Study on online community user motif using web usage mining

Meera Alphy¹ and Dr Ajay Sharma²
¹PhD Scholar, Department of Computer Science and Engineering, SRM University, Haryana, India
²Assistant Professor, Department of Computer Science and Engineering, SRM University, Haryana, India

Corresponding author e-mail: meeraalphy.urumbath@gmail.com

Abstract. The Web usage mining is the application of data mining, which is used to extract useful information from the online community. The World Wide Web contains at least 4.73 billion pages according to Indexed Web and it contains at least 228.52 million pages according Dutch Indexed web on 6th August 2015, Thursday. It’s difficult to get needed data from these billions of web pages in World Wide Web. Here is the importance of web usage mining. Personalizing the search engine helps the web user to identify the most used data in an easy way. It reduces the time consumption; automatic site search and automatic restore the useful sites. This study represents the old techniques to latest techniques used in pattern discovery and analysis in web usage mining from 1996 to 2015. Analyzing user motif helps in the improvement of business, e-commerce, personalisation and improvement of websites.

Keywords. Log file, Pattern analysis, Pattern discovery, Pre-processing

1. Introduction

1.1 Initial Period
As everyone believed Oren Etzioni mentioned web mining in his paper “The World Wide Web: quagmire or gold mine?” According to his Web mining [1] may be broken down into the following subtasks’ Resource Discovery, Information Extraction and Generalization. Resource discovery was unearthing records and favors on the Web. Information Extraction automatically took out particular data from newly revealed network resources. Generalization was revealing motif at individual online community with other online community.

1.2 Later Period
Web mining [21] is the use of data mining techniques to take out information from the online repository. Web mining consists of web content mining, web structure mining and web usage mining. Web content mining is the process of taking out useful data from the collective data of web document. Web structure mining was the process of finding already available structure information from the online community. It can be divided into two types’ hyperlinks and document structure. Web usage mining is the use of data mining techniques to find a useful pattern of data by analyzing web users browsing patterns in the online community.

2. Web usage mining (wum) process
Web usage mining [29] is the process of discovering the patterns used by the web user from the web sites. Web usage having mainly three phases such as pre-processing, pattern discovery and pattern analysis. Collecting the necessary data’s from web server logs, proxy server logs or
browser logs. Log files are automatically stored by the web server which contains the information about the web pages accessed by the web user during browsing the websites. Pre-processing [19] is the first phase in WUM which having sub tasks’ data cleansing, user identification and user session identification. Data cleansing coined unwanted entities to be removed. User identification can be done by internet protocol address, cookies and registration of users. User session identification can be done by analyzing a web user’s browsing activity in a particular time period. Pattern discovery is the main factor in the web usage mining. The pattern can be discovered using statistical analysis, association rules, clustering, classification, sequential pattern, dependency modeling and etc. Pattern analysis is the third phase in web usage mining which gives refine the output patterns from the pattern discovery which can be done using Knowledge query mechanism or visualisation technique.

![Web Usage Mining (WUM) process](image)

**fig 1: Web Usage Mining (WUM) process**

### 3. SURVEY ON WEB Usage MINING

#### 3.1 Knowledge and motif discovery on world wide web

R Cooley [2] et.al proposed web mining can be used for two distinct ways’ web content mining and web usage mining. Web content mining is defined as extracting knowledge from different resources of World Wide Web. Two approaches are mainly used for web content mining; they are agent based approach and database approach. Web usage mining is defined as extracting user
access patterns from the browser. Web usage mining has four processes: pre-processing, transaction identification, pattern discovery tools and pattern analysis tools. In pre-processing, the first task was data cleaning, which removes irrelevant information and the second task was transaction identification which gets all the information about the user browsing for a particular visit. Path analysis, association rules, clustering and classification, one among these can be used for discovering the web transactions. Path analysis could be used to occur in a particular way most regularly look in on track in a Web site. Association rule discovery techniques are in most cases applied to a structured set of data held in a computer, especially one that is accessible in various ways of proceedings where each proceeding was made up of a collection of individual data. Discovering classification rules allow one to start to exist in the collection of individual data to be a member of a specific set according to their common characteristics. Clustering analysis precede one to set of clients together or facts that have similar attributes. Web miner tool is mainly used tool for pattern discovery and analysis.

