

## Design of Open Content Social Learning that Increases Learning Efficiency and Engagement Based on Open Pedagogy

### **Benneaser JOHN**

*SVP, Engineering – AppNexus Inc., New York, USA  
Research Scholar, Karunya University, Coimbatore, India  
benneaser@gmail.com*

### **Dr. V. Thavavel**

*Assistant Professor  
College of Computer and Information Sciences  
Prince Sultan University, Saudi Arabia  
thavavelmurugesanv@gmail.com*

### **Dr. Jayakumar JAYARAJ**

*Associate Professor  
Department of Electrical and Electronics  
Karunya University, Coimbatore, India  
jayakumar@karunya.edu*

### **Dr. A. Muthukumar**

*Associate Professor  
Department of Computer Science & Engineering  
PSG Institute of Technology and Applied Research  
Neelambur, Coimbatore, India  
amk\_in\_2000@yahoo.com*

### **Dr. Poornaselvan Kittu JEEVANANDAM**

*Program Manager, Tata Consultancy Services  
Edison, NJ, USA  
poornakj@gmail.com*

### **ABSTRACT**

Due to the rapid growth in Internet resources, mobile technologies and social media, teaching and learning are increasingly adapting to the notion that 'content is open; learners are social'. The learning materials are open but effective learning is challenging due to the explosion of unstructured content on the web. The effectiveness of learning on the web largely depends on the relevancy of the content and the learner's engagement. This paper's objective is to develop an Open Content Social Learning(OCSL) system, to compare different pedagogical strategies and algorithms on improving effective learning. This paper proposes an enhanced learner-centered online learning experience by matching the content based on learning goals, historical learning preferences and behaviors from other learners with similar goals to increase the learner interaction and engagement.

### **INTRODUCTION**

Open Educational Resources (OERs) are teaching and learning materials that anyone can use and share freely, without charge. Since first being coined by UNESCO in 2002, the term Open Educational Resources has evolved to meet the fast pace of the change and the diverse contexts in which it has now been used (Bossu, Bull, & Brown, 2012). The worldwide OER movement is rooted in the idea of high quality education at no cost. The Cape Town Declaration (2007) states that “Educators worldwide are developing a vast pool of educational resources on the Internet, open and free for all to use. These educators are creating a world where each and every person on earth can access and contribute to the sum of all human knowledge. They are also planting the seeds of a new pedagogy where educators and learners create, shape and evolve knowledge together, deepening their skills and understanding as they go.”

Open learning enables learners to be self-determined and interest-guided. Stacey (2013) educators to “Go beyond open enrollments and use open pedagogies that leverage the entire web not just the specific content in the MOOC platform”. Learners are often unable identify which material is needed, useful, and required at their level. Hence, open content learning design must assimilate the material from various sources and provide a new pedagogy that is appropriate to the needs of today’s learners (Smyth, Bossu & Stagg, 2015). This paper explains the design for an Open Content Social Learning (OCSL) system that leverages Open Content to deliver an adaptive and personalized experience accounting for the pedagogical needs of the learners and similar learners

and the need to recommend learning activities in a pedagogically effective order.

## RELATED RESEARCH

Learner's experiences with open learning do not always contribute to effective learning because some traditional pedagogical strategies are still being used. Over the past decade, researchers have investigated different pedagogical strategies for making the online learning environment effective. Sathiyamurthy & Geetha (2012) state that "The effectiveness of an e-learning system for distance education to a large extent depends on the relevancy and presentation of learning content to the learner". In a recent study, Kim & Reeves (2007) showed that the increase in online courses has definitely helped to reach millions of learners, but the educational effectiveness of online courses is a subject of debate. Learning must be personalized based on the learner's goals and style and compared with "learner-like" learners (individualized and collaborative) as well as adaptive learning resources (organized and filtered), while considering motivation and engagement tools (Cheung, Lam, Szeto, & Yau, 2008). The goal of the adaptive presentation is to adapt the content to the user's goals, knowledge, and other relevant information. The architecture for an Adaptive Hypermedia System adapts the content of a hypermedia page to the user's goals, knowledge, preferences, and other user information for each individual user who is interacting with the system (Stern & Woolf, 2000).

Another aspect of effective search and personalized results is consideration of the learner's profile. All learners are unique; no two will achieve the same learning outcomes across a range of subject areas. Clear guidance can be provided on the diverse learning needs of each student by collecting and continuously updating metadata that is stored for learners in user profiles. Chan (2000) describes that implicit profile creation based on observations of users actions has been used in more recent projects and describes the types of information that is available. This model considers the frequency of visits to a page, the amount of time spent on each page, how recently a page was visited, and whether the page was bookmarked. Paireekreng & Wong (2010) observe that prior knowledge of each learner's activity and an effective user profile is required for personalization. M.P. Cuéllar, M. Delgado, and M.C. Pegalajar (2011) have considered social networks to be a type of Learning Management System (LMS). Social Network Analysis (SNA) is conducted for teachers, learners, learning resources and their interactions. Vassileva, J. (2008) emphasizes that the two main goals of the design of social learning environments should be making them learner-centered and making learning more gratifying. In recent research, association rule-mining algorithms have been used to solve the problem of web page recommendations. A web usage log is used in adaptive association rule-based web mining, which attempts to personalize the results.

Research shows that effective learning requires the following:

1. Learner centric adaptive learning by personalizing with relevant content based on the learner's goals, style, habits and prior knowledge;
2. Learner centric social learning based on the goals, learning style and behavioral patterns of similar learners;

Current Open Content Learning systems include: OER Commons (Yoav Yair 2014, D'Antoni, S 2009), isseek.org (Bansal 2013), Project MERLOT (Malloy & Hanley 2001; Hanley 2015), OCW (Vahdati 2015) and mooc-list (Holotescu, Grosseck, Cretu & Naaji, 2014). Most of these systems are not personalized and do not provide adaptive content. Learners use these platforms as content viewers, and there is no engagement. They do not offer personalized content based on a learner's goals and prior knowledge. To overcome these limitations, the proposed work is to develop an Open Content Repository by consuming the OER content and personalizing the learning experience based on the learner's goals and activities and similar learners' learning activities.

Another aspect of effective search and personalized results is consideration of the learner's profile. All learners are unique; no two will achieve the same learning outcomes across a range of subject areas. Clear guidance can be provided on the diverse learning needs of each student by collecting and continuously updating metadata that is stored for learners in user profiles. Chan (2000) describes that implicit profile creation based on observations of users actions has been used in more recent projects and describes the types of information that is available. This model considers the frequency of visits to a page, the amount of time spent on each page, how recently a page was visited, and whether the page was bookmarked. The user's learning behavior is used to create user profiles in several systems. Paireekreng & Wong (2010) observe that prior knowledge of each learner's activity and an effective user profile is required for personalization. Open pedagogy could be considered to be a blend of personalized adaptive design, algorithms and technologies, and networking among learners, which makes the learning process effective and engaging.

