Optimized Web Mining Technique for Adaptive E-learning Site: A Case Study

Renuka Mahajan, J S Sodhi
Amity University, Noida
AKC Datasystems, Delhi
India
vishalk.mahajan74@gmail.com, jssodhi@akcgroup.com

ABSTRACT: *An important application of web usage mining is mining web log data, where the sequences of web pages accessed by various web users, over a period of time, are recorded on the web server. We propose a new optimized technique in realm of an e-learning site that pre-processes the web log data to recommend the best links for a learner to visit next. We propose a novel methodology, by partitioning the database, on the basis of the learner's knowledge level, to create a specialized suffix tree(s) from the existing sequences of previous 'n' learners' path. Further to reduce the overhead of re-mining the web patterns from the whole web data, we propose that a web traversal pattern should be regarded significant, only if it qualifies the minimum threshold of length and frequency in the database. These significant patterns are added to generalize suffixes. These are then mined, using the most efficient mining algorithm after a comparative analysis of various algorithms, to find the most frequent navigation paths for recommendation to new learner. We conducted experiments in web log mining on a real case study of an Indian e-learning site. The proposed methodology is verified by experiments with promising results on computational time. This speed up obtained, in Web Pattern Mining, is a meaningful approach for building recommender based e-learning system, to predict the future learning paths.*

Keywords: Web Mining, Personalization, Suffix Tree, PL WAP, GSP, FP Growth WAP Mine, Navigation Prediction

Received: 19 May 2015, Revised 24 June 2015, Accepted 30 June 2015

1. Introduction

Web-based educational systems have a lot of information recorded in log files, for example, interactions between students and the online learning systems, details of student successes and failures, student grades, and knowledge levels etc. Analyzing the server logs and the history list can help in understanding the user behaviour and the web structure, thereby improving the design of the website. Web mining [30] is the application of data mining techniques to extract knowledge from web data. According to [1,4,13], web mining is a process of analyzing the data in the web pages or data related to web activity (stored as web log data) and then generating useful and meaningful patterns. Oren Etzioni, was the first person to propose the term of Web mining in his paper [4]. The Web Mining is divided into Web Content Mining, Web Structure Mining, and Web Usage Mining (WUM). Recent years have seen major growth of research in the area of Web Usage Mining. In WUM, Frequent Pattern Mining (FPM) plays an important role for decision making in many real world applications (Xu) such as:

- In e-learning, frequent pattern mining can help educators in analyzing learner behaviour and building recommender system [11].

• In e-business, in discovering customer behaviour and trends, deciding discounts policies or advertisements planning, prediction of potential market opportunities [3] and risk analysis.

• In healthcare, mining Frequent Patterns for decision making on treatments, discovering action of genes [10, 16] and preventing outbreak of diseases.

• In disaster prevention, mining frequent patterns in forecasting of weather by analyzing various environmental factors [19].

• In automobile industry, frequent pattern mining of customers' preferences can help engineers design vehicles.

• In crime detection, frequent pattern mining helps in investigations and in improving public security [2].

In this paper, our aim is to optimize the traversal path pattern mining for navigation prediction in an e-learning case study.

2. Related Work (Web Usage Mining in e-learning)

[32] Explores how learners' behavior can be analyzed in web based learning system which is currently quite prevalent and useful in e-commerce. [33] Proposed a data mining model that captures the user navigation behaviour patterns. [34] Presented an approach to adaptive hypermedia learning environments design and illustrated how learners' knowledge level and individual traits can be exploited to guide the adaptive dimension of a hypermedia system. [36] Proposed an evolving Web-based learning system which can adapt itself to its users, in response to the usage of its learning materials. [37] Proposed a sequential mining algorithm, to analyze learning behaviours. [35] Analysed the sequences patterns of students' web usage after the analysis of log files data.

One of their conclusions was that there still does not exist, a standardized methodology for applying web mining techniques in this field. Hence, this gave us the motivation to research on a new approach by optimizing the use of existing techniques.

The goal of an adaptive e-Learning site from user's point is that the website could intelligently recommend resources and content that would improvise learning outcome based on the cluster learner belongs to. This will help user to traverse less links and use reduced clicks to study the relevant topics/material.