3.2. Interesting navigational patterns can be discovered and utilized to improve the services provided on the site

Georgios Paliouris [3] et.al proposed the group of the people who used an online site into set of people who used similar navigational way of behaving in the online site. The database used in these experiments was collected from the Web site of the Advanced Course on Artificial Intelligence 1999 log files. The unverified machine learning algorithm COBWEB is used to arrange the consumer of the online community, who go after relative paths, into a small group of people having particular characteristics in common. Author given more importance to the elucidation of the common behaviour group that was developed through this procedure. And also COBWEB was an application of clustering which was suitable for symbolic training data such as access sessions. The resulting communities can be used to improve the services provided in the Web site. Author used each trouble-free metric for different navigational behaviour community. This data were used by the site developer to re-arrange it in a useful way that is customized to the importance of each community. Limitation was the representation of user sessions for the learning algorithm. Advantage was reducing the overlap and increasing the coverage.

3.3. Heuristic method for real time web usage mining

Florent Masseglia [4] et.al proposed a heuristic method for real time web usage mining. In view of the rising number of online community’ caller, and the rate at which operates the web site at which consumer’s activities may change from one particular time period to another time period, the output obtained by examining an access online repository may lose its importance, if it is exploited too late. Author proposed a Heuristic based Distributed Miner (HDM) designed to avoid such obstacles and to get regular activities motif. This method allows discovering regular activities motif in real time, which doesn’t depend on connected users. In view of how quick the regular activities motif can vary since the final examination of the access online repository, this output thus provides completely tailored direction-finding schemas for user activity prophecy. Based on a distributed heuristic, author says this method also answers quite a few unfolded issues within the data mining framework are recognizing most popular zone and find out very long chronological motif and collaborative data mining. Corroborate HDM activity and output, the author had verified it with local databases, where each input, arranged in a particular order was allocated a thread (imitating in washer machine) on the evaluating in the machine. The advantage was recognizing a new kind of motif, collaborative data mining and mining vast regular chronological patterns. Disadvantage was limited neighbourhood operators used.

3.4 Ant clustering algorithm and linear genetic programming for web usage mining

Ajith Abraham [5] et.al proposed an ant clustering algorithm to find out web usage mining and a linear genetic programming approach to examine the visitor inclination. First-hand outcome allow
that ant colony clustering perform well when compared to a self organizing feature map for clustering Web usage motifs, even though the performance, quality of state is not that performance efficient when compared to evolutionary-fuzzy clustering approach. An ant colony clustering was proposed to set apart from other visitors and thereafter a linear genetic programming approach to analyze the visitor motif. Data was collected from web access log data at the Monash University's Web site. The inputs were obtained from the statistical text data generated by the online repository from 1st January 2002 to 7th July 2002. Isolating the useful data from the raw file was an important part in the data pre-processing phase. After some preparatory analysis, the author selected the statistical data be made up of domain byte requests, hourly page requests and daily page requests as a focal point of the cluster structures for taking out Web consumer’s usage motifs. The maximum currently used data were recorded with higher index while the minimum currently used data were recorded with a lower index. Disadvantage for these approaches where each of the datasets in the approach was run twice (for f= my to 10, 00,000) in order to test if somehow the output were alike. Hence time consumption is high. The advantages of this approach were classification space was always 2D, non-parametric and toroidal (Toroidal describes something which resembles or relates to torus or toroid. Toroid was a doughnut-shaped object which resembles a torus).

