Adaptive Tool for Searching Open Educational Resources

Adem Keddi
HiLCoE, Computer Science Programme, Ethiopia
ademkeddi@gmail.com

Mesfin Kifle
HiLCoE, Ethiopia
Department of Computer Science, Addis Ababa University, Ethiopia
kiflemestir95@gmail.com

Abstract

The advancement of technology and the popularity of the Internet created many opportunities. Educational institutions have been aware of its massive potential as a learning tool. The development of smart computing devices increased the use of online resources by the public. The Open Educational Resources (OERs) movement originated from developments in Open and Distance Learning (ODL) in the wider context of a culture of open knowledge, open source, free sharing, and peer collaboration. Nowadays, many educational and research institutions and individuals worldwide are producing and publishing OERs. There is great number of open access courseware that provide OERs freely to be used by students and anyone that is engaged in e-Learning activities.

Even though OER repositories provide open access to their databases, still there is no tool that crawls and indexes OERs that are found in different places and make the search adaptable based on the user profile making the search.

In this paper, the most appropriate approach to represent user preferences is selected. A tool that has functionalities of crawling OER databases and indexing OERs, user profiling, providing search interfaces and adapting the search results based on the user's profile is designed and developed. Finally, an experiment is conducted to test the effectiveness of the proposed tool. The tool is developed in small scale for the purpose of testing and evaluation using Open Source software. Experimentation is conducted on the developed tool to validate and verify the functional requirements and the objectives set are met.

Keywords: Personalization; Adaptive Learning; Adaptive Search Tool; Crawling; Indexing; Open Educational Resources

1. Introduction

Since the Internet was adopted and further developed by educational institutions as a means of communication in the 1970s, academics have been aware of its massive potential as a learning tool. The development of smart computing devices increased the use of online resources.

e-Learning is the use of technology to aid and enhance learning as pointed out by Carabaneanu *et al*. [1]. It can be as simple as students watching video lectures that are broadcasted from education centers or as complex as an entire university course provided online. e-Learning began decades ago with the introduction of television and overhead projectors in classrooms and has advanced to include interactive computer programs, 3D simulations, video and telephone conferencing and real-time online discussion groups comprised of students from all over the world. As technology advances, so does e-Learning making the possibilities endless.

In e-Learning, mostly Open Educational Resources (OER), are used as teaching materials. Bissell [2] described OERs as digitized materials offered freely and openly for educators, students, and self-learners to use and reuse for teaching, learning and research. OERs differ from other web resources in that they are openly licensed resources that can be used for teaching or learning. They are usually located in dedicated known repositories.

An increasing number of institutions and individuals worldwide are producing and publishing OER. Amongst these, perhaps the single most known program is MIT's Open Courseware, which provides open access to materials used to teach over 1,800 courses as pointed by Atkins *et al*. [3]. One can find

a growing list of higher education institutions participating in the creation of OER by visiting the Open Education Consortium (OEC) website [4].

OEC declared that as of July 2015 there are more than 280 member organizations worldwide. Besides MIT, University of California, University of Arizona, Cape Town University, Delft University of Technology, and Harvard University are some of the universities that made open educational resources available to the public.

At the start of the e-Learning concept, some educational institutions produced and published educational resources as additional content of the classrooms for their students [5]. After the move by MIT to put the entire course catalog online in 2001, many educational and research institutions joined the movement to make OERs open to the public.

There are too many OER repositories and courseware. These produce huge number of OERs that it is tedious to search for a particular resource in each repository separately and one by one. It is also difficult and time consuming to get a resource that matches the preferences and expertise level of the user making the search.

There are many works done on adaptable systems many of which worked on making the search in a single e-Learning system adaptable. They do not help on how to utilize resources that are found in different OER repositories and make the search adaptable to reflect the user making the query.

In this paper, we designed a tool that indexes open educational resources by crawling through OER databases to make the search in the OER repositories adaptable by making the search in the open access courseware smart enough to know the user's background and preference and hence tailoring the search results to meet the user's needs.