The goals of an adaptive e-Learning site from site administrators' view point is, s/he should be able to categorize a user (novice, intermediate and advance) and then, s/he should be able to help new users to better understand the previous learners' navigation patterns. It will help administrator to design intelligent course planning based on personalized cognitive patterns.

The goals of an adaptive e-Learning site from content and software developer is to track the content/links used by previous users, understand the preferences of the previous site's user's based on the category and then organize the content/links accordingly for the next user based on the category s/he belongs to, thereby making the web application more effective for new learners.

E-learning is powerful, as it allows individuals to learn 'anywhere, anytime' and gives instant access to specific knowledge. But according to [25], in e- Learning, learners face many challenges such as lack of flexibility of the site, lack of adaptability towards learners' needs, lack of effective design of electronic content.. This lack of adaptive learning environments or an environment with adaptive features is often due to the concept "one-size-fits-all". Hence the problem is that e-learning websites cannot teach learners in accordance to their aptitude and provide adaptive material. Moreover, information on the site is often not effectively organized. Hence while navigating; learners often tend to lose their basic aim of inquiry.

An e-learning system is considered to be adaptive [26] "*if it is capable of: monitoring the activities of its users; inferring these preferences out of the interpreted activities, appropriately representing these in associated models; and, finally, acting upon the available knowledge on its learners to facilitate the learning process*". The main benefit of the adaptive presentation is that it reduces the amount of presented information with simplified convenient link structure to the most relevant information for a particular user, solving the "*information overload*" and "*lost in hyperspace*" problem of the e-learning site.

As explained by [26], there are two basic questions in adapting an e-learning site.

a) *Adapting to what*?" - I.e. what aspects of a learner interacting with the system can be taken into account when providing adaptation? The answer includes several learner characteristics, like knowledge, goals, interest, background, learning style etc. The focus of this paper is on the 'knowledge level' as the adaptation criterion.

b) *What can be adapted*?" - According to Author [31] two types of adaptations can be adaptive presentation (content) level and adaptive navigation (link) level. The content level adaptation is used to solve the problem of hypermedia systems which are used by different categories of learners, while link-level adaptation is used to provide some kind of navigation support to prevent learners from getting lost in hyperspace.

According to Author for [38], the e-content can be unclear for a learner who is a beginner level and trivial or boring for an advanced level learner. Also, beginners have almost no knowledge about the educational material available online. Hence they need navigational help to find their way through the web site.

The paper focuses on adapting the actual content, specifically the navigation links.

We conducted experiments using data from an e-Learning portal. Web mining in our case is done on the learners' most potentially visited pages by referring to the visiting histories of other learners, who exhibit similar navigation preferences.
Our research contributions in this paper are:

- We propose the best way to adapt the e-learning by generating a new technique for discovering frequent traversal path patterns from web data.
- We evaluated the computational time of the proposed methodology, experimenting with real-world data and made comparisons with previous work.

We organize our paper as follows. We described our methodology in section 2. We presented the results of the experiments in Section 3. We presented our discussions and conclusions in Section 4 and future area of research work in Section 5.

3. Methodology

Once the course is undertaken, the cumulative number of web log data for each learner was downloaded and compiled for all learner clicks.

Population and Sampling: The sample of 3561 records was selected by random sampling from a population of around 5000 records in the secondary data, of actual web log records of the 'learn-next' portal in December 2012. This study is based on a very good data-set. It is about online courses with large cohorts of students throughout various year groups and subject matters.

Measurement Instruments Design and Development: Each hit against the server, corresponding to an HTTP request, generates an entry in the access log. Each log entry contained fields identifying the client IP address, the user ID, the request time, HTTP status code, the URL requested etc. Data Preparation and Pre-processing were applied to retrieve this suitable data set from raw web log records, to which various pattern mining algorithms could be applied. We removed non useful information (e.g. Total number of bytes transferred, version of the HTTP protocol used) from the web log records. This formed the final input data for the WUM algorithms.