3.5 USIA approach for users and session identification
Zhang Huiying[6], et.al the USIA was used for online users’ recognition and session recognition. There are a lot of online users following a sequence of records come from the same Internet Protocol address in online repository files. Time consumption is higher when checking every record in an online repository. Hence, record efficiency, reduced. If two record internet protocol address is same, then it’s assumed that the two records used by the same web user. The premise of USIA approach was based on two prerequisites as follows by acceptable user’s identification precision and algorithm efficiency. Algorithm’s advantage was good precision, high efficiency and favourable structure of the store. Algorithm’s disadvantage was algorithm need check several elements when it examines whether the current online user is the same online user or not, this results time complexity. The data were collected from the online repository file of Tianjin University's Web site (http://www.tju.edu.cn) and time period of 1st March 2003 to 7th March 2003. The size of the total online repository file was 105M bytes and after pre-processing the size of the file was 378,747 records.

3.6 An individual approach for person name bipolarization
Chien Chin Chen [7], et al. an individual approach that founds out the polarity of persons’ names. Author gave the indication of data in the principal eigen vector of PCA could separate person name to bipolar sets fast. Two techniques’ weighted correlation coefficient and off-topic block elimination were introduced to solve sparseness problem. Data were collected from 33 news documents related to the opening game from Google news and also 166 news documents related to the 2010 World Cup Final. The advantage of the approach was applied to different languages and diverse topic, identify bipolar person names in topic documents correctly without using any external knowledge source and context-oriented. Disadvantage of the approach were person names possessed neutral orientations.

3.7 Motif-based clustering approach to grouping customer transactions
Yinghui Yang [8] et.al using a motif-based clustering approach to assembling consumer transactions. The main idea was to cluster consumer transactions such that motifs generated from each cluster, while similar patterns to be within the cluster, are very different patterns generated from other clusters. The objective of the function was that maximize in order to goal a good clustering of consumer transactions. Advantage was user-centric Web usage data that illustrate that GHIC shows a highly effective clustering of transactions. For a good clustering algorithm had two
measures of cluster qualities’ internal quality and external quality. The internal quality measure that used was the average M measure across all set of two clusters generated by GHIC. The data was collected from market data vendor web browsing data. Disadvantage was algorithm should haven’t considered global objective functions to develop better heuristic functions.

3.8 Semantic web usage mining approach for consumer emotions and behaviours through self-reporting and behavioural tracking

A.C.M. Fong [9] et al. proposed a method a semantic web usage mining approach for consumer emotions and behaviours through self – reporting and behavioural tracking. Thoughts and mindset of the consumer have been found to influence a person’s activity of moving from site to site on the online repository. The proposed approach has four different steps of Personal Web Usage Ontology (PWUO). First step, Personal Web Usage Lattice (PWUL) using input as relative online community records which consists of a group of attributes of a specific time period and a group of attributes on specific subjects necessary for the development of consumer emotions periodic patterns influencing the web surfing activities. A set of attributes at a specific period coined on different time period of a day (e.g., morning, afternoon, etc.), the days of a week (e.g., Monday, Tuesday etc.) and other concepts (e.g., weekdays, weekend, etc.). A set of attributes specific subject defines the useful domain ontological concepts. Second step, Global Web Usage Lattice (GWUL) to entitle all specific time period pattern-based worldwide online repository access activities and relative arranged in order of rank between these activities which can be very large. Third step, Global Web Usage Ontology (GWUO) derived from GWUL step by making a map of World Wide Web access activities and their relative arrangement in order of rank into a category of tasks having common property or attribute. The fourth step, PWUO is the based on information that can subsequently provide meeting the consumer’s needs more effectively and efficiently, making interactions faster and easier and, consequently, increasing consumer satisfaction and the likelihood of repeat visits. Here the inputs were obtained from online repository file Nanyang Technological University, Singapore. The main advantage of this approach was a fast response (low latency) and other were good aid, relevant results. The main limitation of this method was not ease of use.