To achieve our objectives, we used qualitative research method to define adaptability and user preferences list in the context of web searching. Comparative study on different user profile data collection techniques is conducted to select the best method. The optimized user preferences list is developed by gathering data from different books and search engines like Google, Bing and Yahoo. This is used to create a user profile explicitly.

A tool that has functionalities of crawling and indexing open educational resources, and the selected approach to adapt the search results in open educational resources is designed and developed using the agile software development methodology. Finally, software validation and verification experiment is conducted to verify whether the tool is effective and meets its intended purpose.

2. Related Work

Many researches are done on adaptable systems and personalized learning. Greene *et al*. [6] proposed the use of anonymous profiles to develop a framework for an open source Adaptable Personal Learning Environment (APLE) by extending the already existing learning environment known as the Portland Learning Environment by adding the concept of adaptability. Their work is limited to making the existing learning environment adaptable.

Li and Chang [7] designed a personalized e-Learning system which can automatically adapt to the interests and levels of learners based on the IEEE Learning Technology Systems Architecture (IEEE LTSA), a component-based framework for general learning system. Their work is limited to adaptation of a single learning system and employs only the use of implicit information collection technique to construct user profile. This method does not guarantee the user profile created efficiently reflects the user as pointed out in [8].

After reviewing different personalization systems, Sugiyama *et al*. [9] proposed several approaches to make the web search adaptable as opposed to typical search engines which return similar results for the same query made by different individuals with different knowledge and preferences. They conducted experiments using the proposed approaches and evaluated the retrieval accuracy of the approaches to select the best one. From the experiment, the approach that utilizes user profile construction based on the modified collaborative filtering yield the best accuracy by enabling construction of more appropriate user profile and perform fine grained search that is better adapted to each user's preferences. The limitation of this work is that only implicit preferences are used which limits

the representation of the user. The use of web history for user profiling also poses a privacy concern. Another limitation is that the search will be made in the whole Internet which will return thousands of results that are not open educational resources, and filtering them will be a burden to the user.

Some researchers tried to make the search in a single e-Learning system to be adaptable. However, nowadays, there are huge numbers of OERs to be used freely for educational and other purposes. There is no previous research conducted on how to make available all of the OERs that are found in different repositories to be used by students, teachers, researchers, or anyone that embarked on e-Learning, and provide the functionality of adapting the search results. This paper used some of the findings of the above listed researches and fixed the limitations of some to propose an efficient tool that can provide adaptable search results in OER repositories.

3. The Proposed Solution

A tool that has the functionalities of indexing and crawling of OERs, constructing and managing of user profiles, and providing search results adaptation is proposed whose framework is shown in Figure 1.