The process for the optimized tool is a series of steps as under:-

3.1 Learner Clustering

To adapt an e-learning site for the learners, we need to understand how the e-content was accessed by learners having different knowledge levels. As in [3, 16], the authors proposed the influence of 'knowledge level' as major adaptation consideration. Authors [18] proposed assessment marks as the assessment method for determining the knowledge of the user. Web mining on the existing web log data revealed a number of patterns as given in [24]. Past researches indicate that there exist a number of parameters on which individuals may differ. This affects the way in which they interact with an e-learning site [25]. For the content to be maximally effective, it should capitalize on such learner characteristics while delivering personalized e-content. Many personalization parameters are researched in E-learning scenarios [26]. An important personalization parameter of e-learning scenario is their present knowledge level [26]. Most of the systems use this personalization parameter i.e. learner's level of knowledge. It uses the set of values such as novice, intermediate and advanced. Learner's knowledge level, based on their test scores, can be updated into the system and used in the adaptation process (as done by [27]).

We propose the construction of adaptive learning environment by categorizing the learners into 3 categories based on their knowledge levels (found by their test scores) to find relevant subject contents. In e-learning, clustering can be used for finding clusters of students with similar test scores. For this objective, considering their assessment marks as in [17], the learners are categorized as novices, intermediate and advanced using K-means Cluster analysis in SPSS 16.

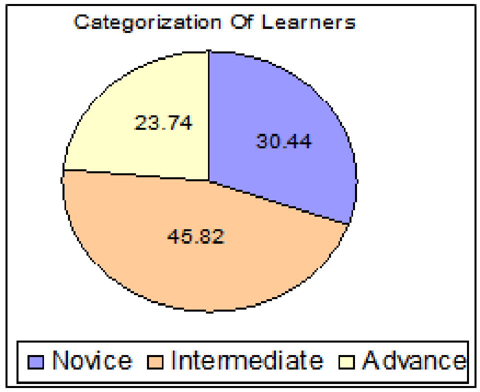


Figure 1. Learner Cluster Analysis

3.2 Suffix Tree Generation

Suffix trees given by [39] are data structures used in many data mining applications [21]. The suffix tree is a data structure that presents the suffixes of a given string in a way that allows for a particularly fast implementation of many important string operations like pattern finding. The suffix tree for a string S is a tree whose edges are labelled with strings, such that each suffix of S corresponds to exactly one path from the tree's root to a leaf. The suffix tree for the string S of length n is defined as a tree such that:

a) The paths from the root to the leaves have a one-to-one relationship with the suffixes of.

b) Edges spell non-empty strings,

c) And all internal nodes (except perhaps the root) have at least two children.

A Generalized Suffix Tree (GST) is a suffix tree made for a set of words instead of only for a single word. It represents all suffixes from this set of words. Each word must be terminated by a different termination symbol or word. [40] Studied the application of generalized suffix trees to mining frequent user access patterns from Web logs.

We extend this generalized suffix tree by including only significant pattern (qualifying minimum frequency as well as minimum length values) to build an optimized data structure to be mined for frequent web patterns. This minimum length parameter is to be decided by the level at which recommendations are required e.g. if navigation prediction is required at the ith level of the recommendation system, where $i = 4$, then minimum length i.e. $i - 1 = 3$ should be included as qualifying length criteria. Any pattern with length < i would be useless in predicting recommendations at ith level. We use disjoint generalized suffix trees (having two parameters of minimum length and frequency of a pattern) pertaining to three different clusters of learners thereby reducing the overhead of mining.

3.3 Comparative Analysis of FP Mining Algorithms

Many algorithms have been introduced in the area of frequent pattern mining in the last decade. Algorithms such as Apriori Algorithm, GSP (generalized sequential pattern mining), and FP (frequent pattern)-Growth, WAP (Web Access Pattern)-Mine and PL (Pre-order Linked)WAP are most popular and efficient algorithms being used for generating various patterns.

A comparison of above mentioned algorithms in the realm of adaptive e-Learning is done and an algorithm that suits best and generates frequent patterns (of topics referred by various learners) efficiently and effectively is used. A comparative study by [9] gives the advantages and drawbacks of all the algorithms. The most suitable algorithm is be used for generating frequent usage pattern and this would be helpful in suggesting recommendations to a new learner(s) in e-learning domain. Our objective is to improve the computation time in generating frequent patterns while maintaining the accuracy. As in [11], all the algorithms mainly differ in two ways:

1) The way in which candidate sequences are generated and stored. Our main aim here is to reduce the number of candidate sequences generated.