3.9 New method to identify navigation related usability problems

Ruli Geng [10] et al. proposed a new method to identify navigation related usability problems. The inputs were obtained from Furniture Giveaway (FG) 2009 website. The method includes three modules where usage pattern extraction, Ideal User Interactive Path (IUIP) modelling and usability problem identification. Usage pattern extraction was done by extract real activity of accurately ascertaining the position and planning and following a route from server repository and finds a repeated design for some typical work data. Usage pattern extraction can be done by process’ Data Preparation and Pre-processing, Transaction Identification and Trail Tree Construction. Data Preparation and Pre-processing consist of data cleaning, User identification, User session identification and path completion. Transaction Identification was sequence of the pages seen by the consumer to purchase the product. Trail Tree Construction used the same path followed by the different consumers make a tree structure which developed by trie algorithm and its frequency taken as trail tree. Meanwhile design IUIP models for same work. IUIP models are based on the perception of consumer action and can be noted by predicting tracks for particular consumer-based on a series of action. IUIP model specifies the maximum time visited by the consumer in a particular page and sequence of pages followed (path). This idea is obtained by reducing the complexity of Adaptive Control of Thought-Rational (ACT-R) model. Usability problem identification helps to find the main cause of web design problems using logical deviation calculation and temporal deviation calculation. Logical deviation was calculated by paths participated by IUIP model, but not select, it can be only count as one deviation and total transaction is equal to the deviation counted by total page visited by a consumer. Temporal
deviation calculated by the consumer visited the page than the maximum time allowed for each page can count as one deviation and total deviation calculated by total deviation counted from total pages visited by a consumer. The main advantage of this model was usability improvement and reducing the complexity in the larger system. The limitation of this method was user satisfaction and presentation aspects of usability.

3.10 The request dependency graph representation of the relationships among HTTP Requests
Jun Liu [11] et.al introduced the request dependency graph was a graph representation of the relationships among HTTP requests. A directed link from A to B in the graph indicates the accessing of online community object B is caused by the accessing of A; hence B depends on A. Methodology to develop such a graph by mining the of time and connecting information among aggregated HTTP requests. The main objective of design and implement algorithm for primary request identification which was a significant task of web usage mining. Advantage was algorithm applied to large-scale real world, where the web access online repository shows that the request dependency graph is a useful tool for web usage mining.

| TITLE                                      | AUTHOR                                      | PROPOSED METHOD                                                                 | LOG FILE                                                                 | ADVANTAGE                                                                 | DISADVANTAGE                                                                 |
|--------------------------------------------|---------------------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------|----------------------------------------------------------------------------|----------------------------------------------------------------------------|
| Knowledge and motif discovery on world wide web | R.Cooley, B.Mobasher, J.Srivastava [2]       | Web Mining can be used for two distinct ways’ web content mining and web usage mining |                                                                          |                                                                            |                                                                            |
| Interesting navigational patterns can be discovered and utilized to improve the services provided on the site | Georgious Paliours, Christos Papatheodorou, Vangelis Karkaletsis [3] | the group of the people who used an online site into set of people who used similar navigational way of behaving in the online site | Web site of the Advanced Course on Artificial Intelligenc e | represention of user sessions for the learning algorithm | reducing the overlap and increasing the coverage |
| Heuristic method for real time web usage mining | Florent Masseglia, Maguelonne Teisseire, Pascal Poncelet [4] | heuristic method for real time web usage mining | local databases | recognizin a new kind of motif, collaborati ve data mining and mining vast regular chronologi cal patterns | limited neighbourhoo d operators used |
| Ant clustering algorithm and linear genetic programming for web usage mining | Ajith Abraham, Vitorino Ramos [5] | ant clustering algorithm to find out web usage mining and a linear genetic programming approach to examine the visitor inclination | Monash University' s Web site. | Classificati on space was always 2D, non-parametric and toroidal | time consumption is high |
| USIA approach for users and session | Zhang Huizing, Liang Wei [6] | USIA was used for online users’ recognition and session recognition | Tianjin University' s Web site | good precision, high efficiency | algorithm need check several elements when it examines |
| identification | describes the online user navigational patterns and favourable structure of the store whether the current online user is the same online user or not, this results time complexity |
|----------------|----------------------------------------------------------------------------------------------------------|
| **An individual approach for person name bipolarization** | Chien Chin Chen, Zhong-Yong Chen, Chen–Yuan Wu [7] individual approach that founds out the polarity of persons’ names 33 news documents related to the opening game from Google news and also 166 news documents related to the 2010 World Cup Final approach was applied to different languages and diverse topic, identify bipolar person names in topic documents correctly without using any external knowledge source and context-oriented. approach were person names possessed neutral orientations. |
| **Motif-based clustering approach to grouping customer transactions** | Yinghui Yang, Balaji Padmanabhan [8] cluster consumer transactions such that motifs generated from each cluster, while similar patterns to be within the cluster, are very different patterns generated from other clusters market data vendor web browsing data user-centric Web usage data that illustrate that GHIC shows a highly effective clustering of transaction s global objective functions haven’t considered to give better huristic function |
| **Semantic web usage mining approach for consumer emotions and behaviours through self-reporting and behavioural tracking** | A.C.M. Fong, Baoyao Zhou, Siu C.Hui, Jie Tang, Guan Y.Hong [9] semantic web usage mining approach for consumer thoughts and mindset of have been found to influence a person’s activity of moving from site to site on the online repository Nanyang Technological University website fast response (low latency), good aid, relevant results Not ease of use. |
| **New method to identify** | Ruli Geng, Jeff Tian [10] new method to identify navigation related usability Furniture Giveaway (FG) usability improvemen t and user satisfaction and |
4. Areas were web usage mining applied

