International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 3, Issue 1 (Jan-Feb 2015), PP. 104-108

COMBI: A NEW COMBINATION OF WEB MINING TECHNIQUES FOR E- LEARNING PERSONALIZATION

Rana Khudhair Abbas Ahmed
Al-Rafidain University College/ Computer Technical Engineering Department
rana_ruc@yahoo.com

Abstract- With the help of web mining technology, the importance of e-learning is in advance. The web mining gives the personalized education. By using various mining techniques, both the learner and the teacher can achieve benefits. Thus, the process of learning becomes simpler and affordable. With the advancement in web mining, the education through eLearning is improved.

In this paper, we proposed "Combi", a new system for combining three web mining techniques with the learner's profile for e-learning personalization. Our personalization is based on learners' preferences, educational background and experience. The goal we took of personalization is learner's satisfaction. Our Combi system helps the learner to find the most suitable related information needed without wasting time and effort in looking through the long list of searched results to find the most suitable information. We discovered the learning behavior pattern to build up a series of feedback and motivation system and we compared the results from our system with an existing related system.

Keywords: System, architecture, personalized, e-learning, web mining.

I. INTRODUCTION

E-Learning is a general term used to refer to a form of learning in which the instructor and learner are separated by space or time where the gap between the two is bridged through the use of online technologies. With web-based learning, it is possible for the learners to learn from anywhere, anytime, at their pace. Web-based learning brings unprecedented level of accessibility to courses in remote area, courses prohibited by budget constraints, courses updated to recently discovered knowledge, qualified instructors, and instruction at any time [1].

Today, E-learning has emerged as a new alternative to conventional learning to achieve the goal of education for all. The concept E-learning has numerous definitions and sometimes confusing interpretations. In our purpose we adopt a definition of E-learning as the use of Internet technologies to provide and enhance students' learning anytime and anywhere. One of its advantages is the learning method which can be more adaptive than conventional learning. In response to individual needs, personalization in education not only facilitates students to learn better by using different ways to create various learning experiences, but also teachers' needs in preparing and designing varied teaching or instructional packages [2].

A key requirement of the contemporary eLearning systems is the personalization that is a function able to adapt the eLearning content or services to the user profile. The personalization include how to find and filter the learning information that fits the user preferences and needs, how to represent it and how to give the user tools to reconfiguration the systems, in consequence, reconfiguration system could be part of personalized environment in some systems [3].

E-Learning systems are grouped in the following approaches [2]:

- Personalization of the learning content, based on learners' preferences, educational background and experience.
- Personalization of the representation manner and the form of the learning content.
- Full personalization, which is a combination of the previous two types.

Personalization in e-learning can be understood as education related technology which is capable of individualizing the interaction between system and learner, based on the personal needs and preferences of learner, and help them to shape their own learning boundaries and to collaborate in terms of thoughts, information and knowledge entities [1].

II. WEB MINING AND PERSONALIZATION

Data Mining is the process of analyzing data from different perspectives and summarizing the results as useful information. Web mining is one of the important branches in data mining. Extraction of useful information or pattern from the web data is called web mining. It can be classified into three following categories as shown in figure (1) [4]:

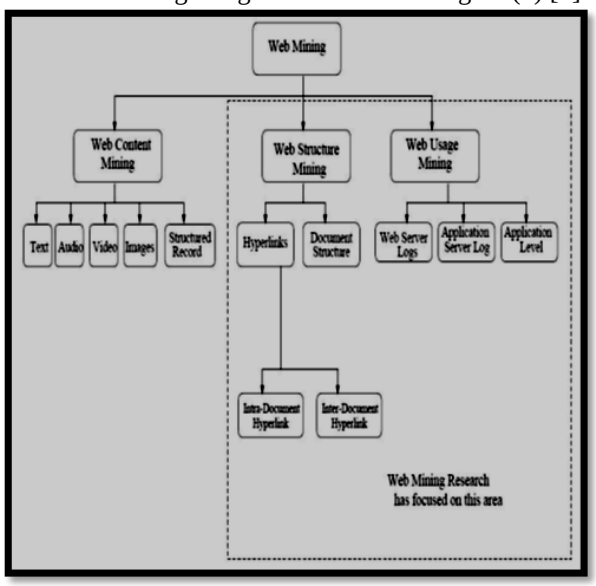

Fig (1): Structure of Web mining [4].