2) The way in which support is counted and how candidate sequences are tested for frequency. The key strategy here is to eliminate any database/ data structure that has to be maintained throughout, for counting purpose.

The GSP [15] (Generalized Sequential Pattern) algorithm was found to be one of the promising candidates in our case study for generating the required frequent web access sequences. It can accomplish the task in lesser time compared to all other algorithms.

Consider the example of a web database D, consisting of 8 transactions of web access sequences. Suppose minimum support count required is = 75 % and minimum length required is > 3.

Further, using the most efficient frequent pattern mining algorithm from comparative analysis of various algorithms, we can find a meaningful frequent hierarchy of navigation paths (10->20->30->40) taken by these learners in less time compared to existing algorithms. As a result, it is possible to identify, which links are used the most and least at different learner levels.

Similarly we can also apply various web usage mining techniques to predict interesting traversal patterns in terms of i) most frequent usage ii) assessment marks iii) average time spent.

This analysis helps in arriving at the most/least preferred sequence of topics by previous learners.

SID	Web access Sequences	Frequent sequence
100	Abdac	Abac
200	Eaebcac	Abcac
300	Babfaec	Babac
400	Babfaec	Abacc
500	Abd	—
600	Ab	—
700	Eca	—
800	Cab	—

Table . 1 Sample Web Access Sequence Dataset D

SID	Length	Events			
212	4	4	7	250	423
214	4	3	7	511	276
215	4	4	7	250	276
217	3	4	7	250	

Table 2. A small sample Input of threshold length filtered Web log Database D2 of the actual e-learning site

The first scan of web database D, we filter the records on the basis of parameter minimum length =3 and derive a reduced database, resulting in new database D' of 4 records. D' is further partitioned into 3 disjoint subsets D1, D2 and D3, mapping records of Advanced Users, Intermediate Users and Novice Users. Now we pre-process it to create disjoint generalized tree(s) from the existing sequences in the database. We extend the generalized suffix tree by including additional information on suffix the nodes. We store at each node the frequency as well as length. We then use the sequential mining using various mining algorithm on this optimized reduced database subsets D1, D2 and D3. The key idea is that we generate as few candidates as possible while maintaining completeness.

4	7	250
7	250	
4	250	
250		
4	7	
7		
4		

Table 3. Output of optimized technique implementations

3.4 Proposed Design and Methodology

The main concept for significance used is that a, pattern p is significant if it satisfies two constraints

a) The length of p > a given length threshold: a significant pattern is sufficiently long enough to carry important pattern information.

b) The frequency of p >= a given support score: a significant pattern must be frequent to be considered.

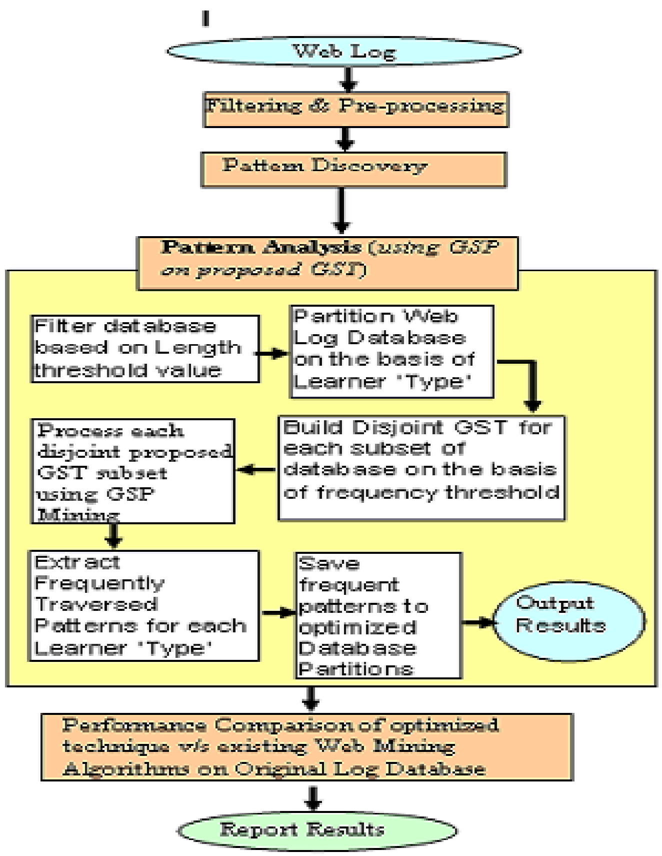


Figure 2. Flow Graph for Mining and Mapping the Web Data from real world data set