E-Learning Personalization [12] recommendation system put forward the skills through study resources were reckon based on the current learner’s recent activity of accurately ascertaining his/her position and planning and following a route history, as well as utilizing likes and dislikes among learners’ predilection and educational content. Intelligent agents to find missing and exploited children [13] provide a user interface system were customizable. Data mining used intelligent agents to find missing and exploited children over the World Wide Web, chat rooms and intranets. Data mining in World Wide Web is known as web mining. Predictive Web usage visualizations [14] help to understand usage information and identify major traffic motif in e-commerce sites. Web customized index combination in commercial web sites [15] helps for self customization and web site usually contains vast amounts of information across hundreds of pages. Without proper direction for the visitor often wanders pointlessly without visiting significant pages, loses attention, and exit the site sooner than expected. Web ad’s effectiveness [17] making it easier for companies to conduct business and transfer information to customers. Online Handwritten Script Recognition [18] helps automatic identification of handwritten script facilitates many important applications such as automatic transcription of multilingual documents and search for documents on the Web containing a particular script. Strong regularities in Web surfing helps to characterize empirical Web usage regularities [20]. Web personalization [22] [26] techniques have three major categories were decision rule-based filtering, content-based filtering, and collaborative filtering [16]. Decision rule-based filtering surveys users to obtain user demographics or static profiles, then lets Web sites manually specify rules based on them. Content-based filtering relies on items being similar to what a user has liked previously. Collaborative filtering uses explicit user ratings of products or preferences to sort user profile information into peer groups. It then tells users what products they might want to buy by combining their personal preferences with those of like-minded individuals E-science in astronomy [23] helps in tracking the web and SQL queries in E-science usage. Unsupervised feature selection for social media data [24] helps to find a tough task due to nonappearance of label information based on which feature relevances

5. WEB USAGE MINING TOOLS

SpeedTracer [27] a Web usage mining and analysis tool, was developed study user surfing pattern by checking out the online repository files with data mining techniques. Rapid Miner 4.6 [28] supports automatic web page visitor session extraction and clustering of the visit. Weka [29] helps

| navigation related usability problems | problems and method includes three modules such as usage pattern extraction, Ideal User Interactive Path (IUIP) modelling and usability problem identification | website | reducing the complexity in the larger system | presentation aspects of usability |
|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----------------------------------------------|----------------------------------|
| The request dependency graph representation of the relationships among HTTP Requests | Jun Liu, Cheng Fang, Nirwan Ansari [11] introduced the request dependency graph was a graph representation of the relationships among HTTP requests and primary request identification which was a significant task. | Local database | applied to large-scale real world | Less accuracy |

Table1 : table of review
in collecting visualization tools and algorithms for data analysis and predictive modeling. Orange [30] is a data mining tool which can be used for web usage mining analyzing.

