COMPARING QUANTITATIVE AND COMMENT-BASED RATINGS FOR RECOMMENDING OPEN EDUCATIONAL RESOURCES

Dalia Hanna
Department of Computer Science
Ryerson University
350 Victoria Street
Toronto, ON, CANADA
dhanna@ryerson.ca

Abdolreza Abhari, PhD
Department of Computer Science
Ryerson University
350 Victoria Street
Toronto, ON, CANADA
aabhari@ryerson.ca

Alexander Ferworn, PhD
Department of Computer Science
Ryerson University
350 Victoria Street
Toronto, ON, CANADA
aferworn@ryerson.ca

ABSTRACT

The recent application of recommender systems for educational resources and e-learning has facilitated online and accessible education on social networks. However, there are currently few studies about the methods for evaluation and performance measurement of these recommender systems in the complicated environment of educational and social networking platforms. The purpose of this research paper is to investigate the effectiveness of using sentiment analysis methods for educational resources based on user comments and compare it with the quantitative approach based on user rating to recommend best open learning resources (OER) available through online OER repositories. The quality of the OER will be justified by comparing the user rating and the users' reviews. The quality of users' reviews is based on calculating the term frequency for selected positive and negative terms, then determining the similarity among the comments. Comments with positive or negative words confirm the high and low ratings respectively.

Keywords: Open Education Resources, OER, Recommender Systems, Sentiment Analysis.

1 INTRODUCTION

According to the UNESCO website (2016), Open Education Resources (OERs) are educational materials that are available in the public domain; anyone can copy, re-share, and use these OERs without permission. There are several types of OERs including textbooks, articles, test, projects, videos, notes and multimedia. Currently, there are millions of OERs on the web; some of these resources are available for free and some are available to use under creative common license with some restrictions. To illustrate, a database may impose a restriction on using a substantive portion of the OER for commercial use. In some repositories, users may provide a numeric rating for each OER used and in others users may provide comments to describe what they liked or disliked about the OER. The UNESCO website lists 367 links to repositories from around the world. Each repository may include thousands of records. Some have numer-

ic rating only, some with comments and some without any user rating at all. With millions of available OERs, there is a need to evaluate their quality and their use in teaching and learning. By comparing the user ratings and the associated comments, we could confirm the quality of the OER in a certain subject area. Then, users will have a more informed input about the quality of the OER and recommend its use to the other users.

With the increased number of online repositories, there is a need to develop methodologies to evaluate the quality of these OERs. For example, through engagement on social media, users could provide comments that could promote and recommend an OER, even when there is no rating system provided to support its quality. Additionally, educators may depend on user ratings to choose appropriate OERs for their courses, however, supporting their choice with well-informed user comments enriches their decisions. Moreover, understanding the usefulness of a comment in a specific subject area may help students in selecting a solid learning path that is based on evidence of quality in this subject area.

In this study, we used Multimedia Educational Resource for Learning and Online Teaching (MERLOT), which is a peer-reviewed repository with over 78,000.00 OERs. MERLOT utilizes both quantitative and qualitative means for feedback as well as an expert peer rating for each OER. We start with a literature review of some studies that used MERLOT and other used social media, like blogs and Facebook, to recommend educational resources to users. We then, in the methodology section, use sample data from high rated OERs and low rated OERs and compare these rating with their comments. We use sentiment analysis to analyze the comments as described in section 2.

1.1 Literature Review

(Leskovec et al. 2014) introduced two methods for recommender systems: content-based and collaborative filtering. Content-based systems provide recommendations to users based on user ratings of other similar items. The main advantage of this approach is that it highlights the features that caused the item to be recommended; a disadvantage is that it is hard to find an appropriate feature. Alternatively, collaborative filtering approaches recommend items based on user's preference and what they liked. One of the disadvantages of this approach is that it can't recommend an item if that item was not rated or liked before.

(Karampiperis et al. 2014) used the MERLOT repository as an example for their recommender system and analyzed the feedback about the peer reviews and users' comments using collaborative filtering recommender system. The study analyzed the recommendations that are based on the users' opinion which are listed as comments (text) on the website and also as a rating system. The process started first with classifying the ratings, where 1 to 2 represents a negative opinion, 3 is neutral and 4 to 5 as positive. Then, defined the terms in a corpus (T) and assigned scores to the positive terms as (PosScore) and negative terms as (NegScore). Finally, applying the sentiment analysis process by preparing the corpus, eliminating stop words, then building the training and evaluation data sets using quality centric collaborative filtering recommender.

The calculation algorithm used is listed below:

$$Sentiment\ (T, P, N) = \frac{5}{2} \left(1 + \frac{PosScore(T, P) - NegScore(T, N)}{\max\{PosScore(T, P), NegScore(T, N)\}} \right)$$

(Karampiperis et al. 2014, p.19)

Where *P* is the term with positive popularity, *N* is the term with negative popularity and *T* is the set of text that will be examined.