- Web structure mining
- Web content mining
- Web usage mining
- Web Structure mining:

Web structure mining targets on analysis of the web and one of its uses is to identify more preferable documents. It

helps to discover similarities between web sites or discovering significant sites for a specific topic or branch or in discovering web communities. It is also used to reveal the schema of web pages [4].

- Web Content mining:

Web content mining refers to the extraction of useful information from the Web document's content. The content of the web page consist of text, images, audio, video etc., it includes the techniques like clustering or associating web pages according to the respective branches. It also helps in discovering patterns in web pages to mine useful data [4].

- Web Usage mining:

The web log files can be generated when the user visits the website. The extraction of information from web log files is known as the web usage mining. The Usage data captures the identity of Web users along with their browsing behavior at a Web site. In general this type of mining includes several steps: data collection, data pretreatment, and knowledge discovery and pattern analysis in order to understand and better serve the needs of Web-based applications [4].

III. APPLYING WEB MINING FOR E-LEARNING

From a direct perspective, information extraction is one of the forms of learning. Search engine technologies use the web mining techniques to extract or retrieve the most pertinent and important pages. However, the contribution of Web mining has not been restricted to such explicitly available information such as page content. The learning is frequently supported with inclusion of other kinds of data such as concept pecking order on which a Web structure is based or web usage information. These kinds of data do not directly reflect the information in the page but help in building the context and circumstances in which such information is sought [4].

Web usage mining techniques can be used to discover user navigation patterns. However, it is the usage information that actually reflects how a user is navigating or learning from the Web site. Such usage information can not only serve as a useful feedback to the experts about the learners approach, but can also suggest to learners from the 'navigation experience' of other user's on what they found useful. Initial work on analyzing Web logs to discover patterns and associations between Web pages visited provided the right direction for such kind of analysis, but did not especially address the issue of eLearning [4].

These kinds of analysis can be done either offline or online, or integrating both. Web mining techniques coupled with integrated meta-information such as author info, download info, and other additional info explicitly defined by a domain expert helps to improve the learning process. Given a large, knowledge-dense website and a non-expert user seeking information, recommending relevant content becomes a significant challenge. Web mining is also referred as a useful tool for providing expert-driven recommendations to non-experts, helping them understand what they need to know, as opposed to what is popular among other users. Web Mining also has focused on modeling user navigation behavior [4].

IV. RELATED WORK

[5] presented an architecture with the use of Web mining for Web personalization. He proposed a system that provided a new approach with combination of web usage mining, HITS algorithm and web content mining. It combines hits results on user logs and web page contents with a clustering algorithm. We proposed "Combi", a new system for combining three web mining techniques with the learner's profile for e-learning personalization. This system helps the learner to find the most suitable related information needed without wasting time and effort in looking through the long list of searched results to find the most suitable information. We discovered the learning behavior pattern to build up a series of feedback and motivation system.

V. THE PROPOSED COMBI SYSTEM ARCHITECTURE

Our Combi system, saves information about the user's learning requirements without asking him explicitly. This information helps the system to make personalization of the learning content based on the learner's preferences, educational background and experience. And this information is stored in a profile.

Our Combi system combines three web mining techniques (web content mining, web structure mining, web usage mining). Figure (2), represents the structure of our Combi system.