For a given length threshold ,to find the most frequent pattern for arriving at a particular node 'X' in a log database 'D', we use the concept of generalized suffix tree.

1) Before constructing the suffix tree, we filter the records in the database for a minimum length threshold. The generalized suffix tree would be built, if we consider only the records containing a particular minimal length and removing remaining records from database.

2) We partition the log database 'D' into three independent subsets, D1, D2, D3 on the basis of Learner 'Type'

3) Then we build Disjoint Generalized Suffix tree (GST) for each subset of database for particular learner type.

4) While constructing the suffix tree, at each node, for the corresponding pattern, store the frequency (i.e. the number of occurrences of pattern in the database): we increment the frequency by one at each visited node during the addition of a new suffix. A node is considered frequent, if it has frequency> min_freq .

5) We hypothesize, that a node is considered frequent, if it has frequency> min_sup and length > min_length.

6) For each web record, we keep adding a node to the temporary GST data structure, only if it qualifies minimum threshold values for length and frequency.

7) Thus, we obtain reduced database, for each learner 'type'.

8) Process each Disjoint GST subset using GSP Mining for pruning of our tree.

9) Extract Frequently Traversed Patterns for each Learner 'Type'.

10) Save these most frequent patterns that have highest likelihood to a new learner, to optimized new database partitions.

11) Performance Comparison with existing Web Mining Algorithms on Original Log Database.

12) In order to evaluate the effectiveness of proposed methodology and to evaluate the discovered learner access patterns, we conducted experiments on real world data sets and made comparisons with remaining algorithm e.g apriori-based GSP, pattern growth based WAP mine and early pruning based PL WAP algorithm.

4. Experimental Results

This section presents a set of experiments to compare the performance of the three algorithms (Apriori –based, Pattern growth and Early Pruning) in the context of the e-learning recommender system with respect to the optimized tool developed and evaluate the execution time at different support by experimentation. For this we use the same database from the web log of an e-learning site as mentioned earlier. This secondary data is the web usage data from an e-learning portal called learn-next (for class VIth -XIIth), and is collected automatically by the Web servers . It represents the navigational patterns of e-learners of this e-learning site.

The algorithms from all three categories of Apriori-based, Pattern growth and Early pruning i.e. GSP, WAP mine and PL WAP were implemented in C++ on the original database as in [8] and then compared with the new proposed technique(i.e. the optimized tool mentioned in section 2).

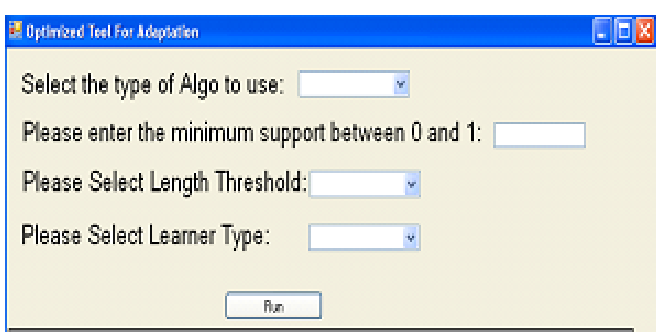
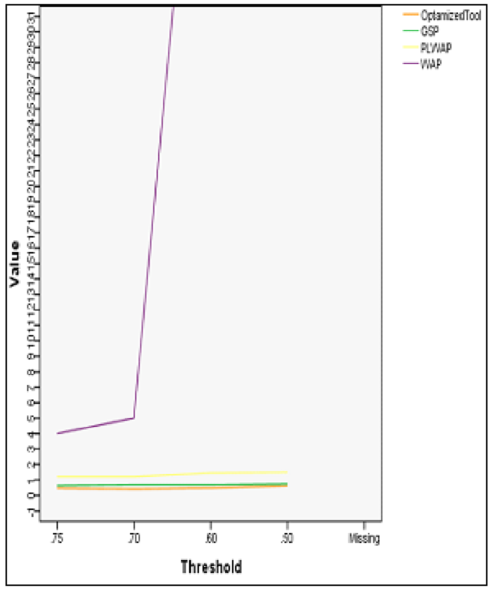