### OPEN PEDAGOGY AND LEARNER-CENTERED LEARNING

Some early MOOC experiments were based on a pedagogy of connectivist learning (Milligan, Littlejohn, & Margaryan, 2013), which connects many people in a loose online network that enables them to share their ideas and learn together. While this approach harnesses the power of many voices and technologies, it is difficult to manage at a large scale and requires learners to know how to navigate the web resources and engage with their peers (de Waard, Koutropoulos, Keskin, Abajian, Hogue, Rodriguez, & Gallagher, 2011). So which pedagogies actually improve with scale? Some effective methods of teaching, such as personal tutoring, cannot scale up to thousands of learners without enormous costs, even though researchers in artificial intelligence have been attempting for many years to develop computer-based tutors. In contrast, methods of direct instruction scale well – a good educational television program can inform a hundred people, or a million – but they are not very effective at engaging people in active and reflective learning. There is a general theory of scale that can be applied to education. The Network Effect proposes that the value of a networked product or service increases with the number of people who use it (Sharples, Adams, Ferguson, Gaved, McAndrew, Rienties, Weller & Whitelock, 2014). For example, a telephone system becomes more valuable when we connect millions or billions of phone users worldwide. The worldwide web benefits from interconnecting millions of people through their computers. But people are not solely points in a network; we have knowledge and perspectives to share. Thus, the Social Learning Effect can be stated as such: the value of a networked learning system increases as it enables people to learn easily and successfully from each other. Another difficulty experienced by many who have participated in connectivist MOOCs (Milligan, Littlejohn, & Margaryan, 2013) is the feeling of being ‘lost in hyperspace,’ of having too many options and possibilities and not knowing where they are in a learning activity, who to engage with, and where to go next.

Most existing e-learning platforms and tools focus on technology without rigorous investigation of the pedagogical issues or quality control of the e-learning material. The motivation to learn and engage with the e-Learning solution is key to its effectiveness, especially when the effectiveness is defined as the time spent using the product: ‘Results suggest the importance of motivation to learn and workload in determining aggregate time spent in e-learning courses’ (Brown, 2005). Open pedagogy could be considered to be a blend of personalized adaptive design, algorithms and technologies, and networking among learners, which makes the learning process effective and engaging.

### OPEN CONTENT SOCIAL (OCSL) SYSTEM

This section summarizes the general overall system architecture and design of OCSL before discussing the individual modules in detail. OCSL is a personalized learning system represented in figure 1 uses complex algorithms to automatically learn a learner’s interests with respect to learning activities. It then makes highly personalized content recommendations based on the goals, past activity and similar learners’ activities.

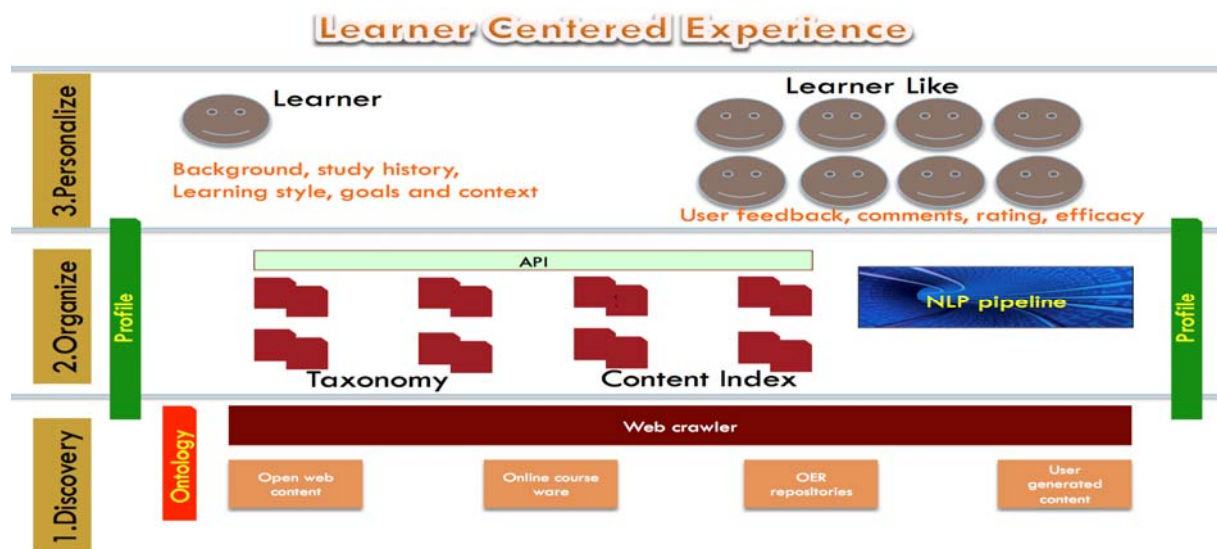


Figure 1. Overview of the Learner-Centered Learning Experience leveraging Open Content.

Research shows that most of the Open Content learning platforms currently use standard search techniques by combining conventional information retrieval techniques that are based on page content, such as word vector space (Salton, & McGill, 1983), with link analysis techniques based on the hypertext structure of the Web, such

as PageRank (Brin & Page, 1998) and HITS (Devi, Gupta, & Dixit, 2014). The PageRank algorithm (Brin & Page, 1998) attempts to provide an objective estimate of the Web page importance. However, the importance of the Web pages is subjective for different users. The true relevancy of a page depends on the interests, goals and existing knowledge of the individual users; a global ranking of a Web page might not necessarily capture the importance of a page for a given individual user. OCSL expands the scope of the search to generate more personalized results and greater learning engagement using the following two modules:

**A. Offline Process:**

1. The content manager reads the content (Crawling, API calls, Streaming API).
2. The content classification engine analyzes the content.
3. The system sends 20% of the content to the Natural Language Processing NLP API.
4. After categorization, the content is verified by Amazon Mechanical Turk through APIs.
5. The remaining 80% of the content is classified using the Naïve Bayes classifier (Patil & Pawar 2012) algorithm.
6. Once the content is classified with attributes (meta-data), it is loaded into the content index.

The content index indexes the attributes and stores it inside the Apache Solr container. This content index is updated periodically through an offline process.

**2. Online Process:**

1. The learner inputs his/her goals, learning style, and relevant content.
2. The pedagogy engine formulates the query to retrieve content in three ways, depending on the historical information and the learner’s goals:
  - a. Conventional search using an inverted index and page ranking algorithm.
  - b. Improved results based on the Content Hierarchy and Learner attribute-based Matching (CHLAM) of the OCSL system.
  - c. Superior results based on CHLAM and Similar Learners Attribute-based Matching (CHSLAM) of the OCSL system.
3. Filter the content results.
4. Implicitly capture the learner’s activity and use it as a feedback loop to apply to the learner’s profile attributes.

