224 hamzeheil@int.ir "http://www.server.adro.co/pixel.aspx?id=1723&param=
[09/May/2016 13:35:20] 172.27.102.114 tajvarpoura@int.ir "نامه دانشجویان بورسیه به روحانی" Bartarinha.IR | بترین ها | http://www.bartarinha.ir/fa/news/326711/%D9%86%D8%
[09/May/2016 13:35:24] 172.17.8.30 nazarpourf@int.ir "TSETMC: : شرکت مدیریت فناوری بورس تهران" http://www.tsetmc.com/Loader.aspx?ParTree=15
[09/May/2016 13:35:24] 172.27.102.114 tajvarpoura@int.ir "اسامی نهایی شهدای سپاه در خان طومان" Bartarinha.IR | بترین ها | http://www.bartarinha.ir/fa/news/326708/%D8%A7
[09/May/2016 13:35:26] 172.17.8.205 amoozesh "TSETMC: : شرکت مدیریت فناوری بورس تهران" http://www.tsetmc.com/Loader.aspx?ParTree=15
[09/May/2016 13:35:28] 172.17.7.159 nourmunda@int.ir "ورزش سه: جدول لیگ برتر (۹۵-۹۶) جام خلیج فارس" http://www.varzesh3.com/table/%D8%AC%D8%AF%D9%88%D9%84-%D9%84%D8%B
[09/May/2016 13:35:28] 172.27.101.89 jamshidigh@int.ir "برای فارس ایران AccuWeather لار آب و هوای می ۲۰۱۶ - پیشبینی" http://www.accuweather.com/fa/ir/lar/208371/may
[09/May/2016 13:35:30] 172.27.103.117 vahidim@int.ir "زنده گیری راسو نیم متری گونشخوار در یک خودرو سواری + عکس" http://www.yjc.ir/fa/news/5599747/%D8%B2%D9%86%D8%AF%D9%87
[09/May/2016 13:35:30] 172.17.7.43 hasanir@int.ir "فرم ثبت نام ویژه داوطلبان غیر راه اهدی" http://azmoon.railjob.ir/FreeRegister
[09/May/2016 13:35:30] 172.17.8.111 taghipourm@int.ir "Untitled Page" http://172.16.2.161/setsession.aspx
[09/May/2016 13:35:30] 172.27.102.107 estakhrp@int.ir "آمار اداره برنامه ریزی" http://172.16.0.181/RepDaeri_shar_weekly.aspx
[09/May/2016 13:35:31] 172.17.7.128 pouyaa@int.ir "مضاحیه نکتری" http://phdreg.pnu.ac.ir/
[09/May/2016 13:35:31] 172.17.7.211 parvah@int.ir "صفحه نخست - خبرآنلاین" http://www.khabaronline.ir/
[09/May/2016 13:35:32] 172.27.102.114 tajvarpoura@int.ir "تایید سازمان یافته بودن قاجاق در کشور" Bartarinha.IR | بترین ها | http://www.bartarinha.ir/fa/news/326704/%D8%AA
[09/May/2016 13:35:33] 172.17.3.201 Abonman3 "Untitled Page" http://172.16.2.163/setsession.aspx