CONCLUSION

Tim Berners-Lee invented the World Wide Web in 1989. In 2009 Tim burners lee coined the term Linked Data in his semantic web project. Linked data describe the relation of a web page to another web page. This made World Wide Web semi structured and easy to access related data. Limited numbers of data’s are accessed under this linked data in online communities. Hence unlimited data’s in online community gives Web usage mining importance in day to day life. Web usage mining can be applied to every area in the online community to extract useful information. Vast areas of research available in web usage mining to develop user friendly online community and can be generated more structured data’s in world wide web.

ACKNOWLEDGEMENT

Firstly, I would like to thank almighty for making me motivated. Secondly, Dr Ajit Kumar and Dr jawahar Lal, SRM university, Haryana, Delhi-NCR, Sonepat for their kind help.

REFERENCES

[1] Oren Etzioni “The World Wide Web: quagmire or gold mine?” Appears in Comm. of ACM, Nov. ’96
[2] R. Cooley, B. Mobasher, and J. Srivastava “Web Mining: Information and Pattern Discovery on the World Wide Web” IEEE 1997
[3] Georgios Paliouras, Christos Papatheodorou, Vangelis Karkaletsis, Costantine Spyropoulos, Panayiotis Tzitziras “From Web Usage Statistics to Web Usage Analysis” volume : 2 Pages: 159 - 164 IEEE 1999
[4] Florent Masseglia, Maguelonne Teisseire, Pascal Poncelet ” Real Time Web Usage Mining: a Heuristic based Distributed Miner” pages: 288 - 297 volume:1 IEEE 2002
[5] Ajith Abraham, Vitorino Ramos “Web Usage Mining Using Artificial Ant Colony Clustering and Linear Genetic Programming” pages: 1384 - 1391 Volume:2 IEEE 2003
[6] Zhang Huiying, Liang Wei “An Intelligent Algorithm of Data Pre-processing in Web Usage Mining ”, Proceedings of the 5th World Congress an Intelligent Control and Automation, Pages: 3119 - 3123 Vol.4 June 15-19, 2004, Hangzhou, P.R. China
[7] Chien Chin Chen,Zhong-Yong Chen, and Chen –Yuan Wu “ An Unsupervised Approach for Person Bipolarization using Principal Component Analysis”, IEEE Transactions on Knowledge and Data Engineering,Volume:24, Issue: 11 Pages: 1963 - 1976 ,November 2012
[8] Yinghui Yang and Balaji Padmanabhan” GHIC: A Hierarchical Pattern-Based Clustering Algorithm for Grouping Web Transactions” IEEE Transactions On Knowledge And Data Engineering, Vol. 17, No. 9, September 2005
[9] A. C. M. Fong, Baoyao Zhou, Siu C. Hui, Jie Tang and Guan Y. Hong “Generation of Personalized Ontology based on Consumer Emotion and Behavior Analysis”, IEEE transaction on Affective Computing,Volume:5, Issue: 2,Pages: 152 - 164, IEEE 2012
[10] Ruili Geng and Jeff Tian “Improving Web Navigation Usability by Comparing Actual and Anticipated Usage”, IEEE Transactions on Human-Machine Systems, Volume: 45, Issue: 1,Pages: 84 – 94, IEEE 2014.
[11] Jun Liu, Member, IEEE, Cheng Fang, and Nirwan Ansari, Fellow, IEEE “Request Dependency Graph: A Model for Web Usage Mining in Large-scale Web of Things” IEEE Internet of Things Journal, Volume: PP, Issue: 99,Pages: 1 – 1, IEEE 2015
[12] Mohamed Koutheûr Khribi, Mohamed Jemni and Olfa Nasraoui “Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval” Advanced Learning Technologies, 2008, ICALT ’08,Eighth IEEE International Conference 2008,Pages: 241 – 245.