Figure 3. New Optimized frequent pattern mining Tool



Graph 1. Performance comparison graph at different support (for GSP, PLWAP and WAP mine and proposed technique)

2	5	2	7	368	270
7	5	5	7	79	62
8	5	2	5	250	273
10	5	1	7	191	360
14	5	3	7	197	144
15	5	4	5	359	195
17	5	3	7	121	98
20	5	1	7	338	216
22	5	3	5	285	189
27	5	3	7	64	184
31	5	4	7	549	183
40	5	3	5	452	318
41	5	3	7	324	425
46	5	2	5	250	181
54	5	1	7	104	377
56	5	1	5	261	127
62	5	4	5	359	333
66	5	2	5	486	193
69	5	2	7	229	390
76	5	3	5	285	20
80	5	5	7	403	255
83	5	3	7	511	357
90	5	1	7	502	317
94	5	1	7	557	161
100	5	1	7	151	123
104	5	2	5	125	106
106	5	5	5	455	345
115	5	1	5	261	200
118	5	5	7	267	110
119	5	4	5	288	190
120	5	3	7	511	56
123	5	4	5	531	99
131	5	4	7	335	191
134	5	2	5	23	406

Figure 4. Sample Input

Threshold	Optimized Tool	GSP	PLWAP	WAP
0.75	0.469	0.641	1.219	4
0.7	0.422	0.688	1.219	5
0.6	0.484	0.688	1.437	57
0.5	0.625	0.735	1.484	72
0.4	0.719	1.047	1.703	139

Figure 5. Comparative Execution times at different Support levels

5	6	
6		
5	1	
1		
5	1	5
1	5	
5	5	5
5	5	
5		
5	7	
7		

Figure 6. Sample Output

As shown in the figure 5, out of the three existing web mining algorithms, the GSP approach outperforms two other web mining approaches. We also observe that the WAP's performance execution time was the maximum, and is proved from the experimental result shown in graph. The results were quite encouraging for the optimized tool used (using the proposed methodology) as shown in the figure 5.

The results are verified using k-fold cross validation with $k = 3$.The original sample is partitioned into $K = 3$ sub samples. Of the 3 sub samples, a single sub sample is retained as the validation data for testing the model. The remaining (3-1) sub samples are used as the training data. The inference from the graph is also shown.

4. Discussion and Conclusions

Based on our experiment on the e-learning portal, it has been observed that the performance of the proposed optimized tool is the best among the mining algorithms being tested, because it eliminates the infrequent patterns based on the node length recommendation and enable algorithm to operate on smaller partitioned search space. It thus prunes uninteresting patterns and hence uses less memory which enables it to optimize the execution time.

The GSP is the next most efficient when used on mining short patterns. This is because the GSP iteratively scans database depending on the length of longest frequent sequence. Hence the execution time was less in our case, when the sample dataset contained small frequent sequences [23].The experiment confirms that WAP algorithm takes maximum execution time and it further deteriorates as minimum threshold value is decreased, means as sequences grow larger, the performance of WAP algorithm decreases.

In our study, we have proposed and developed a methodology to optimize the computational time, while maintaining the accuracy of pattern mining. The new optimized tool creates a suggestion mechanism by suggesting shortcuts to frequently visited pages based on previous learner activities. Thus, layout of course can be reorganized to better suit learner's needs. It suggests topics to improve the performance based on online assessment results.

This extended tree can be created and fetched into the memory. It is used to mine significant patterns which are both sufficiently long as well as appears frequently enough in a database over user defined threshold values. Results of experimentation show performance gains of this technique on optimized data structure, which is processed and interpreted usefully.