Each module performs its defined function and exchanges information with other modules, as shown in figure 2.

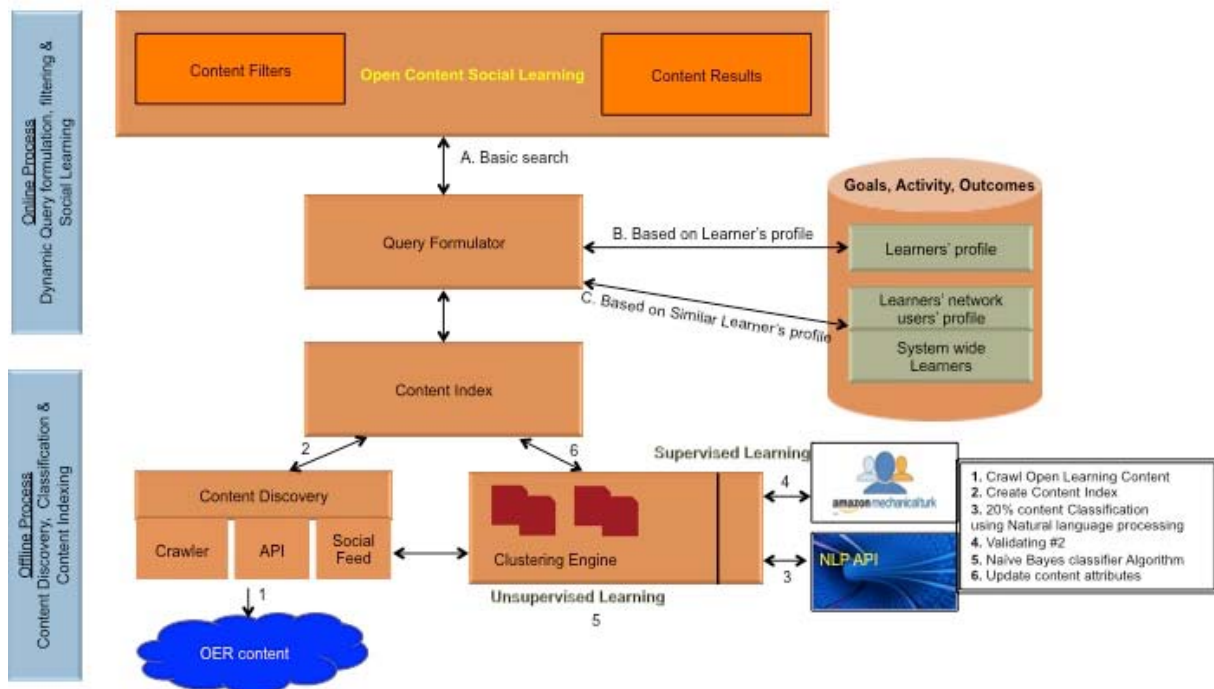


Figure 2. System Architecture of the OCSL Work

The role of content discovery is to crawl open content from the Internet, i.e., the World Wide Web and social media, and to locate content to present to the user. The content manager is configured to collect content from three sources: 1. Crawling OER content sites 2. Streaming API against social media platforms 3. API calls



against learning platforms such as MERLOT (Hanley, 2015), OER Commons, Gooru learning.

Content clustering entails grouping similar uncategorized documents together based on similarity measures. Content classification categorizes and organizes content by combining multiple methods of context-sensitive analysis. The clustering engine consumes content from multiple sources (Nutch Crawler, Federated API search, and Streaming API for social media feeds) and performs the following steps:

1. Alchemy’s machine learning APIs (Quercia, Askham, & Crowcroft, 2012) are used for categorizing the content. OCSL uses the Taxonomy API to perform classification. The Entity API calls fetch the desired Internet web page, normalizes it, and extract named entities, topics, and other content.

a. [http://www.alchemyapi.com/api/taxonomy\\_calls/urls.html](http://www.alchemyapi.com/api/taxonomy_calls/urls.html)

b. <http://www.alchemyapi.com/api/entity/urls.html#rurl>

Using the Taxonomy and Entity API, content metadata is updated in the Solr content repository.

2. As recommended by Wang, Kraska, Franklin, & Feng (2012), OCSL leveraged a hybrid human-machine approach in which machines are used to perform an initial, coarse pass over all of the data, and people are used to verify only the most likely matching pairs. OCSL integrates with the Amazon Mechanical Turk API to verify the classified content.

3. Using the Apache Mahout framework and Naive Bayes classifier algorithm (Patil & Pawar 2012), OCSL automatically classifies documents using a training set developed from the previous two steps. The training set includes documents that are already associated with a category. Using this set, the classifier determines, for each word, the probability that it reflects a document that belongs to each of the considered categories. To compute the probability that a document belongs to a category, the classifier multiplies together the individual probabilities of having each of its words in this category. The category that has the highest probability is the category that the document is most likely to belong to.

4. OCSL updates the content index engine with all of the taxonomy attributes (URL, content category, content sub category, content type, last modified, and many more).

The Dynamic Query Formulator is the core component of the OCSL system design. Most conventional search engines function with a search query that is limited and not as good as searching by phrases. The pedagogical engine uses a dynamic query formulator algorithm that was developed through this research to navigate a learner’s learning experience by analyzing his/her user interactions and prior learning knowledge on any given topic. The OCSL pedagogical engine also dynamically generates a query based on similar learners’ learning experiences.

Learner Attribute-based Matching (LAM) enhances the conventional search experience by building a user profile to provide more personalized search results based on learning style, type of content, recent activity, content categories, or other interests of the users. To build an intelligent pedagogical learning engine based on attributes, this system ensures that both users and documents are tagged with the same types of attributes. We are implicitly and explicitly collecting information from learners about their learning behaviors, learning goals, and other criteria. Basically, the pedagogy engine is responsible for figuring out both the most appropriate way to construct the queries and which data to use in them to optimize the relevancy of the learner’s learning experience. While a conventional search engine builds a sparse matrix of terms that are mapped to documents in the content index, OCSL enhances the design to map the user’s behavior to those documents. The Learner Attribute-based Search enables the system to classify users and content into a hierarchy that goes from more general to more specific categories, but it is further possible to query this hierarchy and apply a stronger relevancy weight to more specific matches:

*Learner\_Profile:*{

*MostLikelyCategory:*"engineering.computerscience.artificialintelligence",

*2ndMostLikelyCategory:*"engineering.computerscience.datastructures",

*3rdMostLikelyCategory:*"engineering.mathematics.algebra", ... }

First, each category from a learner’s profile can be broken into three terms in the query, with each term corresponding to a level of specificity in the classification:

*(engineering.computerscience.artificialintelligence vs. . engineering.computerscience.datastructures vs. engineering.mathematics.algebra).*

Second, each term is assigned a different query weight, with higher weights assigned to more specific terms. This arrangement serves the purpose of boosting the more specific (and presumably better) matches higher in the search results. Third, there are three distinct sets of queries, which correspond to the three potential classifications that are listed on a learner’s profile:

(*engineering.computerscience.artificialintelligence, engineering.computerscience.datastructures, engineering.mathematics.algebra*).