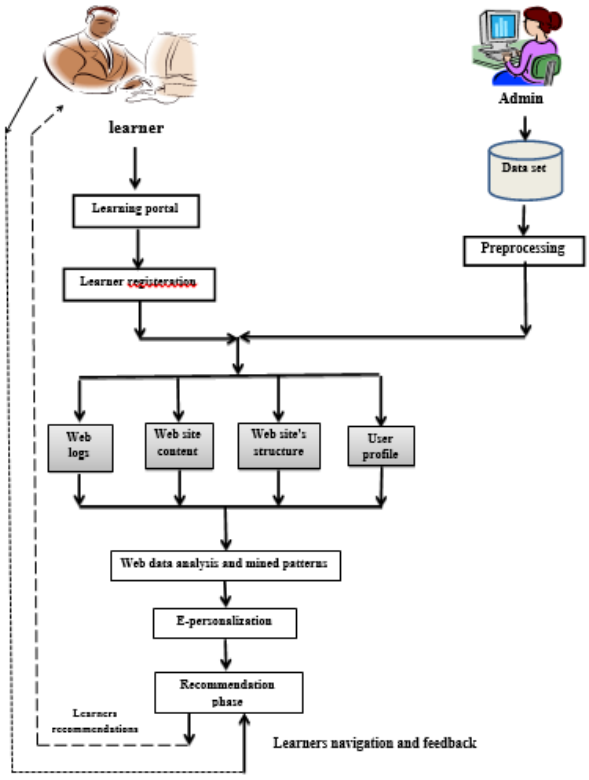

Figure (2): The architecture of the proposed Combi system.

VI. ALGORITHM OF THE PROPOSED COMBI SYSTEM

In this section, we explain the algorithm used in our Combi system for e- personalization process, as shown in figure (3).

Algorithm: e- personalization of information.
Input: search query.
Output: recommended learner's URLs and text.
Process:
Step1: begin.
Step2: calculate both the sum of hub score of page and authority score of page (to find the maximum hits).
Step3: preprocessing the dataset.
Remove unnecessary entries in the web log (image entries in the log having extensions .gif, .jpeg, etc) and remove the records that have failed HTTP status code.
Step4: apply Lingo clustering algorithm.
Step5: output clusters.
Step6: input clusters to the web mining techniques with user profile.
Obtain data from various sources (usage data, content data structure data with user data).
Step7: data analysis and mining the interesting patterns.
Step8: e-personalization (Learner and Session Identification).
Step9: recommendation phase.
mining of learners' profile, crawling and indexing of learning resources, extracting user preferences from the learner's active session, computing relevant links to recommend for the active learner and finally, give recommendation to the learner's and then waiting for his feedback.
Step9: end of algorithm.

Figure (3): Algorithm of the Combi system.

If the learner (user) is new, he first makes registration to the learning portal. The registration process saves personal information about the learner in a profile. When the learner (user) login to the Combi system using his own user name and password, he then makes his search.

Admin is responsible for forming a standard dataset including learning objects. The data can be obtained from various sources:

1) Usage Data: The log data collected automatically by Server represents the navigational behavior of learners. Each hit, represents a single entry in the log. Each log entry consists of time, date of request, IP address of the learner, the resource requested, mode of request(POST/GET), HTTP status ,no of hits, cookie etc. This data needs to be transformed and aggregated at different levels of abstraction. A session is the sequence of page views by a single user during a single visit.
2) Content Data: It includes static HTML/XML pages, multimedia files, collection of records from operational databases.
3) Structure Data: Web designer's view of content organization, within the site e.g. HTML/XML pages can be represented as a tree structure forms the structure data.
4) User Data: It includes profile data and the demographic information about registered learners, learner ratings, etc.

Web data preprocessing is the first step in Web Mining. It is necessary to convert the data to an appropriate form for solving a specific educational problem. This includes choosing what data to collect, focusing on the questions to be answered, and making sure the data align with the questions.

This was applied to retrieve this suitable data set from raw web log records, to which various data mining & statistical techniques could be applied. This dataset is preprocessed by data cleaning, learner and session identification and page view identification.