5. Future Work

Future research work includes direct data entry into a database, so that data may be examined by mining algorithms in real time. The set of frequent patterns derived by most of the current pattern mining methods is still huge for effective usage. Much research is still needed to substantially reduce the size of derived pattern sets and enhance the quality of retained patterns with better visualizations.

This information on actual web usage of a learner can help in adapting a website to suit another similar potential learner. This, in turn, will enable an e-learning site to reorganize the site content more efficiently or to recommend topics.

Acknowledgments

The authors' wishes to thank the EVP - Learn Next e-learning portal for the support in providing us real time data in undertaking the research work.

References

[1] Anand, S., Buchner, A. (1998). Decision Support Using Data Mining. *Financial Times Management*, 184.

[2] Bagui, S. (2006). An approach to mining crime patterns. *International Journal of Data Warehousing and Mining*, 50-80.

[3] Botsios, S., Georgiou, D. (2008). Recent Adaptive E-Learning Contributions towards a Standard Ready Architecture, *In*: Proc. of IADIS International Conference e-Learning, 226-230.

[4] Chen, H., Zhang, Y., Houston A. L. (1998). Semanticindexing and searching using a Hopfield net, *Journal of Information Science*; 24. 3-18

[5]Chou, P. B., Grossman, E., Gunopulos, D., Kamesam, P. (2000). Knowledge Discovery and Data Mining. *In* : Proceedings of Sixth ACM SIGKDD International Conference, 447-456.

[6] Etzioni, O. (1996). The World Wide Web: Quagmire or gold mine. Communications of the ACM, 39 (11), 65-68.

[7] EZEIFE, C. I. (2005). Mining Web Log Sequential Patterns with Position Coded Pre-Order Linked WAP-Tree. *Data Mining and Knowledge Discovery*, 5-38.

[8] Ezeife, C. I. (2005). PLWAP sequential mining: open source code. 1st International Workshop on Open Source Data Mining: *Frequent Pattern Mining Implementations*, 26-35.

[9] Ezeife, C. L. (2003). Position coded pre-order linked WAP-tree for web log sequential pattern mining. *In*: Proceedings of the 7th Pacific-Asia conference on Advances in knowledge discovery and data mining, 337-349.

[10] Lu, Y. (2005). PLWAP sequential mining: open source code. Proceedings of the 1st international workshop on open source data mining: *frequent pattern mining implementations*, (p. 26-35).

[11] MABROUKEH, N. R., EZEIFE, C. I. (2010). Taxonomy of Sequential Pattern Mining Algorithms, University of Windsor, *ACM Computing Surveys*, 43 (1), Article 3.

[12] Oyama, T., Kitano, K., Satou , K., Ito, T. (2000). Mining Association Rules related to protein protein interactions, *Genome Informatics* 11, (358–359).

[13] Cooley , R, . Mobasher , B., Srivastava, J. (1997). Web Mining: Information and Pattern Discovery on the World Wide Web, *IEEE*, 558-567

[14] Mahajan,R., J. S. (2012). Mining User Access Patterns Efficiently for Adaptive e-Learning Environment. International *Journal of e-Education, e-Business, e-Management and e-Learning*, 277-279.

[15] Ramakrishnan Srikant, R. (1996). Agarwal. Mining Sequential Patterns: Generalizations and Performance Improvements, *Springer Advances in Database Technology* — EDBT '96, 1057, 1-17.

[16] Sfenrianto, Hasibuan, Z. A., Suhartanto, H. (2011). The Influence Factors of Inherent Structure in e-Learning Process. *International Journal of e-Education, e-Business, e-Management and e-Learning,* 1 (3), 217-222.

[17] Esichaikul, V., Lamnoi, S. (2011). C Bechter. Student Modelling in Adaptive E-Learning Systems, *Knowledge Management & E-Learning. An International Journal (KM&EL)*, 3 (3), 342-355.

[18] Watson, J., Ahmed, P. K., Hardaker, G. (2007).. Creating domain independent adaptive e-learning systems using the sharable content object reference model, Emerald Group Publishing Limited, *Campus-Wide Information Systems*, 24 Iss: 145 – 71.

[19] Wetjen, T. (2002). Discovery of frequent gene patterns in microbial genomes. University of Bremen, Germany. TZI report, Technologie Zentrum InofrmatiK.