[09/May/2016 13:35:34] 172.17.8.224 hamzeheil@int.ir "http://172.16.1.2/FormM/WebModalForm.aspx?Id=Show1,WebForm501_46164_Frame85_16927_WebModalFormSI
[09/May/2016 13:35:34] 172.17.7.211 parvah@int.ir "HE - جشنواره ۱۲۰" http://ads.goldiran.ir/show.aspx?s=53&u=635722216960711534
[09/May/2016 13:35:35] 172.17.8.30 nazarpourf@int.ir "TSETMC: : شرکت مدیریت فناوری بورس تهران" http://www.tsetmc.com/Loader.aspx?ParTree=15
[09/May/2016 13:35:36] 172.17.8.205 amoozesh "TSETMC: : شرکت مدیریت فناوری بورس تهران" http://www.tsetmc.com/Loader.aspx?ParTree=15
[09/May/2016 13:35:36] 172.17.8.224 hamzeheil@int.ir "http://www.server.adro.co/pixel.aspx?id=1723&param=
[09/May/2016 13:35:37] 172.17.8.239 alizadeha@int.ir "داستان های شاشا در تونال - لوازم مورد نیاز برای کوهنوردی در تابستان" http://tochaltrip.blogfa.com/post/104.aspx
[09/May/2016 13:35:38] 172.27.102.95 hoseini@int.ir "http://172.16.1.2/FormM/WebModalForm.aspx?Id=121,WebForm501_46164_Frame85_16927_WebModalFormSho
[09/May/2016 13:35:42] 172.27.32.56 ensafm@int.ir "Untitled Page" http://172.16.2.165/setsession.aspx
```

Figure 2: Stored logs format of users in server

By selecting the button Open Log, a window is opened like the following figure and the user is asked to give the address of storage location of server logs.

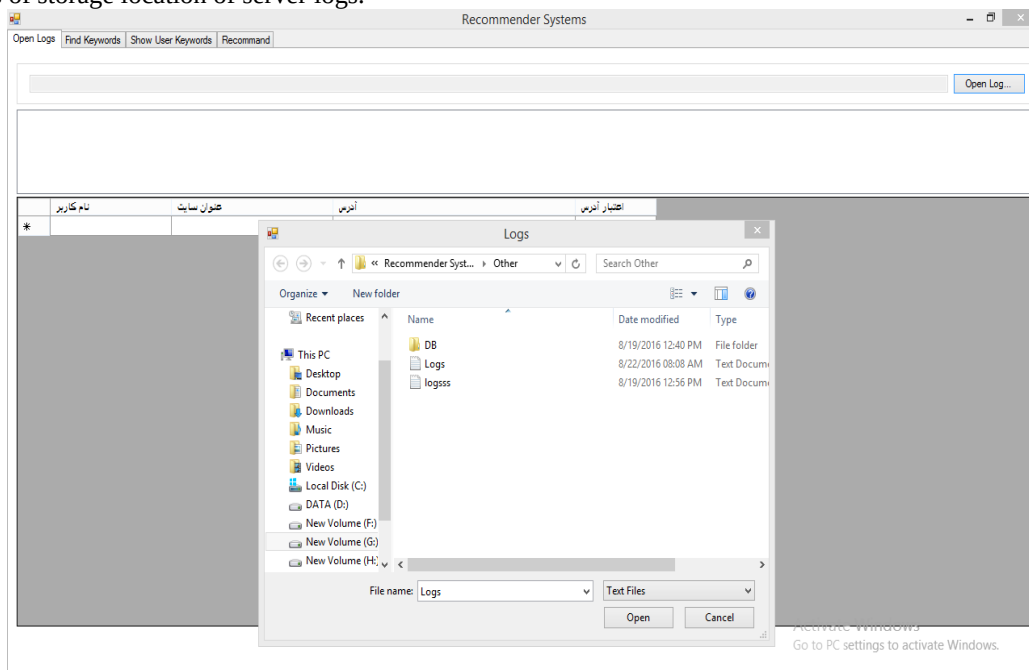
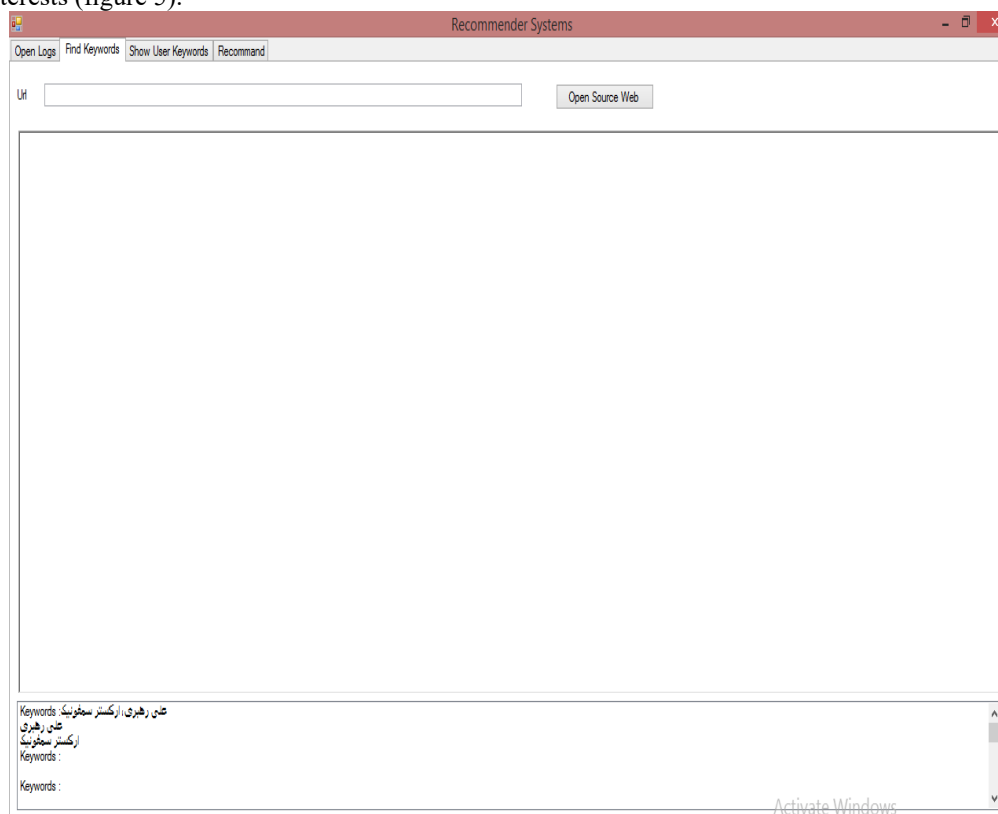


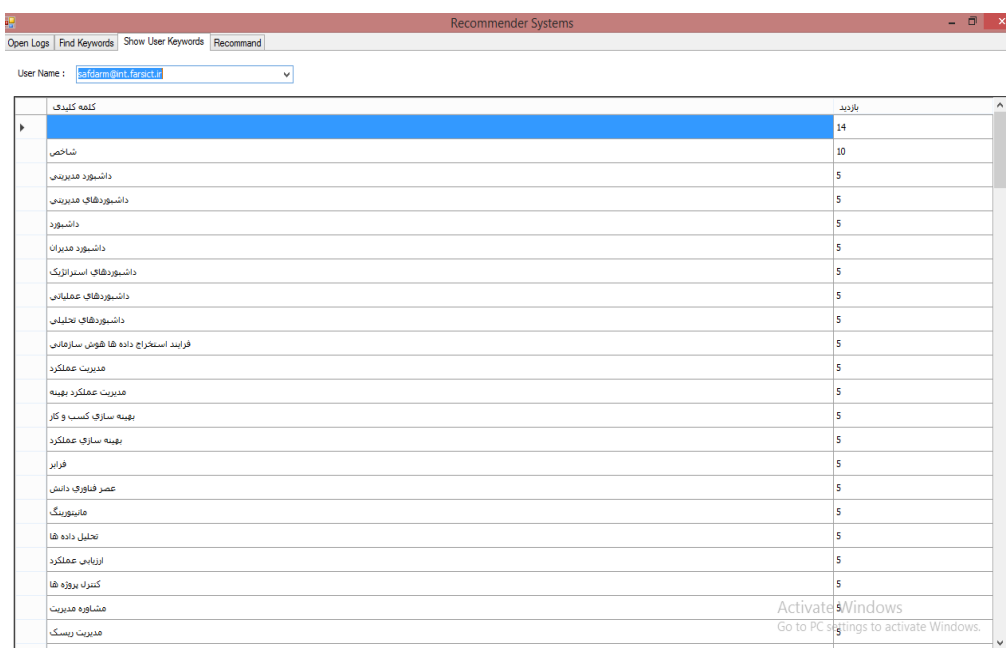
Figure 3: Selection window of stored logs location

After loading of recalled logs, system starts to check some information such as accuracy of entered addresses, exiting usernames in determined range and links searched by each user (figure 4).

In the next step in “find keyword” tab, all keywords of each site is extracted to find site content and understand users’ interests (figure 5).



In “show user keyword”, the most number of keywords searched by user are shown. For this purpose, first, a user is selected and then keyword list will be shown in table automatically (figure 6).