The end result is that by using query weights on terms that combine a measure of their probability (most likely to least likely) and their specificity (most descriptive to least descriptive), a fuzzy query can be constructed to match documents that match any of the criteria; at the same time, it boosts documents to the top of the search results that match the best combinations of those attributes within the hierarchy.

The query parameter also allows the author to weight the fields differently. This parameter can be used to make a query match in one field more significant than a query match in another field.

$$qf = field_1^{v_1} + field_2^{v_2} + \dots + field_n^{v_n}$$

where  $qf$  is the Query Fields, and  $v$  is the weight for each field, based on the learner's goals and interests as calculated and applied dynamically. In our approach, we personalize PageRank scores by assigning weights to the fields based on matched goals and activities based on the learner and similar learners. At the query time, the user's profile matches with the corresponding personalized values.

By mapping the learning behavior of users to documents, OCSL system is effectively creating links in the index between documents. Klačnja-Milićević, Vesin, Ivanović, & Budimac (2011) recommended that similar users learn similar content, which means that documents that are mapped to similar users are likely related. To make use of these relationships to recommend learning items to a new user, we find other similar users and recommend other items. OCSL provides a mechanism to form a social network among the learners who have similar learning interests, preferences, and learning experiences based on the data collected. A learning group in OCSL is a group of learners who share common learning goals and mutually recommend learning content that meet those goals. OCSL uses User-based Collaborative filtering and Item-based Collaborative filtering (Drachler, Hummel & Koper, 2008) to filter the learning content and recommend learning activities in a pedagogically effective order.

To evaluate our design, we conducted a Web crawl against Open Educational Resources (OER) and implemented a dynamic query formulator engine. We performed an experimental study that focused on Science, Technology, Engineering, Mathematics (STEM) engineering students. Our study explored the results of the following three algorithms, to validate the idea of effective learning by personalizing the content results. The study lasted for almost three months. Learners were grouped into 15 groups.

1. Algorithm 1 – Basic search using inverted index and page ranking conventional algorithm
2. Algorithm 2 – Search based on the Content Hierarchy and Learner Attribute-based Matching (CHLAM) of the OCSL system
3. Algorithm 3 – Search based on CHLAM and Similar Learners Attribute-based Matching (CHSLAM) of the OCSL system

We asked each learner to use our OCSL system after they entered their goals and profiles into our system. We did not provide any information about the main goal of the system. The learners were expected to use the platform and learn based on their choice of preferences. A results page was shown with the recommended content based on the three different types of algorithms mentioned above. Figure 3 is a screen shot of the OCSL system.

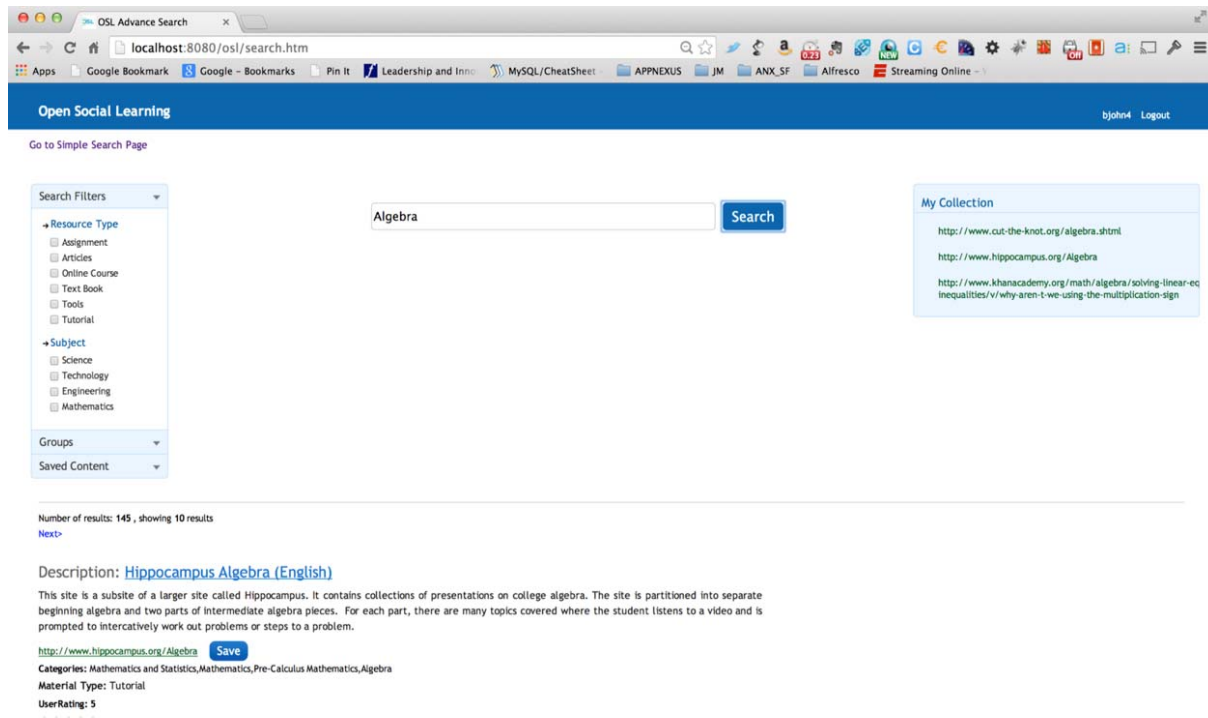


Figure 3. OCSL System screen shot

## TESTING APPROACH AND RESULTS

Comparing search results and recommendation systems is difficult. The best way to experiment with different relevancy parameters is to run A/B experiments that randomly divide users into groups over the same time period, with each group interacting with a different algorithm. Another common method for measuring the relative performance of algorithms involves generating test data and performing comparative analysis using the generated log data (Khosla, & Bhojane, 2013). To experiment with learning activities in detail, behavioral patterns were extracted from the log files and user activity database table.