[20] Xu, L. (2011). A Study of Frequent Pattern Mining in Transaction Datasets.

[21] Chen, Zhixiang, Fowler, R. H., Fu., A W C., Wang. C. (n.d.). Efficient web mining for traversal path patterns.

[22] Zhang, Zhongnan.,W. W. (2004). Mining dynamic interdimension association rules for local scale weather prediction. *Twenty-eighth Annual International Computer software and Applications Conference*, 146-149.

[23] Pei, Jian J., Han, Asl, B M ., Wang, J., Pinto, H., Chen, Q ., Dayal, U., Hsu, M. C. (2004). Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach, *IEEE Transactions on Knowledge and Data Engineering*, 16 (10).

[24] Mahajan, R., Sodhi, J. S., Mahajan. V. (2014). Usage patterns discovery from a web log in an Indian e-learning site,, A case study, Springer, *Education and Information Technologies*, 19, 1-26.

[25] Mitchell, T. (2005). Chen. Hypermedia learning and prior knowledge: domain expertise vs. system expertise, *Journal of Computer Assisted Learning*.

[26] Goyal, M., Yadav, D., Choubey, A. (2012). E-learning: Current State of Art and Future Prospects, *International Journal of Computer Science Issues*, 9 (3), (2), May .

[27] Giridharan, A. (2005). Adaptive e-Learning Environment for Students with Divergent Knowledge Levels, ELELTECH 2005, Hyderabad.

[28] Radenkovi, B., Despotovi, M,. Bogdanovi, Z., Bara, D. (2006). Creating Adaptive Environment for e-Learning Courses, *JIOS*, 33 (1).

[29] Popescu, E. (2008). DYNAMIC ADAPTIVE HYPERMEDIA SYSTEMS FOR E-LEARNING, University of Craiova, Romania.

[30] Srivastava, J., Cooley, R., Deshpande, M., Tan, P. (2000). Web usage mining: Discovery and applications of usage patterns from web data, in SIGKDD Explorations, 1 (2), 12–23.

[31] Brusilovsky, P. (1996) , Methods and techniques of adaptive hypermedia , User Modeling and User Adapted Interaction, (Special issue on adaptive hypertext and hypermedia), 6 (2-3), 87-129.

[32] Osmar, R. Zaiane. (2001), Web Usage Mining for a Better Web-Based Learning Environment (2001), Technical Report TR01-05, Department of Computing Science, University of Alberta.

[33] Borges, Jose., Mark Levene. (2000). Data Mining of User Navigation Patterns, Springer Web Usage Analysis and User Profiling , *Lecture Notes in Computer Science* 1836, 92-112.

[34] Papanikolaou, K., Grigoriadou, M., Magoulas, G. D., Kornilakis, H. (2002). Towards New Forms of Knowledge Communication: the Adaptive Dimension of a Web-based Learning Environment. *Computers and Education*, 39 (4), 333-360.

[35] Cristóbal Romero, Sebastián Ventura, Enrique García, Data mining in course management systems: Moodle case study and tutorial, Elsevier, *Computer and Education* 51 (1), 368-384.

[36] Ya TANG, Tiffany., MCCALLA, Gordon (2003) Smart Recommendation for an Evolving E-Learning System Workshop on Technologies for Electronic Documents for Supporting Learning, International Conference on Artificial Intelligence in Education (AIED)

[37] Zhou, Baoyao., Hui , Siu Cheung., Cheuk, Alvis., Fong, Ming (2006). Efficient sequential access pattern mining for web recommendations, *International Journal of Knowledge-based and Intelligent Engineering Systems*, 10 (2), 155-168.

[38] Czaja, SJ, Hammond, K, Blascovich, J.J, Swede, H. (1989). Age related differences in learning to use a text-editing system. *Behaviour and Information Technology*. 8 (4), 309–319.

[39] Ukkonen, E. (1995). On-line construction of suffix trees. Algorithmica. 1995, 14 (3), 249–260.

[40] Xiao, Y., Dunham, M. (2001). Efficient mining of traversal patterns. *Data & Knowledge Engineering*, 39, 191-214.