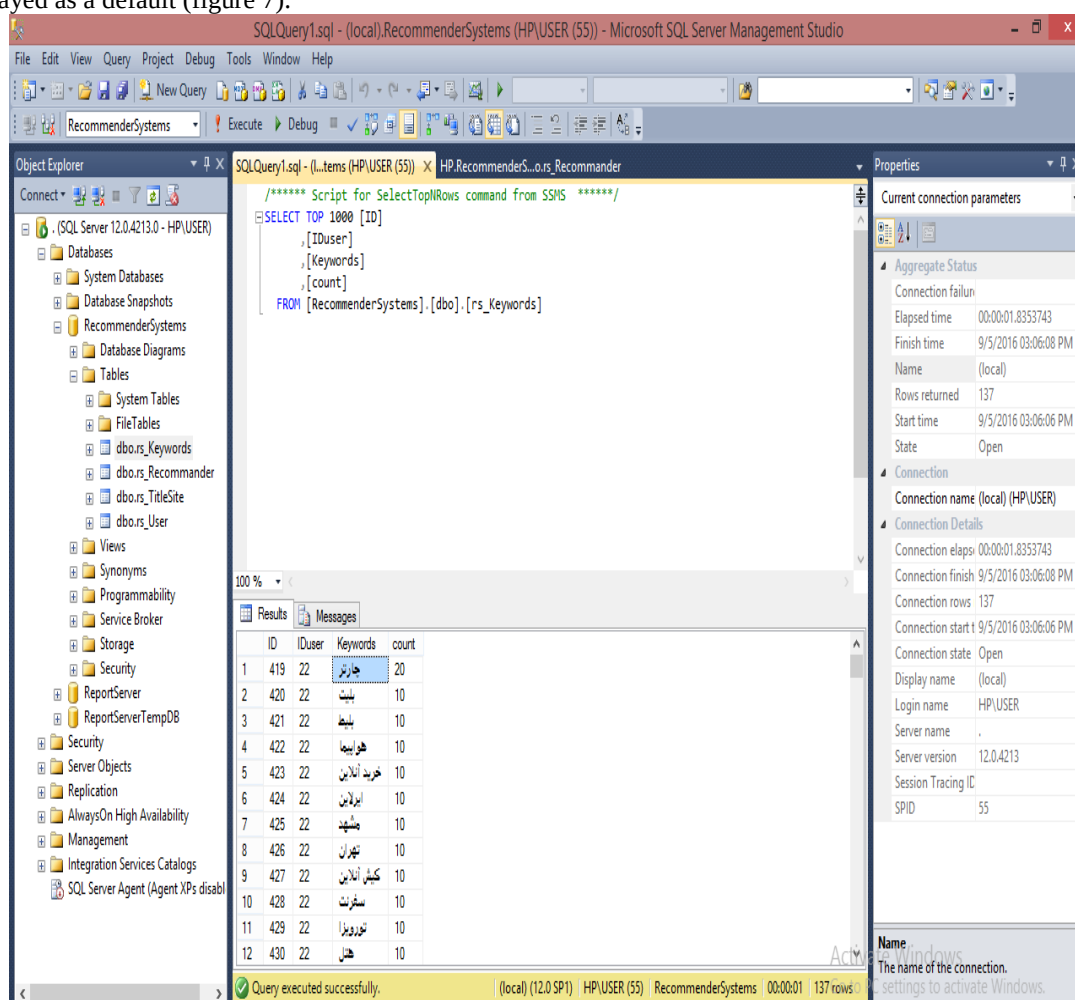


The screenshot shows a web application titled 'Recommender Systems'. It has a menu bar with 'Open Logs', 'Find Keywords', 'Show User Keywords', and 'Recommender'. Below the menu, there is a 'User Name' dropdown menu with 'sajadarm@nt.faract.ir' selected. The main content area is a table with two columns: 'کلمه کلیدی' (Keyword) and 'بازدید' (Views). The table lists various keywords and their corresponding view counts.

کلمه کلیدی	بازدید
شخص	14
شخص	10
دانشور مدیریتی	5
دانشورهای مدیریتی	5
دانشور	5
دانشور مدیران	5
دانشورهای استراتژیک	5
دانشورهای عملیاتی	5
دانشورهای تحلیلی	5
فرایند استخراج داده ها روش سازگاری	5
مدیریت عملکرد	5
مدیریت عملکرد بهینه	5
بهینه سازی کسب و کار	5
بهینه سازی عملکرد	5
فرایند	5
عصر فناوری دانش	5
ماینبریگ	5
تحلیل داده ها	5
ارزیابی عملکرد	5
کنترل پروژه ها	5
مشاوره مدیریت	5
مدیریت ریسک	5

Figure 6: Showing user's keywords

After mining keywords related to each searched site by user, we will be able to determine the favorite site of user correctly. For this purpose, keywords and searched link should be inserted into sql program and relevant table and in this regard, if the user searches for favorite keyword, the site consisting of searched information will be displayed as a default (figure 7).



The screenshot shows the Microsoft SQL Server Management Studio interface. The 'Object Explorer' on the left shows the database structure for 'RecommenderSystems'. The 'Query Editor' in the center shows a SQL query that selects the top 1000 rows from the 'rs_Keywords' table, ordered by 'count' in descending order. The 'Results' pane at the bottom shows the output of the query, which is a table with columns 'ID', 'IDuser', 'Keywords', and 'count'.

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [ID]
      ,[IDuser]
      ,[Keywords]
      ,[count]
FROM [RecommenderSystems].[dbo].[rs_Keywords]
    
```