There are two aspects of a search result set that determine the quality of the results, the precision and recall, as Powers and David (Powers & David, 2011) suggest. Precision is the fraction of the retrieved documents that are relevant. A precision of 1.0 means that every result that is returned by the search is relevant, but there could be other relevant documents that were not a part of the search result.

$$precision = \frac{| \{relevant\ documents\} \cap \{retrieved\ documents\} |}{| \{retrieved\ documents\} |}$$

Recall is the fraction of the relevant documents that are retrieved. A recall of 1.0 means that all of the relevant documents were retrieved by the search, irrespective of the irrelevant documents also included in the result set.

$$recall = \frac{| \{relevant\ documents\} \cap \{retrieved\ documents\} |}{| \{relevant\ documents\} |}$$

If all of the documents are retrieved, then the recall is perfect but the precision may not be good. On the other hand, if the document set contains only a single relevant document and that relevant document is retrieved in the search, then the precision is perfect but again the result set may not be good. This relationship shows a trade-off between the precision and recall, in which they are inversely related.

The F-score is a measure of a test's accuracy. It considers both the precision  $p$  and the recall  $r$  of the test to compute the score:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

In this approach, we can take previously saved user behavior data from log files and test how well each of the candidate algorithms predicts the results that were previously acted on by the users. In the case of OCSL, we take the list of search results for every search or recommendation run for the user and plot them in aggregate on a precision versus recall graph, showing whether the algorithm made the correct prediction based on the user's

historical behavior. For example, the correct prediction might be defined in terms of which learning materials a user consumed, and thus, any query model that resulted in higher precision and recall for that learning content would be considered to be a better algorithm.

We analyzed the system logs and calculated the Precision, Recall and F-Score based on the learner’s activity for each algorithm. In the following results table, each row indicates the aggregated result of a group of learners who interacted with the system. The Learning activity indicates the number of times each learner interacted with the system. The Total recommendations show the number of learning (retrieved) documents that were displayed to the learners, while the Total documents indicate the possible number of documents (relevant documents) that were related to the search.

Table 1. Conventional search using an inverted index and page ranking algorithm

Group #	# of interactions	# of recommendations	Total no. of documents	Precision	Recall	F-Score
1	12	510	17519	0.0235	0.0007	0.001368981
2	7	2939	15090	0.0024	0.0006	0.000927337
3	4	722	17307	0.0055	0.0002	0.000462127
4	9	560	17469	0.0161	0.0005	0.001029866
5	38	103	16993	0.0367	0.0024	0.004462451
6	35	146	17883	0.2397	0.002	0.003906686
7	99	172	17857	0.5756	0.0056	0.011026955
8	32	660	17369	0.0485	0.0019	0.00367795
9	4	459	17570	0.0087	0.0002	0.000455224
10	24	1609	16420	0.0149	0.0016	0.002918998
11	20	830	17199	0.0241	0.0012	0.002323015
12	77	876	17153	0.0879	0.0047	0.008937899
13	32	137	17892	0.2336	0.0018	0.003570632
14	30	168	16344	0.0178	0.002	0.003664306
15	51	80	17949	0.6375	0.0028	0.005666667

Table 2. Search based on the Content Hierarchical and Learner Attribute-based Matching (CHLAM) of OCSL

Group #	# of interactions	# of recommendations	Total no. of documents	Precision	Recall	F-Score
1	123	810	17219	0.1519	0.0074	0.014185215
2	160	616	17413	0.2597	0.0094	0.018209754
3	140	439	17590	0.3189	0.0081	0.015792442
4	120	218	17811	0.5505	0.0068	0.013384641
5	230	443	17586	0.5192	0.0132	0.025819488
6	230	266	17763	0.8647	0.013	0.025565498
7	211	612	17417	0.3448	0.0124	0.023939188
8	227	389	17640	0.5835	0.013	0.025409974
9	211	411	17618	0.5134	0.0121	0.023669303
10	220	409	17620	0.5379	0.0126	0.024663677
11	166	260	17769	0.6385	0.0094	0.018511291
12	121	120	16829	0.1008	0.0077	0.014277286
13	177	934	17095	0.1895	0.0108	0.0204956
14	156	303	17726	0.5149	0.0089	0.017447713
15	110	900	17129	0.1222	0.0067	0.012761761

Table 3. Search based on CHLAM and on Similar Learners Attribute-based Matching (CHSLAM)

Group #	# of interactions	# of recommendations	Total no. of documents	Precision	Recall	F-Score
1	298	330	17699	0.903	0.0169	0.033116631
2	199	260	17769	0.7654	0.0112	0.02215049
3	82	76	17953	1.0789	0.0046	0.009093429
4	120	140	17889	0.8571	0.0067	0.01332667
5	310	311	17718	0.9968	0.0175	0.034390947
6	215	217	17812	0.9908	0.0121	0.023853109
7	120	124	17905	0.9677	0.0067	0.01331484
8	307	330	17699	0.9303	0.0174	0.034099745
9	111	101	17928	1.099	0.0062	0.012306669
10	130	150	17879	0.8667	0.0073	0.014437226
11	144	166	17863	0.8675	0.0081	0.01599378
12	168	172	17857	0.9767	0.0094	0.018640777
13	141	146	17883	0.9658	0.0079	0.015645806
14	318	320	17709	0.9938	0.018	0.035280413
15	119	120	17909	0.9917	0.0066	0.013201686

The data in the table represents aggregate precision and recall calculations that are based on the learners in 15 different groups. Table 3 shows that the learning groups that used OCSL with the CHSLAM algorithm had an effective learning experience by interacting with the system more than the user groups that used the OCSL with the conventional and CHLAM algorithms. The precision is calculated as (# correct matches) / (# total results



returned), and the recall is calculated as (# correct matches) / (# correct matches + # missed matches). Although the precision and recall are not perfectly negatively correlated, there is a natural tension between the two in such a way that improvements in one often lead to declines in the other. The data from the table can be easily turned into a graph. All three tables are generated as graphs in Figure 4, Figure 5, and Figure 6, which show that the CHSLAM algorithm of OCSL generates improved results.