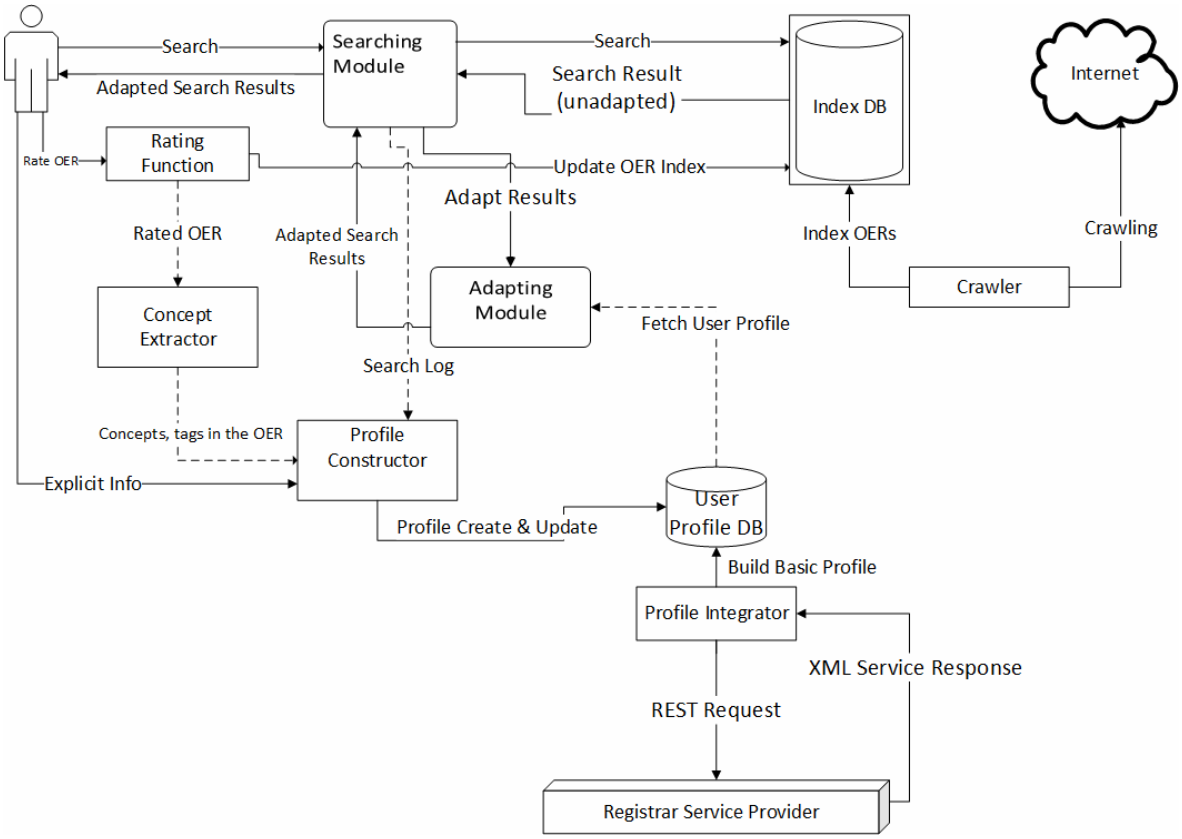

Figure 1: The Proposed System Framework with Emphasis given to User Profile Construction

As shown in Figure 1, the user searches for OERs by entering a search string in the system. The search string is passed to a controller in the search module. The controller validates the search string and performs search keyword extraction to create search keys. The search keys are then used to search in the OERs index database. Then the search module will pass the search results to the Adapting Module.

The Adapting Module retrieves the user profile from the User Profile Database and performs adaptation and returns the adapted search results to the Search Module. The search module returns the final adapted search result to the user. The user can rate OERs using the rating mechanism found in the search results page.

In the proposed tool, there are three core modules:

i. User profile construction and updating,

ii. Crawling and indexing OERs, and

iii. Searching results adaptation.

The details of each is presented below.

i. User Profile Construction and Updating

The proposed system uses user profile to adapt the search results. The following are the different clusters of data used for user model creation and management:

- *explicit preferences*: defined by the user such as sex, educational level, language skills, hobbies, topics of interest, etc.
- *implicit preferences*: that the system defines automatically by analyzing data about user behavior and activities.
- *user behavior data*: generated from the user statistics.

Among the available implicit information collection techniques, the search log is used to collect information of the user implicitly. Other techniques such as browser agents and desktop agents add some burden on clients since they are essentially client-side approaches [10]. Another advantage of using search logs is that it does not bring privacy issues to the user since it only collects information related to search.

For the user profile construction and updating, both implicit and explicit information collection methods are used to yield better result as previously conducted researches show that there is no clear answer whether implicitly created profiles are more or less accurate than explicitly created profiles [10].

ii. Crawling and Indexing OERs

The OERs databases are limited in number and the web addresses of most of them can be found from the consortium of OERs website and other related repository pages. For this application, we proposed developing a web page that contains URLs of OERs databases. If there is new digital library or OERs repository, the web address for the repository has to be added to this web page with ease. So, this web page will be used to build a URL seed list that is input to the crawler.

The high-level architecture of the crawler used for this tool shown in Figure 2 is adapted from the basic crawler cited in [11]. The basic difference is that this crawler only crawls OER repositories listed in the seed page.