The study introduced qualitative information into collaborative filtering recommender system over social platforms. Additionally, it concluded that the quality of a resource is better evaluated by using rich analysis provided about the user-provided comments. It indicated a limitation with regards to the ratio of positive to negative ratings as high rated OERs had more comments than the low rated OERs.

Cechinel et al. (2014) performed two experiments using the educational objects posted on the MERLOT website as well. In the first experiment, Data Mining Classification Algorithms (DMCA) were used to classify the resources based on their ratings. The best model produced lower Mean Absolute Error (MAEs) and higher kappa coefficients in classifying resources as good, average and poor. In the second experiment, they used the resource's statistical profile as well but included other collected metrics of the material type and its relation to a specific discipline through the use of Artificial Neural Networks (ANN). The outcome of these experiments depended on the method used, with a conclusion that neural networks are best to be used when there is no need for reasoning when using a certain model. More studies are needed by using the proposed method on other repositories.

A study by Perryman and Coughlan (2013) looked at six Commonwealth groups using Facebook as a bulletin board for their activities. These groups used Facebook to share their object-focused behaviours, which is indicated by sharing documents and links and relationship-focused behaviours which is shown by the number of likes and comments on a post. They concluded that it is possible to use Facebook to broaden the access to OERs based on the comments and ratings.

Lopez-Vargas, Piedra, Chicaiza, and Tovar (2015) developed a recommender system based on social network analysis (SNA) to find educational resources posted on Twitter. Their research helped in informing the efforts of collaborators in the StartUp project (https://www.facebook.com/StartupProject/) which is a web space to support young entrepreneurs in launching their start-ups. The methodology used: (1) identify a topic list, (2) develop queries, (3) extract data from Twitter using TAW Crawler, (4) normalize the data, (5) enhance the URLs associated with the OERs, (6) filter the OERs, and finally (7) rank the OERs. Their research has shown that it is feasible to obtain OERs from social media networks to recommend resources for young entrepreneurships seeking education in start-ups.

Liu and Qin (2016) used clustering methods to detect the circle of friends in social media microblogs for a specific user. In this study, k-means algorithm was used as a baseline method in detecting the community. Clustering is subjective as it is based on the business need or model. Partitioned clustering (k-means) which starts with identifying the desired number of clusters (K), assign objects to the nearest centre and then continue to re-assign the objects to a closer cluster until there the object no longer can move (Lin n.d.). Various clustering algorithms are widely used in search engines queries such as Google. Unlike classification methods, clustering doesn't include pre-classified data, rather it investigates a pattern of similar behaviour of a group and how it maybe dissimilar to another group (Elmasri and Sham 2016). The method used by Liu et al. provided the common methodology for analysing comments therefore we used it to investigate its effectiveness for processing of OER comments as described as our methodology listed below in section 2.

Clustering based on similarity between the text already has been used on social networks. For example, in our research lab http://dsmp.ryerson.ca (Ahmed and Abhari 2014, Almeshari et al. 2013) we used text based analysis to cluster users' tweets on the social network "Twitter." In this work, we propose a method for clustering comments related to online educational resources.

2   METHODOLOGY

2.1   Select Data

We used the data in the MERLOT website. The data listed on the website as of November 10, 2016 is as follows:

- Number of Resources (OERs): 78,142.
- Number of Members: 143,331.
- Number of Resources with User Comments: 6,046.
- Number of Resources with User Ratings: 8,355.
- Bookmark collection: 19,530 (Objects grouped into categories).
- Number of Comments: 25,832.

2.2   Prepare Query and Data List

We selected the records with both user ratings and comments to be able to compare both ways of ratings. The number of the sample data selected is 8355 OER. The user rating is between 1 to 5, with 1 indicating poor quality and thus low recommendation and 5 is highly recommended. The data includes the OER title, description, main category, overall rating by users. This dataset represents OERs with ratings and could be ranked from highest to lower. At the time writing the paper we didn't have access to the complete data of user comments associated with the selected OERs, therefore we selected the OERs having comments from the website. Based on Cechinel et al. (2014) method, we selected comments associated with OERs in a certain subject area which is "Science and Technology" as listed in Table 1. It is clear from Table 1 that there are much more comments associated with the high ranked resources that the low ranked resources.

Table 1: Range of ratings in the selected data Set from the MERLOT website in the subject area of "Science and Technology" with a total of 2898 OER.

Range of Ratings	Number of Associated OERs
Ratings form 4 to 5	2158
Ratings from 2.1 to 3.9	552
Ratings from 1 to 2	188

2.3   Selection of Sample Data

We selected a sample data of five OERs from the high ratings list, which are OERs with average user ratings between 4 and 5. Table 2 lists the number of comments for each OER selected and its overall average rating by the users. The total number of comments selected randomly constitute the corpus in our sample of high ratings which is 112 comments with an overall average rating of 4.58.