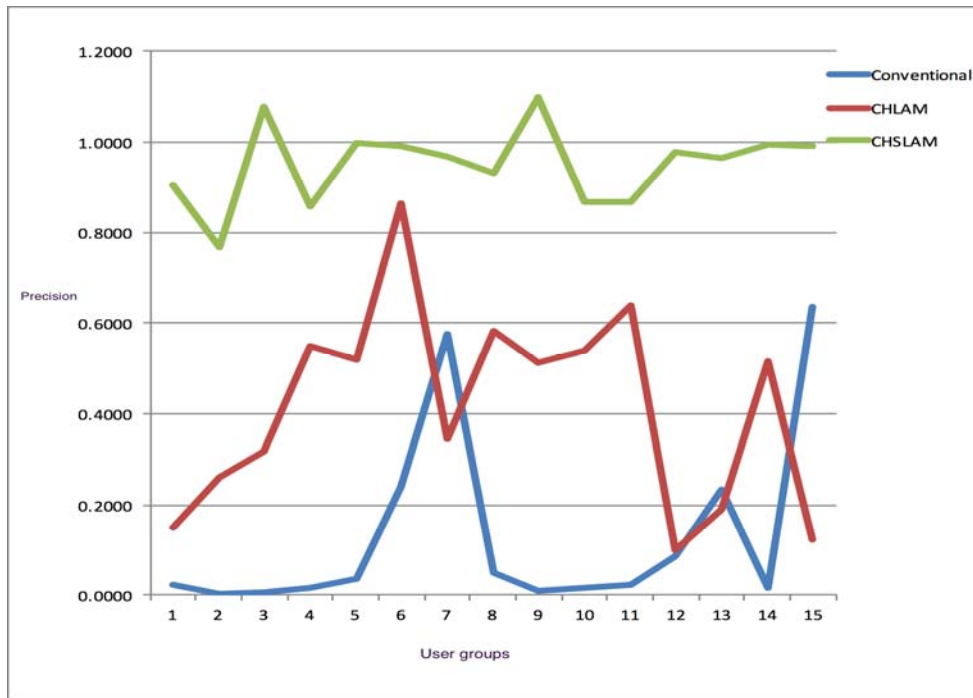


Figure 4. Precision values for Conventional, CHLAM and CHSLAM of OCSL algorithms

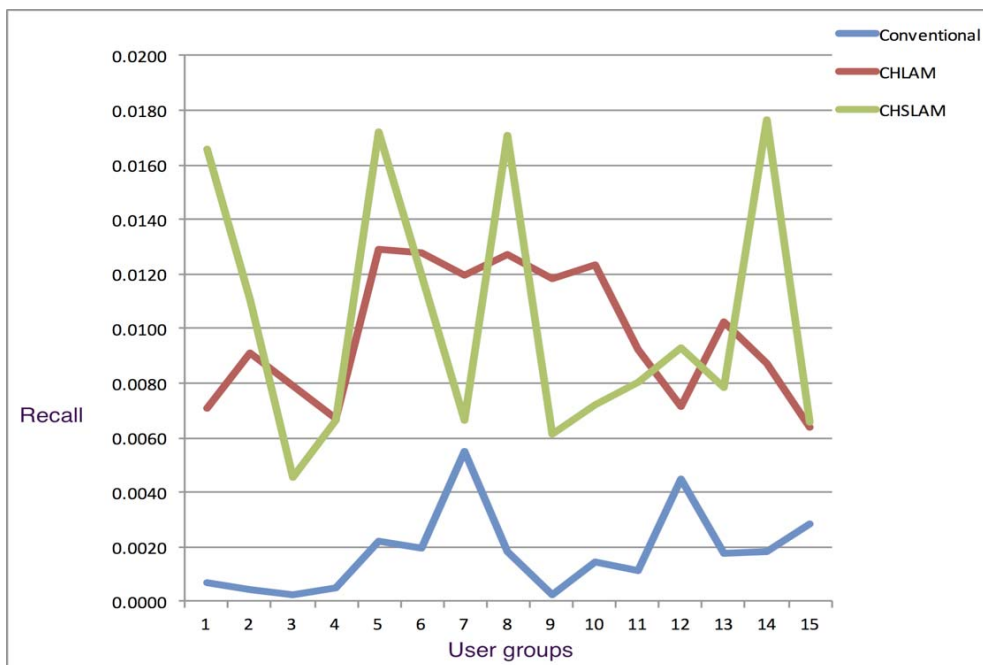


Figure 5. Recall values for Conventional, CHLAM and CHSLAM of OCSL algorithms

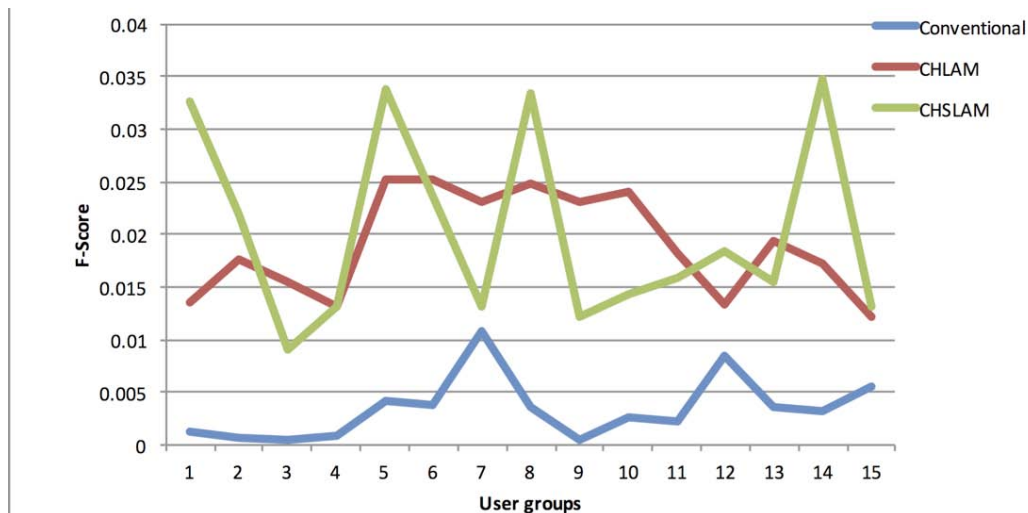


Figure 6. F-Score values for Conventional, CHLAM and CHSLAM of OCSL algorithms

The F-score shows an absolute score for an algorithm that strives for good balance between the precision and recall. Figure 6 shows that the learners engaged more successfully based on the CHSLAM algorithm compared to the CHLAM and conventional algorithms. The F-Score can be interpreted as a weighted average of the precision and recall, where an F-Score reaches its best value at 1 and worst at 0. The average F-Score value for conventional algorithm was 0.0034, and for CHLAM algorithm it was 0.0190 and for CHSLAM algorithm it was 0.0203. Based on the tests, CHSLAM algorithm yielded better F-Score results. To obtain a subjective evaluation of the OCSL system, we organized a non-mandatory questionnaire that collected information on learners with respect to the main features of the system. More than 65% of the learners reported that the system recommended personalized results and was able to focus on the correct content. Overall, the system showed remarkable improvement in self-learning. The learners were able to focus more time on studying the correct content and less time on searching for the content.