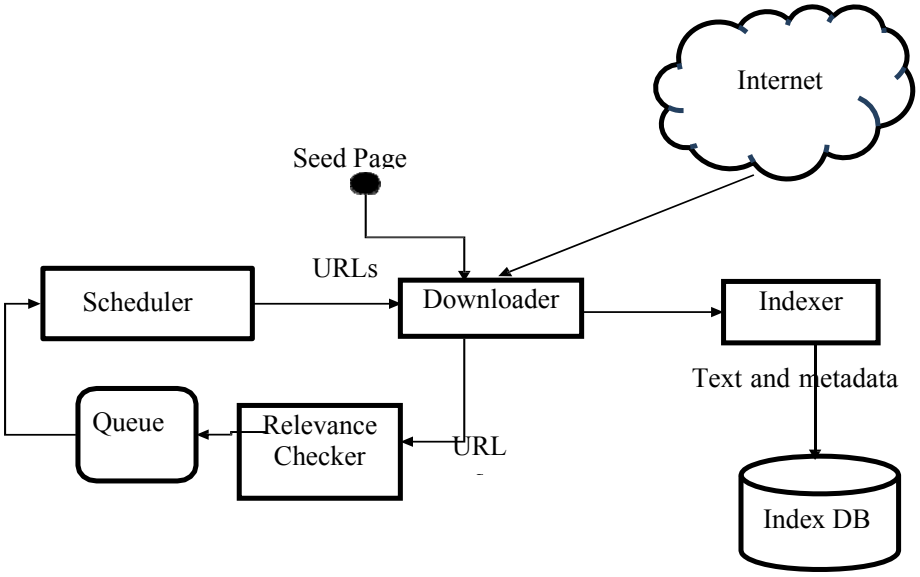

Figure 2: Adapted High Level Architecture of the Crawler

The seed page is maintained manually as there are limited number of OER repositories. The downloader gets the first URL from the seed page and downloads it. It then retrieves all the URLs in the downloaded page and passes the links to the Relevance Checker. The downloader also passes the downloaded file to the Indexer so that the file will be indexed to become suitable for searching.

The Relevance Checker checks and determines if the URL passed to it is an OER in the OER repositories. If it is determined to be an OER, the URL will be passed to the Queue and then to the

Scheduler. This process will be repeated until there are no URLs scheduled to be downloaded.

Algorithm of the Indexer

i. Get the title of the downloaded file
ii. Check if the crawled OER has keywords defined in the metadata
    a. If there are keywords in the metadata, copy them into indexed_file_keyword table
    b. Otherwise, skip
iii. Get the first page, get the first word (token), remove suffixes, capitalization, apostrophes, and put in the token list with frequency 1
iv. Get the next token, and remove prefixes, suffixes and capitalizations and compare with the list of tokens
    a. If match found, increment the frequency of the match by 1.
    b. Otherwise, add the token to the list of tokens and make its frequency
v. Go to step iv and continue until EOF is reached

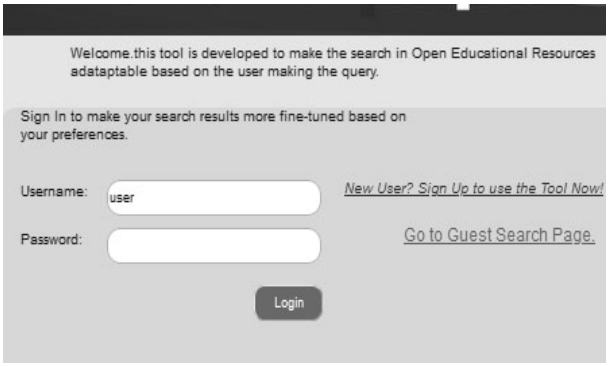

Figure 3: Login Page

4. Implementation

The system is developed in Open Source Software like MySQL Server and JavaServer Faces technology. Apache Tika library 1.2 is used for metadata extraction from the crawled resources.