A learner profile is generated from learner related information; this includes learner's preferences, learner demographics learner knowledge, and learning styles. Our Combi system generally deals with unstructured data related to the items. This data has been processed, into relational form such as a set-of-words representation commonly used for textual data.

After data cleaning, filtering, pre-processing, integrating from multiple sources, we transformed the integrated data into a warehouse, suitable to be used as input to various data mining techniques with the learner's profile.

Our Combi system proceeds in three web mining techniques:

1) Web structure mining: This process searches for the similarities between web sites or significant sites related to the learner's search.
2) Web content mining: This process extracts useful information from the web document's content. The content of the web page consists of text, images, audio and video.
3) Web usage mining: Through this process, the server save's the learner web log files (Ip address, topic id, blog id, access time of the learner).

Interesting patterns could then be mined. For learner identification, one needs to identify who accessed web site and what pages were accessed. In session Identification, these page accesses of each learner are divided into individual learner sessions.

In the e-learning personalization process, there's much information available about not only learner's interaction rather about activities, such as reading, writing, taking assignments and communication with peers. Analyzing the server logs and the history list can help to understand the user behavior and the web structure, thereby improving the design of the website

Web mining methods were used to incorporate previous learners' feedback for making recommendations. Our Combi system uses the learner's profile to recognize student goals and predict a recommendation list.

The recommendation procedure is performed by mining of learners' profile, comparison of extracted features from unseen items with content descriptions in the user profile (content descriptions are textual features extracted from web pages), crawling and indexing of learning resources (crawling the entire learning resources available in a course repository and forming an inverted index mapping each keyword to a set of pages in which it is contained), extracting learner preferences from the learner's active session (set of URLs or list of terms extracted from these URLs), and computing relevant links to recommend for the active learner.

In this preference feedback, the learner is provided with a list of recommendations and is required to choose one of the recommendations that best suit his requirement. Our Combi system then uses this feedback to present the learner with other, similar objects. The iterations continue until the learner finds an object of interest or abandons the search. After we discovered the learning behavior pattern we then can build up a series of feedback and motivation system to reach the learner's satisfaction.

VII. RESULTS

We evaluated the effectiveness of our combi system approach and the performance is measured using two factors, Random Index and Precision (P) [5]:

The percentage of retrieved documents that is in fact relevant to the query (i.e., "correct" responses).

Precision (accuracy) = $\frac{|\{Relevant\}\cap\{Retrieved\}|}{|\{Retrieved\}|}$

Random Index (RI), the Rand index measures the percentage of decisions that are accurate [5]:

$$R1 = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

A true positive (TP) decision assigns two similar documents same cluster. True negative (TN) decision assigns two dissimilar documents to different clusters. (FP) decisions assign two dissimilar documents to same cluster. A (FN) decision assigns two similar documents to different clusters [5].

We have used ambient (ambiguous entries) dataset as standard dataset for testing and finding search results, which is the same used in the previous related system in order to compare its results with our Combi system. Table (1), shows a sample of the dataset queries and the results of the previous related system. Table (2), shows the using of the same sample of the dataset queries and the results of our Combi system.

The comparison of the results between our proposed Combi system and the existing related system are shown in table (3), (4) and (5). The performance of our Combi system compared with the previous related system (which is based only on the combination of web usage mining and web content mining), shows that our Combi system has better performance improvement as compared to the previous related existing system as shown in figure (4).

Sr. No.	Topic	Total links	Correct	Incorrect	Related system%
1	Java	26	13	13	50
2	Object	16	10	6	65
3	Php	19	11	8	58
4	Php forms	16	8	8	50
5	Sql	22	13	9	60
6	Asp.net	20	13	7	65
7	Xml	24	10	14	40
8	Javascript	15	9	6	60
9	Mico	12	5	7	40
10	events	14	7	7	50

Table (1): The results from the previous related system.