## CONCLUSIONS

We presented a design and implementation of an end-to-end implementation model and conducted several experiments to test our system. Our system starts with a clustering engine that processes the content from various OER sources to properly map it to the taxonomy we built to support STEM (science, technology, engineering, and mathematics) content. It then generates personalized search results based on the content hierarchy (e.g., content type, content category) and learner attributes (e.g., learning style, recent activity). We took the learner experience from the logs and database and plotted them in aggregate on a precision versus recall graph, which showed whether the algorithm made the correct prediction based on the learner’s historical behavior as well as similar learners’ learning behaviors. Here, the precision and recall are not perfectly negatively correlated; there is a natural tension between the two in such a way that improvements in one often lead to declines in the other. We found that a search that was based on the historical learning of learners and similar learners’ behaviors (CHSLAM of OCSL) yielded better F-Score results compared with the conventional search as well as a search based only on Content Hierarchical and Learning Attribute-based Learning (CHLAM). In the future, we plan to expand the system by creating peer groups with complex algorithms by leveraging similar learners’ data from OCSL. We will explore extending the personalized mechanism and pedagogical aspects of OCSL to increase the engagement of learners by having the influencers and mentors interact with the peer group.

## REFERENCES

- Antunes, C., & Oliveira, A. (2002, July). Using context-free grammars to constrain apriori-based algorithms for mining temporal association rules. In Proc. Workshop on Temporal Data Mining.
- Banerjee, A. V., & Duflo, E. (2014). (Dis) organization and Success in an Economics MOOC. *The American Economic Review*, 104(5), 514-518.
- Bansal, T., Chabra, S., & Joshi, D. (2013). Current Initiatives and Challenges to OERs in Indian Higher Education.
- Barrett, R., Maglio, P. P., & Kellem, D. C. (1997, March). How to personalize the Web. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (pp. 75-82). ACM.
- Bledsoe, T. S., Harmeyer, D., & Wu, S. F. (2014). Utilizing Twitter and# Hashtags Toward Enhancing Student Learning in an Online Course Environment. *International Journal of Distance Education Technologies*

- (IJDET), 12(3), 75-83. doi:10.4018/ijdet.2014070106
- Brin, S., & Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18), 3825-3833.
- Brusilovsky, P., Kobsa, A., & Nejdl, W. (2007). *The adaptive web: methods and strategies of web personalization* (Vol. 4321). Springer Science & Business Media.
- Chan, P. K. (2000). Constructing web user profiles: a non-invasive learning approach. In *Web usage analysis and user profiling* (pp. 39-55). Springer Berlin Heidelberg.
- Cuéllar, M. P., Delgado, M., & Pegalajar, M. C. (2011). Improving learning management through semantic web and social networks in e-learning environments. *Expert Systems with Applications*, 38(4), 4181-4189.
- Cheung, K. S., Lam, J., Im, T., Szeto, R., & Yau, J. (2008, December). Exploring a pedagogy-driven approach to e-courses development. In *Education Technology and Training, 2008. and 2008 International Workshop on Geoscience and Remote Sensing. ETT and GRS 2008. International Workshop on* (Vol. 1, pp. 22-25). IEEE.
- Cheung, B. S., Hui, L. C. K., Yiu, S. M., Lee, J. K., Kwok, S. L., & Leung, K. C. (2003). Content Engineering Agent: A TBL-Based E-Course Development Tool with TQM. *IJDET*, 1(2), 57-71. doi:10.4018/978-1-59904-935-9.ch096
- Chiou, Y., & Shih, T. K. (2015). Auto Grouping and Peer Grading System in Massive Open Online Course (MOOC). *International Journal of Distance Education Technologies (IJDET)*, 13(3), 25-43. doi:10.4018/IJDET.2015070102
- D'Antoni, S. (2009). Open Educational Resources: reviewing initiatives and issues. *Open Learning: The Journal of Open, Distance and e-Learning*, 24(1), 3-10.
- de Waard, I., Koutropoulos, A., Keskin, N., Abajian, S. C., Hogue, R., Rodriguez, O., & Gallagher, M. S. (2011). Exploring the MOOC format as a pedagogical approach for mLearning. *Proceedings from mLearn*.
- Devi, P., Gupta, A., & Dixit, A. (2014). Comparative Study of HITS and PageRank Link based Ranking Algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 3(2).
- Di Noia, T., Mirizzi, R., Ostuni, V. C., Romito, D., & Zanker, M. (2012, September). Linked open data to support content-based recommender systems. In *Proceedings of the 8th International Conference on Semantic Systems* (pp. 1-8). ACM.
- Drachler, H., Hummel, H. G., & Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. *International Journal of Learning Technology*, 3(4), 404-423.
- Fardinpour, A., Pedram, M. M., & Burkle, M. (2014). Intelligent Learning Management Systems: Definition, Features and Measurement of Intelligence. *International Journal of Distance Education Technologies (IJDET)*, 12(4), 19-31. doi:10.4018/ijdet.2014100102
- Ferguson, R., & Sharples, M. (2014). Innovative pedagogy at massive scale: teaching and learning in MOOCs. In *Open Learning and Teaching in Educational Communities* (pp. 98-111). Springer International Publishing.
- Ganesh, S., Jayaraj, M., Kalyan, V., Murthy, S., & Aghila, G. (2004, April). Ontology-based web crawler. In *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on* (Vol. 2, pp. 337-341). IEEE.
- Gauch, S., Chaffee, J., & Pretschner, A. (2003). Ontology-based personalized search and browsing. *Web Intelligence and Agent Systems*, 1(3-4), 219-234.
- Goodwin, B., & Miller, K. (2013). Evidence on flipped classrooms is still coming in. *Educational Leadership*, 70(6), 78-80.
- Guarino, N., Masolo, C., & Vetere, G. (1999). Ontoseek: Content-based access to the web. *Intelligent Systems and Their Applications*, IEEE, 14(3), 70-80.
- Hanley, G. L. (2015). MOOCs, MERLOT, AND OPEN EDUCATIONAL SERVICES. *MOOCs and Open Education Around the World*, 33.
- Hafidi, M., & Bensebaa, T. (2015). Architecture for an Adaptive and Intelligent Tutoring System that Considers the Learner's Multiple Intelligences. *International Journal of Distance Education Technologies (IJDET)*, 13(1), 1-21. doi:10.4018/ijdet.2015010101
- Holotescu, C., Grosseck, G., Cretu, V., & Naaji, A. (2014, October). Integrating MOOCs in Blended Courses. In *The International Scientific Conference eLearning and Software for Education* (Vol. 4, p. 243). "Carol I" National Defence University.
- Iiyoshi, T., & Kumar, M. V. (2008). *Opening up education: The collective advancement of education through open technology, open content, and open knowledge*. Mit Press.
- Khosla, S., & Bhojane, V. (2013). Performing Web Log Analysis and Predicting Intelligent Navigation Behavior Based on Student Accessing Distance Education System. In *Advances in Computing, Communication,*