ID	IDuser	Keywords	count
1	419	چارتر	20
2	420	بلیت	10
3	421	بلیط	10
4	422	هواپیمای	10
5	423	خرید آنلاین	10
6	424	ایرلاین	10
7	425	مسئله	10
8	426	تجارت	10
9	427	کشتی آنلاین	10
10	428	سفر	10
11	429	تور ویزا	10
12	430	هتل	10

Figure 7: Display database table of recommender system

At the last step and in “Recommend” tab, the proposed site for selected user is prepared based on the favorites of user in sql table and is displayed to user (figure 8).

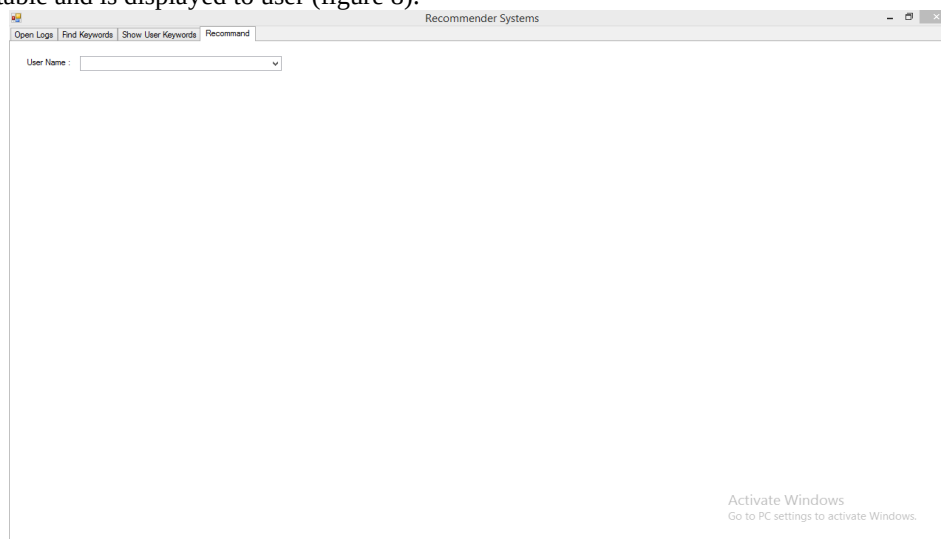


Figure 8: Recommendation window to user

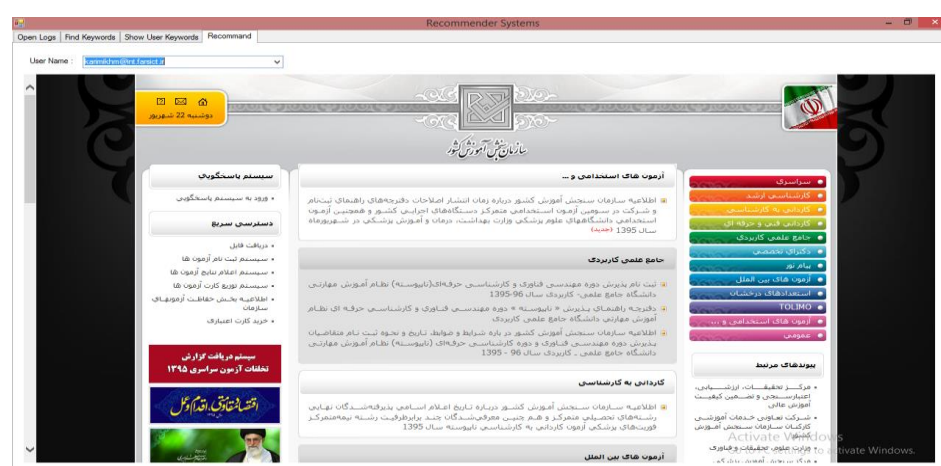


Figure 9: A recommended page to user

If there is not any site recommended to user, browsing engine of Google is displayed as default (figure 10).

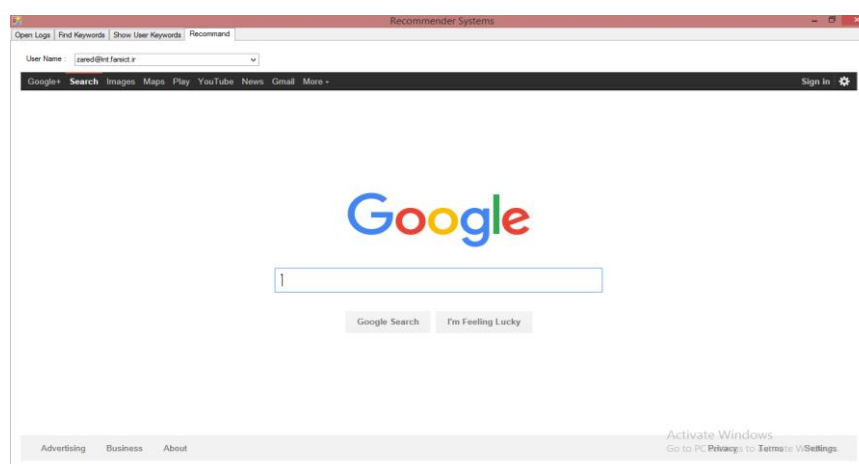


Figure 10: Displaying Google page to user

For instance, recommendation steps to users are shown as follows:

First, a user with fixed name is selected (figure 11).