[13] Steve G. Romaniuk “Using intelligent agents to identify missing and exploited children” IEEE Intelligent Agent, Volume: 15, Issue: 2, Pages: 27 - 30, March/April 2000

[14] Ed H. Chi “Improving Web Usability through Visualization” IEEE Internet Computing, Volume: 6, Issue: 2, Pages: 64 - 71, April 2002

[15] Mamata Jenamani, Pratap K.J. Mohapatra, and Sujoy Ghose “Online Customized Index Synthesis in Commercial Web Sites” IEEE Intelligent Systems, Volume: 17, Issue: 6, Pages: 20 -26, November/December 2002

[16] Sung Ho Ha “Helping Online Customers Decide through Web Personalization” IEEE Intelligent Systems, Volume: 17, Issue: 6, Pages: 34 - 43, November/December 2002

[17] Sung Min Bae, Sang Chan Park and Sung Ho Ha “Fuzzy Web Ad Selector Based on Web Usage Mining” IEEE Intelligent Systems, Volume: 18, Issue: 6, Pages: 62 - 69, November/December 2003

[18] Anoop M. Nambooti and Anil K. Jain “Online Handwritten Script Recognition” IEEE Transactions On Pattern Analysis And Machine Intelligence, Volume: 26, Issue: 1, Pages: 124 - 130, January 2004

[19] Doru Tanasa and Brigitte Trousse “Advanced Data Preprocessing for Intersites Web Usage Mining” IEEE Intelligent Systems, Volume: 19, Issue: 2, Pages: 59 - 65, March/April 2004

[20] Jiming Liu, Shiwu Zhang and Jie Yang “Characterizing Web Usage Regularities with Information Foraging Agents” IEEE Transactions on Knowledge and Data Engineering, Volume: 16, Issue: 5, Pages: 566 – 584, May 2004

[21] Pranam Kolari And Anupam Joshi “Web mining: Research and Practice” the IEEE CS and AIP, Volume: 6, Issue: 4, Pages: 49 - 53, July/August 2004

[22] Sung Ho Ha” Helping Online Customers Decide through Web Personalization” IEEE Intelligent Systems , Volume : 17, Issue: 6, Pages: 34 - 43, November/December 2002

[23] M. Jordan Raddick, Ani R. Thakar, and Alexander S. Szalay “Ten Years of SkyServer I: Tracking Web and SQL e-Science Usage” IEEE CS and the AIP, Volume: 16, Issue: 4, Pages: 22 - 31, July/August 2014

[24] Jiliang Tang and Huan Liu “An Unsupervised Feature Selection Framework for Social Media Data” IEEE Transactions On Knowledge And Data Engineering, Volume: 26, Issue: 12, Pages: 2914 – 2927, December 2014

[25] Mehdi Adda and Petko Valtchev” Toward Recommendation Based on Ontology Powered Web-Usage Mining” IEEE Computer Society, Volume : 11, Issue: 4, Pages: 45 - 52, July/August 2007

[26] Dimitrios Pierrakos, Member, IEEE, and Georgios Paliouras “Personalizing Web Directories with the Aid of Web Usage Data” IEEE Transactions On Knowledge And Data Engineering, Volume: 22, Issue: 9, Pages: 1331 - 1344, September 2010

[27] Wu,K.-L.; Yu,P.S.; ” SpeedTracer : A Web usage mining and analysis tool” IBM system journal, Volume: 37, Issue: 1, Pages: 89 – 105,IEEE1998.

[28] https://rapidminer.com/

[29] http://www.cs.waikato.ac.nz/ml/weka/

[30] http://orange.biolab.si/