The first page that is shown when the application is opened is the login page as shown in Figure 3. In this page, a user can make the search as guest user or login to the system. Logging into the system is not just for authentication purpose, but also it serves as the source of information for user profile generation in that it passes the user's ID to the system so that the necessary user profile is retrieved.

Figure 4 shows the explicit information filling or updating page.

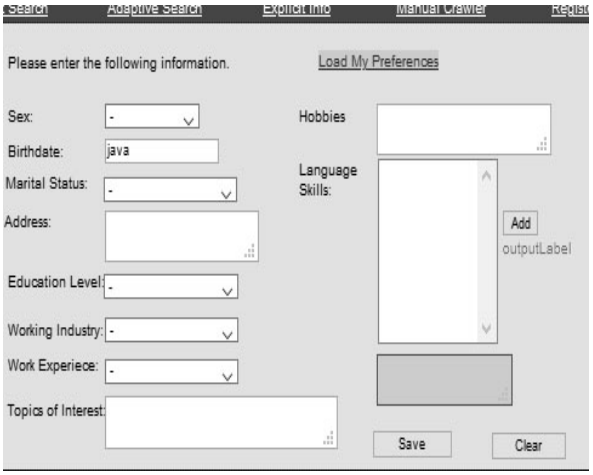

Figure 4: Explicit Information Filling Page

Figure 5 shows the search result for query string "java" made by a user with user profile of language skill of English and French and education level of graduate. So OERs that match both the language skill of the user (English and French) and education level (graduate level) are returned in the first top rows, each prioritized by their previous rates.

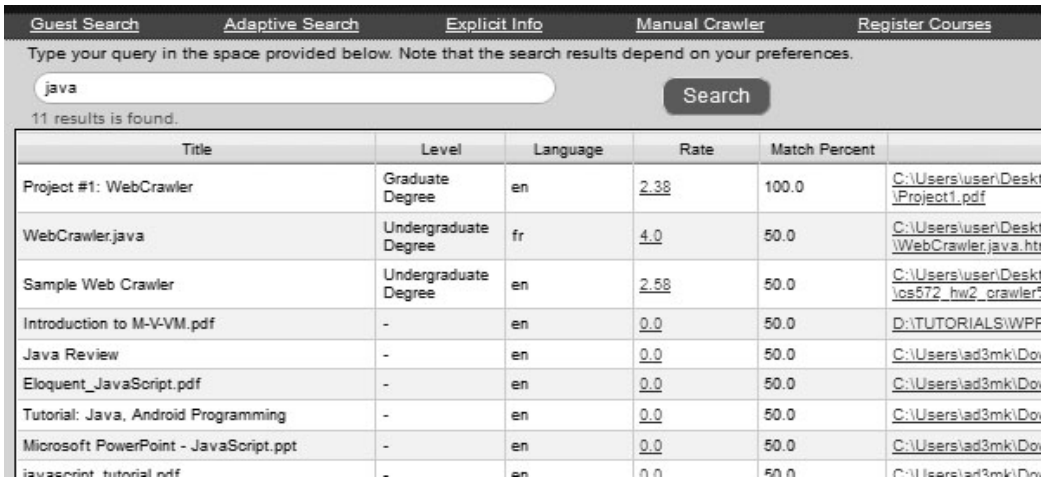

Figure 5: Search Result for Search String 'Java' for a User with Skill English and French and Educational Level of Graduate

Figure 6 shows the result of the search made by querying for "java" by a user that has only English language skill and education level of undergraduate. So OERs that are written in English and have level of undergraduate are prioritized at full matches and returned in the first rows, while OERs that match only by one of either of the two (language or education level) follow and they are prioritized by the rates given to them. Thirdly, OERs that match the search string but not the user's language and educational level will follow ordered by their rates. This ensures that all OERs that are returned by the search will be displayed to the user but adapted based on language skill, education level, and geographical location of the user making the query.