Table 2: Number of comments associated with the OERs with high average user rating (between 4 and 5).

OER #	Number of Users Provided Ratings	Overall Numeric Rating	Number of Users Provided Comments
1	41	4.7	11
2	175	4.54	55
3	14	4.67	12
4	56	5	17
5	57	4	17
Total	**343**	**4.58**	**112**

Similarly, we selected five OERs from the low ratings list, with overall average user rating of 1 to 2. Table 3 lists the number of comments associated with the five OERs selected. The total number of comments selected randomly, which constitute the corpus in our sample of low ratings corpus of nine comments with an overall average rating of 1.1. Table 3 clearly illustrates the lake of comments associated with the low rated OERs.

Table 3: Number of comments associated with the OERs with low average user rating (between 1 and 2).

OER #	Number of Users Provided Ratings	Overall Rating	Number of Users Provided Comments
6	3	1.67	4
7	2	1.5	1
8	1	1	1
9	1	1	1
10	1	2	1
Total	8	1.4	8

2.4 Identify the Positive and Negative Terms

2.4.1 Positive Terms

As defined in Elmasri and Sham (2016) the term frequency (TF) is how frequently a term appears in a document. TF are weighted relative to their frequency counts in the document to reflect the rank of terms in the calculation based on their relevance degree. Using a TF calculator developed in python and then aggregated in Excel, we calculated the Term Frequency (TF) in the corpus of the sample data for the five selected OERs with a total number of 112 comments as listed in Table 2. Then, we removed the stop words and identified the positive terms that indicated a positive reaction towards the OER. We select the positive words based on the affective learning domain which describes the manner we interact with objects emotionally, such as values, feelings, appreciation, motivations, enthusiasms, and attitudes (Clark 2015). Figure 1 provides a visual representation of all terms in the entire corpus that consist of all comments associated with the selected OERs. We identified six words that represent a positive reaction about a resource. The words selected are useful, great, excellent, recommend, helpful and good. Table 4 lists these terms and its associated frequencies in the corpus. The terms are highlighted as red dots in Figure 1.
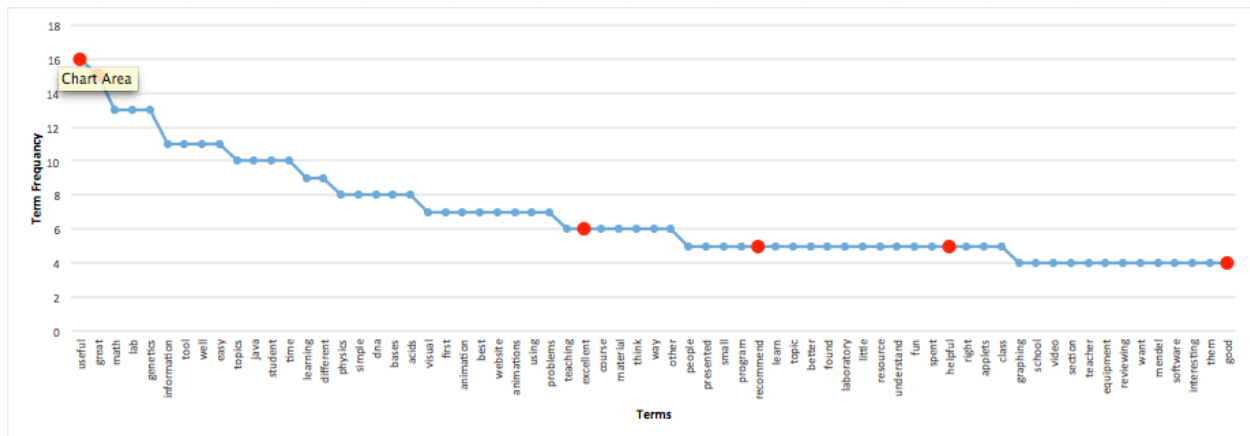

Figure 1: Term frequencies of high rated OERs in the corpus.

Table 4: Term frequency in the corpus which was selected for the high average rates between 4 to 5.