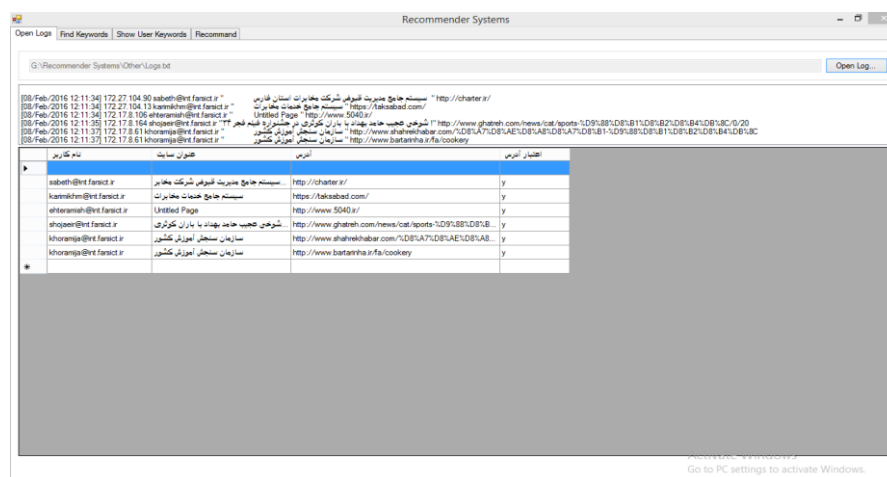


Figure 11: The recalled log of users

After acceptance of keyword information that contains airplane ticket, charter, Kish ticket, Mahan Air, etc., user data log will be mined (figure 12).