Sr. No.	Topic	Total links	Correct	Incorrect	Combi system%
1	Java	33	17	14	55
2	Object	19	18	8	69
3	Php	26	15	9	63
4	Php forms	22	11	9	54
5	Sql	28	18	11	67
6	Asp.net	25	17	9	71
7	Xml	29	16	16	40
8	Javascript	18	13	8	65
9	Mico	20	10	9	45
10	events	21	12	9	50

Table (2): The results from our Combi system.

Combi system%	Related system%
55	50
69	65
63	58
54	50
67	60
71	65
40	40
65	60
45	40
50	50

Table (3): The Comparison of results from the previous related system and our Combi system.

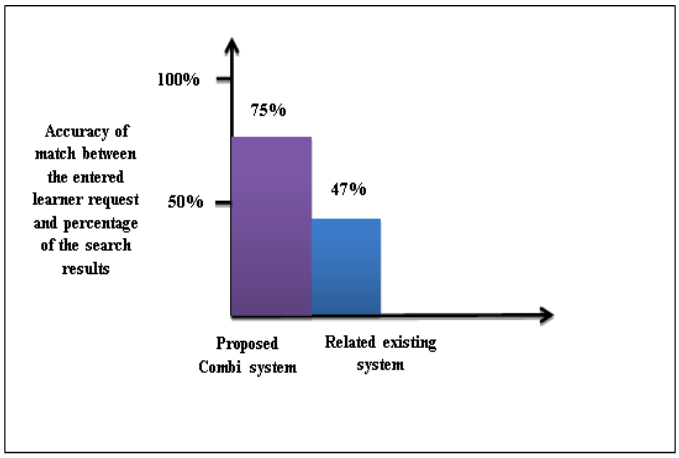

Figure (4): Comparative performance compared with the related existing system

VIII. CONCLUSION

Web data mining is the series of task used for mining or extracting useful information from the web pages or web sites. It provides intrinsic knowledge of teaching and learning process for effective education planning by applying various techniques. The ultimate goal we took of personalization is learner's satisfaction. Our Combi system saves information about the user's learning requirements without asking him explicitly. This information helps the system to make personalization of the learning content based on the learner's preferences, educational background and experience. Our Combi system then uses this feedback to present the learner with other, similar objects. The iterations continue until the learner finds an object of interest or abandons the search. After we discovered the learning behavior pattern we then can build up a series of feedback and motivation system. Our Combi system has better performance improvement and learner's satisfaction as compared to the related existing system, because we used three data mining techniques rather than two techniques, associated with the learner's profile and then used this whole information for recommendation and feedback from the learner for precision results.

IX. FUTURE WORK

For future work, research can be done in developing integration strategies for approaches that can accurately predict student performance in courses and approaches that help a select a subject or courses based on student interests

and learning objectives. This information on actual web usage by a learner can help in adapting a website to suit another similar potential learner.

REFERENCES

[1] Sandesh Jain, Dhanander K. Jain, Harihar Bhojak, Ankit Bhilwar, and Mamatha J., "Personalization of e-Learning Services using Web Mining and Semantic Web", International Journal of Machine Learning and Computing, Vol. 2, No. 5, October 2012.

[2] Essaid El Bachari1, El Hassan Abelwahed2 and Mohammed El Adnani, "E-learning personalization Based on Dynamic Learners' Preference", International Journal of Computer Science & Information Technology (IJCSIT), vol 3, No 3, June 2011, DOI : 10.5121/ijcsit.2011.3314200e-learning personalization.

[3] Radoslav Pavlov, Sofia, Bulgaria, Acad. G. Bonchev and Desislava Paneva, "Personalized and Adaptive eLearning - Approaches and solutions", Institute of Mathematics and Informatics, BAS, Bulgaria, Acad. G. Bonchev str., Sofia dessi@cc.bas.bg 2005.

[4] Umadevi, B.Uma Maheswari, P.Nithya, "Design of E-Learning Application through Web Mining", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 8, August 2014.

[5] Manasi Chakurkar1, "A Web Mining Approach for Personalized E-Learning System", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 3, 2014 Page 51, www.ijacsa.thesai.org.