- and Control (pp. 70-81). Springer Berlin Heidelberg.
- Kim, B., & Reeves, T. C. (2007). Reframing research on learning with technology: In search of the meaning of cognitive tools. *Instructional Science*, 35(3), 207-256.
- Kim, J. K., Kim, H. K., & Cho, Y. H. (2008). A user-oriented contents recommendation system in peer-to-peer architecture. *Expert Systems with Applications*, 34(1), 300-312.
- Klašnja-Milićević, A., Vesin, B., Ivanović, M., & Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education*, 56(3), 885-899.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5), 604-632.
- Kumari, T., Gupta, A., & Dixit, A. (2014). Comparative Study of Page Rank and Weighted Page Rank Algorithm. *Proceedings of International Journal of Innovative Research in Computer and Communication Engineering*, 2(2).
- MacDonald, M. (2015). The Battle for Open by Martin Weller. *Journal Of Perspectives In Applied Academic Practice*, 3(1). doi:10.14297/jpaap.v3i1.139
- Malloy, T. E., & Hanley, G. L. (2001). MERLOT: A faculty-focused Web site of educational resources. *Behavior Research Methods, Instruments, & Computers*, 33(2), 274-276.
- McGreal, R., Kinuthia, W., Marshall, S., & McNamara, T. (2013). Perspectives on Open and Distance Learning: Open Educational Resources: Innovation, Research and Practice.
- Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. *MERLOT Journal of Online Learning and Teaching*, 9(2).
- Monge, S., Ovelar, R., & Azpeitia, I. (2008). Repository 2.0: Social dynamics to support community building in learning object repositories. *Interdisciplinary Journal of E-Learning and Learning Objects*, 4(1), 191-204.
- Ng'ambi, D., & Bozalek, V. (2015). Editorial: Massive open online courses (MOOCs): Disrupting teaching and learning practices in higher education. *British Journal of Educational Technology*, 46(3), 451-454. doi: 10.1111/bjet.12281
- O'Donnell, E., Lawless, S., Sharp, M., & Wade, V. P. (2015). A Review of Personalised E-Learning: Towards Supporting Learner Diversity. *International Journal of Distance Education Technologies (IJDET)*, 13(1), 22-47. doi:10.4018/ijdet.2015010102
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: bringing order to the Web.
- Paireekreng, W., & Wong, K. W. (2010, January). Mobile content personalisation using intelligent user profile approach. In *Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on* (pp. 241-244). IEEE.
- Parr, C. (2013). MOOC completion rates 'below 7%,'. *Times Higher Education*, 9.
- Passant, A., Samwald, M., Breslin, J., & Decker, S. (2009). Federating distributed social data to build an interlinked online information society.
- Patil, A. S., & Pawar, B. V. (2012, March). Automated classification of web sites using Naive Bayesian algorithm. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 1).
- Porcello, D., & Hsi, S. (2013). Crowdsourcing and curating online education resources. *Science*, 341(6143), 240-241.
- Quercia, D., Askham, H., & Crowcroft, J. (2012, June). TweetLDA: supervised topic classification and link prediction in Twitter. In *Proceedings of the 4th Annual ACM Web Science Conference* (pp. 247-250). ACM.
- Rai, L., & Chunrao, D. (2016). Influencing Factors of Success and Failure in MOOC and General Analysis of Learner Behavior. *International Journal of Information and Education Technology*, 6(4), 262.
- Rensing, C., de Freitas, S., Ley, T., & Muñoz-Merino, P. J. (Eds.). (2014). *Open Learning and Teaching in Educational Communities: 9th European Conference on Technology Enhanced Learning, EC-TEL 2014, Graz, Austria, September 16-19, 2014, Proceedings* (Vol. 8719). Springer.
- Sathiyamurthy, K., & Geetha, T. V. (2012). Automatic Organization and Generation of Presentation Slides for E-Learning. *International Journal of Distance Education Technologies (IJDET)*, 10(3), 35-52.
- Schmidt, S., Scholl, P., Rensing, C., & Steinmetz, R. (2011). Cross-lingual recommendations in a resource-based learning scenario. In *Towards Ubiquitous Learning* (pp. 356-369). Springer Berlin Heidelberg.
- Şenyuva, E., & Taşocak, G. (2014). Implementation of Web-Based Distance Education in Nursing Education in Turkey: A Sample Lesson in Patient Education. *International Journal of Distance Education Technologies (IJDET)*, 12(3), 1-13.
- Siemens, G. (2005). *Connectivism: A Learning Theory for the Digital Age* <http://www.elearnspace.org/Articles/connectivism.htm> (Accessed: 01/03/2007).
- Sharples, M., Adams, A., Ferguson, R., Gaved, M., McAndrew, P., Rienties, B., ... & Whitelock, D. (2014).



- Innovating Pedagogy 2014.
- Smyth, R., Bossu, C., & Stagg, A. (2015). Toward an Open Empowered Learning Model of pedagogy in higher education. *Open learning and formal credentialing in higher education: Curriculum models and institutional policies*. Hershey: IGI Global.
- Stern, M. K., & Woolf, B. P. (2000). Adaptive content in an online lecture system. In P. Brusilovsky, O. Stock, & C. Strapparava (Ed.), *Adaptive Hypermedia and Adaptive Web-based systems* (pp. 225-238). Berlin: Springer-Verlag.
- Stacey, P. (2013). The Pedagogy Of MOOCs. *The International Journal for Innovation and Quality in Learning*.
- Tang, J., Wu, S., Sun, J., & Su, H. (2012, August). Cross-domain collaboration recommendation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1285-1293). ACM.
- Toledo, R. Y., & Mota, Y. C. (2014). An e-Learning Collaborative Filtering Approach to Suggest Problems to Solve in Programming Online Judges. *International Journal of Distance Education Technologies (IJDET)*, 12(2), 51-65.
- Town, C. (2007). The Cape Town open education declaration: Unlocking the promise of open educational resources. Retrieved October, 28, 2010.
- Vassileva, J. (2008). Toward social learning environments. *Learning Technologies, IEEE Transactions on*, 1(4), 199-214.
- Wanaskar, U., Vij, S., & Mukhopadhyay, D. (2013). A Hybrid Web Recommendation System Based on the Improved Association Rule Mining Algorithm. arXiv preprint arXiv:1311.7204.
- Wang, L. C., Meng, X. W., & Zhang, Y. J. (2012). Context-aware recommender systems. *Ruanjian Xuebao/Journal of Software*, 23(1), 1-20.
- Wang, J., Kraska, T., Franklin, M. J., & Feng, J. (2012). Crowder: Crowdsourcing entity resolution. *Proceedings of the VLDB Endowment*, 5(11), 1483-1494.
- Yair, Y. (2014). Open educational resources: reasons to be cheerful?. *ACM Inroads*, 5(4), 37-38. <http://doi.acm.org/10.1145/2684721.2684729>
- Zaïane, O. R., Li, J., & Hayward, R. (2006). Mission-based navigational behaviour modeling for web recommender systems. In *Advances in Web Mining and Web Usage Analysis* (pp. 37-55). Springer Berlin Heidelberg.