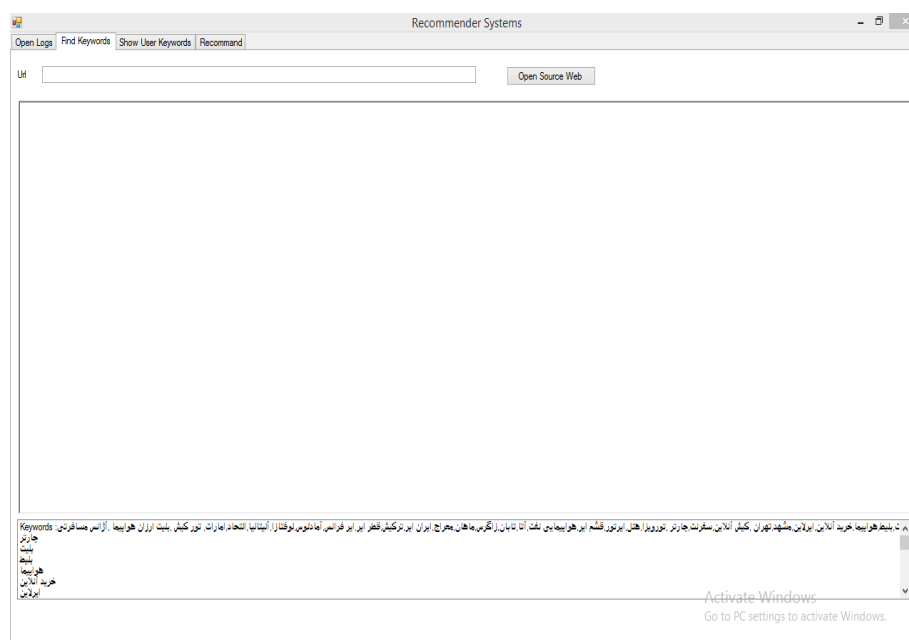


Figure 12. User's keywords

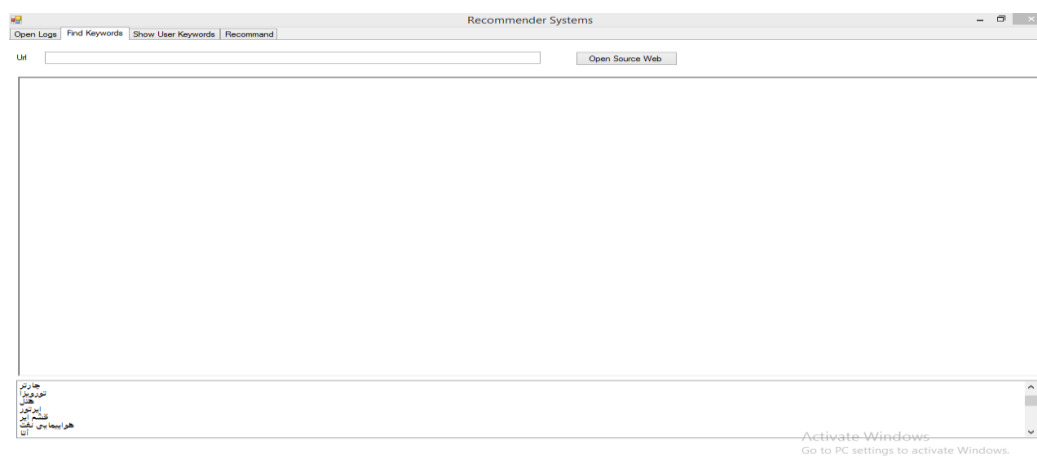
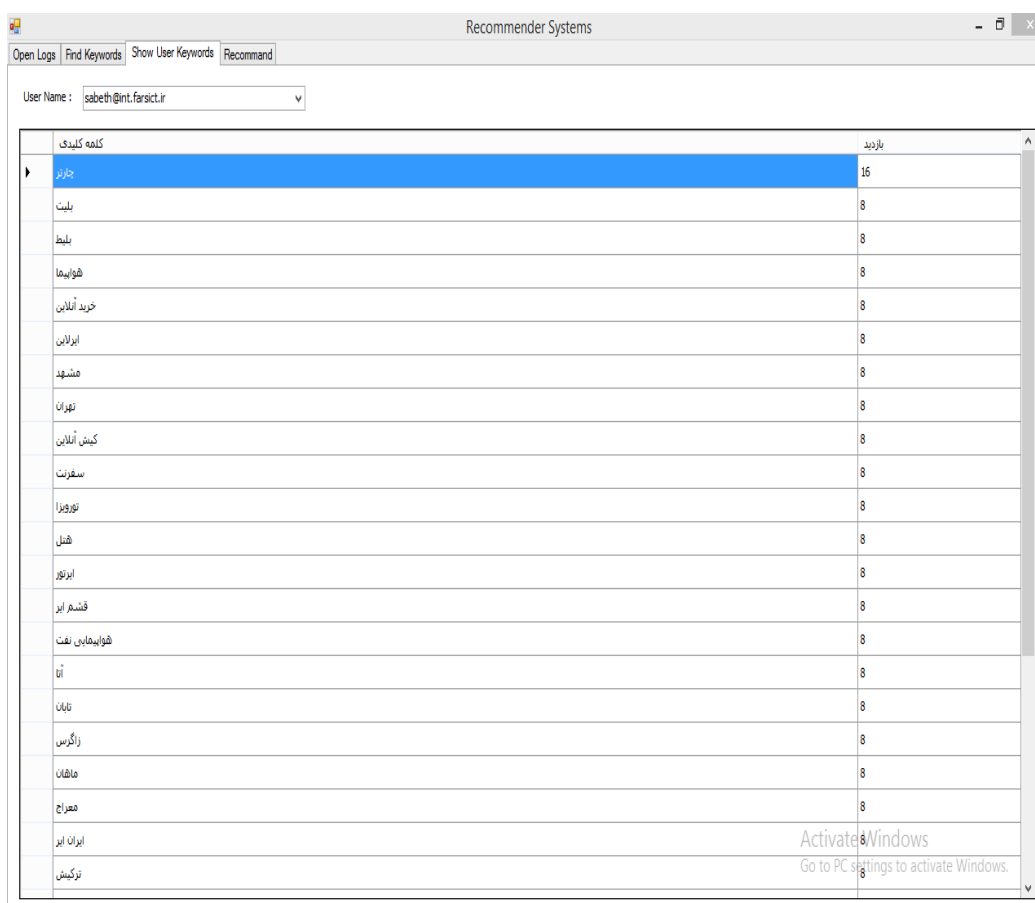


Figure 13: Displaying users' keywords

As is clear, Charter word has been more searched and is on the highest position of keywords; therefore, it can be found that the selected user had been searching for airline Reservation site. Therefore, Charter Site is recommended to this user (figures 14, 15).



کلمه کلیدی	تاریخ
چارتِر	15
بلیت	8
بلیط	8
هواپیما	8
خرید آنلاین	8
ایرلاین	8
مشهد	8
تهران	8
کیش آنلاین	8
سفرنت	8
تورویزا	8
شیل	8
ایران	8
قسمت ایر	8
هواپیمایی نفت	8
آنا	8
تایان	8
زاگرس	8
ماهان	8
مهرج	8
ایران ایر	8
ترکیش	8

Figure 14: Highest priority of keywords



Figure 15: Displaying Charter Site to user

V. Conclusion

Expanding increase in number of online stores and huge volume of provided products in these stores and increasing websites used in organizations, all firms should find a way to understand users' interest and present personalized recommendations in order to persuade users to use provided services. Creating profile for users is the first step in this regard, which includes some information such as organizational-environmental features, visiting experience of each user and given ranks based on user's search. The next step is use of a system with intelligent process that determined users' profiles and interest. Recommender systems help users to find their targeted items. Naturally, these systems are capable of recommending if they have not enough information about users and considered items of users. Finding valuable and structured information among a large volume of existing unstructured information can be used for many of cases. Growth and expansion of social networks has created a new opportunity for users to share their ideas and interests with each other. In this regard, recommender systems can automatically can find favorite information of users and recommend them. The purpose of recommender systems is indeed ranking system items in terms of users' interests in order to propose high-rank items to user. This process can increase system efficiency and reduce user's confusion among large volume of information in virtual space. Communicational networks contribute to simple access to information. Meanwhile, increasing information has led to "information overflow". In this regard, recommender systems have been emerged in response to overloaded information. These systems recommend some contents matched with