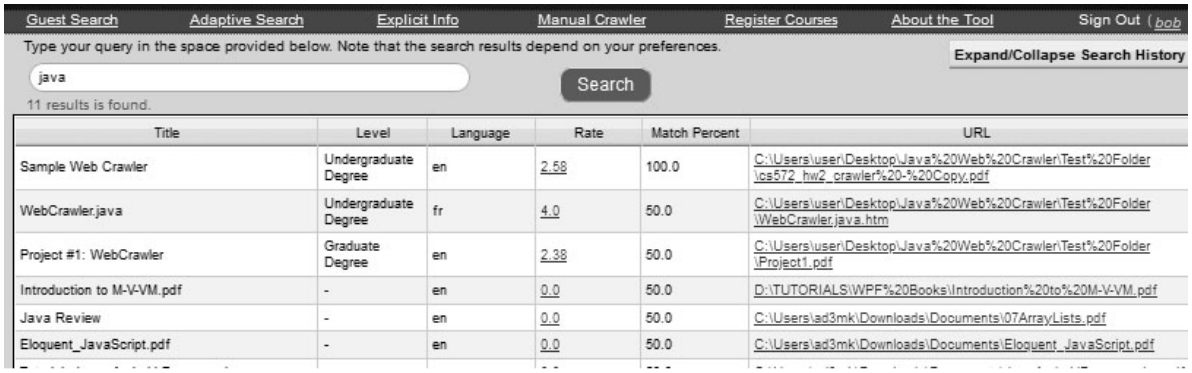

Figure 6: Search Result for Search String 'Java' for a User that Only Reads English and Educational Level of Undergraduate

For a guest user that searched for "java" without logging into the system, the search result looks as shown in Figure 7.

Note that in the search result, all are displayed without being ranked and prioritized since the system has no information about the user making the search. But here, the system can make adaptation based on the geographical location of the user which is retrieved from the IP address of the user.

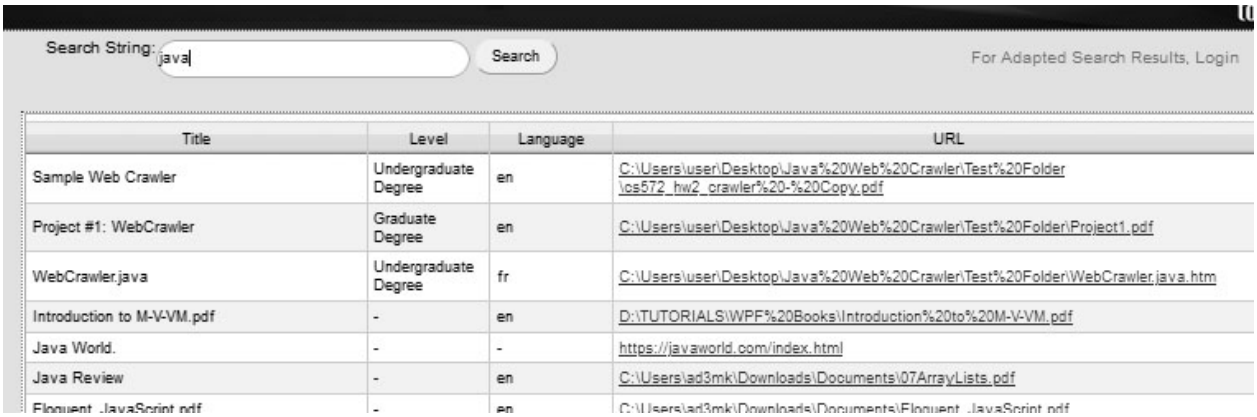

Figure 7: Search Result for Search String 'Java' for a User that is not Logged into the System

5. Software Validation and Verification

Software validation and verification is done using experimental method. For this purpose, 10 students are selected randomly with the objective to check the functional and non-functional requirements of the software and check and confirm the effectiveness of the tool in adapting search results based on the profiles of the users using the tool.

Using manual crawling functionality of the tool, few OER repositories are put into the crawler and around 300 OERs are indexed and used for the experiments.

The users have different profiles as shown below.

- Education level and field of study (Undergraduate, Graduate Computer Science, Graduate Software Engineering),
- Language they can read (e.g., English, Arabic, French, German, etc.),
- Guest users (no information about the users).

User's profile is created using the application's explicit profile collection functionality.

Out of the 10 users, 8 of them were made to perform the search in two modes: registered user with full profile and guest user. Two users used the tool as guest users only. The users are made to perform search using similar search strings. Finally, they are made to express their findings of the tool.

Those users that used the tool in the two modes found out that the results are ranked based on their language proficiency and education level filled during the explicit information collection. They also expressed the tool prioritizes the results based on the rates of the OERs. They also noted that when they use the tool in the guest mode, the tool prioritizes the search results based on the rate given to the tool.

The two users that used the tool as a guest found that the tool displays the results in unspecified order. But as the users that used the tool in both modes spotted, the search results are based on the rates given by the registered users to the OERs.

All of the users that tested the tool as registered users identified that the tool performed in conformance to their profiles and the tool's search logging functionality is promising.

6. Conclusion and Future Work

For evaluation purpose, a tool that provides all the functionalities (in small scale) is developed. Software validation and verification testing is conducted on few randomly selected users.

The main attributes used fore validation and verification are:

- the user's educational level,
- language skills,
- geographical location, etc.

From the result of the experiment, the system is verified to adapt based on the user performing the search. The evaluation of the proposed tool is done and is found to be satisfactory even though few users participated and the user profile is not mature.

The search algorithm used in this tool is basic and it should be developed more to take more user profiles in adapting and predicting the user's needs. The tool builds and updates user profiles as users

interact with the system. This is one of the outstanding features of the proposed tool.

To speak of the effectiveness of the tool in quantitative way, all of the modules will be implemented including the indexer, user profiler and user profile integrator. Evaluating the efficiency of the tool as a whole will also be done in the future.

References

[1] Luciana Carabaneanu, RomicaTrandafir, and Ion Mierlus Mazilu, “Trends in e-Learning,” in Proceedings of MMT2006, January 2006.

[2] Bissell, A. “Permission granted: Open licensing for educational resources”, The Journal of Open and Distance Learning, Vol. 24, 2009, pp. 97-106.

[3] Daniel E. Atkins, John Seely Brown, and Allen L. Hammond, “A Review of the Open Educational Resources (OER) Movement: Achievements, Challenges, and New Opportunities,” Retrieved from https://hewlett.org/wp-content/uploads/2016/08/ReviewoftheOERMovement.pdf, Last accessed on July 2015.

[4] Open Education Consortium Official website, http://www.oeconsortium.org, Last accessed on July 13, 2015.

[5] Hal Abelson, “The Creation of OpenCourseWare at MIT”, Journal of Science Education and Technology, Vol. 17, No. 2, pp. 164-174, April 2008.

[6] Green, S., Nacheva-Skopalik, L., and Pearson, E., "An Adaptable Personal Learning Environment for e-Learning and e-Assessment", in Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing, CompSysTech'08, 2008.

[7] Xin Li and Shi-Kuo Chang, “A Personalized E-Learning System Based on User Profile Constructed Using Information Fusion,” in Proceedings of the 11th International Conference on Distributed Multimedia Systems, DMS 2005.

[8] Liliana Ardissono, Cristina Gena, Pietro Torasso, Fabio Bellifemine, Angelo Difino, and Barbara Negro, “User Modeling and Recommendation Techniques for Personalized Electronic Program Guides”.

[9] K. Sugiyama, K. Hatano, M. Yoshikawa, and S. Uemura, “Adaptive Web Search Based on User‟s Implicit Preference”, DEWS2004, 5-B-04.

[10] Susan Gauch, Mirco Speretta, and Aravind Chandramouliand Alessandro Micarelli, “User Profiles for Personalized Information Access”, in The Adaptive Web, Methods and Strategies of Web Personalization, January 2007.

[11] Shyam Nandan Kumar, “World towards Advance Web Mining: A Review”, American Journal of Systems and Software, 2015, Vol. 3, No. 2, pp. 44-61.