users' needs. Recommender systems need accurate models of features, preferences, needs, system's knowledge about user and users' activities so that these options have led to emergence of a group of recommender systems to provide users with appropriate recommendation. There are different types of information resources in system based of the system application. This information might include users' scores to items, personal information of users, content related to system's items, communications existing in social networks and information related user's situation. Recommender systems cope with overloaded information through automatic recommending to users based on their interests and benefiting from statistical technics and knowledge discovery to recommend products and services to users. Recommender system would enable user to find suitable and favorite option without facing any repetitive or unhelpful information rapidly. The advantage of this system is performing based on the users' activity and collecting their behavior and interest. There are various resources and methods for data collection. Explicit method is a data collection method, in which user explicitly expresses favorite options. In this method, user highlights and enters interest level in system and indeed enhances information level of system and system can predict priorities and interests of users based on these scores. The other data collection method is implicit method that is a little harder and system should track and monitor behaviors and activities of users to find his/her interests and favorite options. This information consists of click paths, times spent on each page, closed pages, etc. in addition to explicit and implicit information, some systems use personal information of users. Studies indicate that recommender systems do this action very well; however, they face some deficits. Hence, it is vital to improve system and organization performance and results providing efficient algorithms. An approach of content recommender system based on users' logs existing on proxy server was presented in this study considering interests of users. This system not only was effective in search speed of sites considered by users to meet their needs but also reduced network traffic at organizational level considerably and network bandwidth enjoyed less traffic. Therefore, users could access to their required sites with the least time cost and this expanded quality and satisfaction among organization coworkers.

References

- [1] Forsati, R. M. Mohammadreza, An algorithm based on link structure of pages and users' data usage for webpages recommendation, Second Data Mining Conference, Iran, Tehran, Amirkabir University of Technology, Institute for Research on Data Processing Gita
- [2] Sarah, S., Ali, H. A., Massoud, R. A. (2013). Providing a developed version of PageRank algorithm to rank web pages, National Conference on Computer Engineering and Sustainable Development with a focus on computer networks, modeling and system security, Mashhad, institutions of higher education of KHAVARAN
- [3]. Cooley, R., B. Mobasher, and J. Srivastava, Data preparation for mining World Wide Web browsing patterns. Knowledge and information systems, 1999, 1(1): p. 5-32.
- [4]. Etzioni, O., The World-Wide Web: quagmire or gold mine? Communications of the ACM, 1996, 39(11): p. 65-68.
- [5]. Cooley, R., B. Mobasher, and J. Srivastava. Web mining: Information and pattern discovery on the World Wide Web in Tools with Artificial Intelligence, 1997. Proceedings, Ninth IEEE International Conference on. 1997. IEEE.
- [6]. Kazienko, P. Filtering of web recommendation lists using positive and negative usage patterns. in International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. 2007. Springer.
- [7] Srivastava, J., et al., Web usage mining: Discovery and applications of usage patterns from web data. ACM Sigkdd Explorations Newsletter, 2000, 1(2): p. 12-23.
- [8]. Gibson, D., J. Kleinberg, and P. Raghavan, Inferring web communities from link topology. in Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space---structure in hypermedia systems: links, objects, time and space---structure in hypermedia systems. 1998. ACM.
- [9]. Singh, B. and H.K. Singh. Web data mining research: a survey, In Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on. 2010. IEEE.
- [10]. Pabarskaite, Z. and A. Raudys, A process of knowledge discovery from web log data: Systematization and critical review. Journal of Intelligent Information Systems, 2007. 28(1): p. 79-104.

- [11]. Shivaprasad, G., et al., Neuro-Fuzzy Based Hybrid Model for Web Usage Mining. *Procedia Computer Science*, 2015. 54: p. 327-334.
- [12]. Dixit, D. and J. Gadge, Automatic recommendation for online users using web usage mining. *arXiv preprint arXiv:1009.0433*, 2010.
- [13]. Bommepally, K., et al. Internet activity analysis through proxy log. in *Communications (NCC), 2010 National Conference on*. 2010. IEEE.
- [13]. Tao, Y.-H., et al., A practical extension of web usage mining with intentional browsing data toward usage. *Expert Systems with Applications*, 2009, 36(2): p. 3937-3945. http://www.civilica.com/Paper-IDMC02-IDMC02_016.html, 1387.
- [14]. Haveliwala, T.H. Topic-sensitive pagerank, in *Proceedings of the 11th international conference on World Wide Web*. 2002. ACM.
- [15]. Mobasher, B., et al., Discovery and evaluation of aggregate usage profiles for web personalization. *Data mining and knowledge discovery*, 2002, 6(1): p. 61-82.
- [16]. Aktas, M.S., M.A. Nacar, and F. Menczer, Personalizing pagerank based on domain profiles. in *Proc. of WebKDD*. 2004.
- [17]. Kiatkumjounwong, N., et al. Analysis and classification of web proxy logs based on patterns of traffic rates, in *TENCON 2014-2014 IEEE Region 10 Conference*. 2014. IEEE.
- [18]. Zeng, Z., An intelligent e-commerce recommender system based on web.