Term	Frequency
useful	17
great	15
excellent	6
recommend	7
helpful	5
good	4

2.4.2 Negative Terms

Similarly, we repeated the same process to select the negative terms in the corpus in which we selected four negative terms used to indicate negative reaction about a resource. To be consistent with the previous process. We selected the affective learning domain words, thus eliminating the words "no" and "not" as these words are used to describe other words rather than provide reaction or a value. In addition, part of the corpus cleaning process was to eliminate the stop words with three characters or less. We used a python program to extract the terms from the corpus and imported the result to Excel for aggregation. Figure 2 provides visual representations of the corpus with low ratings and Table 5 lists the selected negative terms in the corpus.
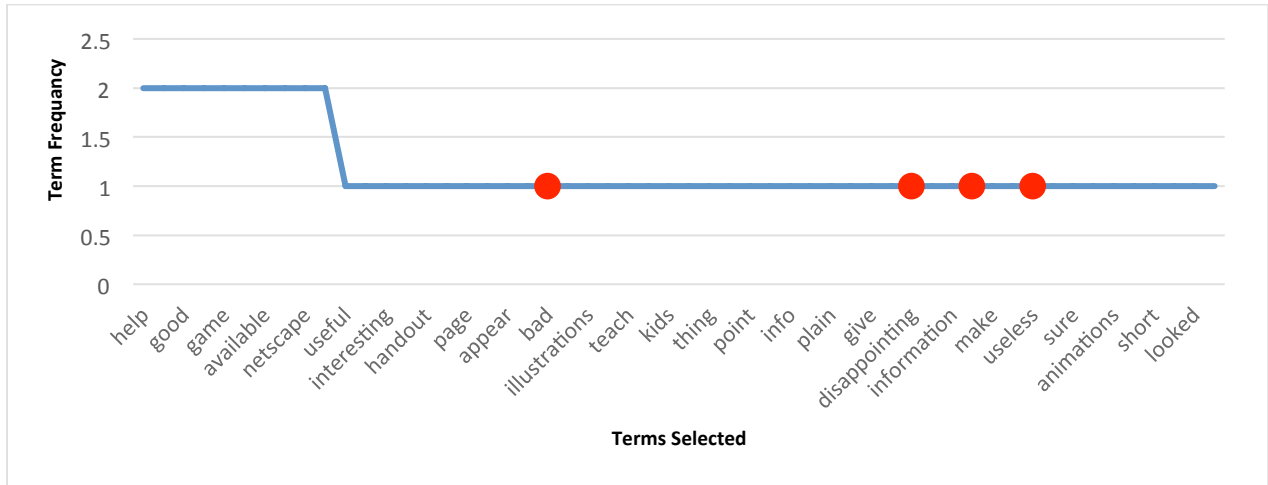

Figure 2: List of all terms in the corpus with comments associated with low user rating

Due to the very few comments for the low average ratings, the data sample indicates low number of term frequency of negative words in the corpus, which is appearing only once as listed in Table 5. This may affect the comparison to the average rating as it may not be enough data to provide a recommendation or an opinion about the object. (Karampiperis et al. 2014) experienced the same challenge in their study. Based on this, we will only continue with investigating the positive terms going forward.

Table 5: Term Frequency in the entire corpus which was selected form the low average rating which is between 1 and 2.

Term	Term Frequency
bad	1
disappointing	1
terrible	1
useless	1

2.5 Clustering by Calculation of Similarity

Elmasri and Sham (2016) stated that the use of the cosine similarity function is helpful in calculating the similarity when we are taking into account the frequency of words in a corpus to find the matching weights. As this is a fixed collection of documents, then the TF and inverse document frequency IDF values can be easily calculated.

First, we calculate TF.IDF as listed in equation (1) which is taking the important words in the document where TF is the number of terms in the document and IDF is the number of documents counted in searched collection and are indexed by term.

$$tf*idf(t,d,D) = tf(t,d)*idf(t,D) \tag{1}$$

To calculate the Term Frequency (TF) in the selected corpus of 112 comments, we ran a python code to calculate the TF of all terms in the corpus, normalized the outcome and then calculate IDF using equation (2). In our sample data, the data set has five objects, which means that N of the sample data is equal to five.

$$IDF(t,D) = Log2(N/df) \tag{2}$$

Second, we used the cosine similarity equation to calculate the angle between 2 vectors. Therefore, the calculated similarity would be between 0 and 1, where 1 indicates that the two vectors are exactly the same as the cosine of 0 is 1. The values were calculated using a python code and verified manually.

$$A \cdot B = \|A\|\|B\| \cos \theta \tag{3}$$
$$cosSim(A,B) = \vec{V}(A) \cdot \vec{V}(B) / |\vec{V}(A)\|\vec{V}(B)|$$
$$= \sum Ai \times Bi \; n \; i{=}1 \; \sqrt{\sum (Ai)} \; n \; 2 \; i{=}1 \times \sqrt{\sum (Bi)} \; n \; 2$$

After, we calculated the quantitative similarity, which is based on user ratings between same comments. We used manual calculations to be able to verify that the process is correct. Table 6 lists the both the ratings and the comments dissimilarity. Figure 3 illustrates the dissimilarity among the same OERs using user ratings.

Table 6: Dissimilarity between the ratings and comments.

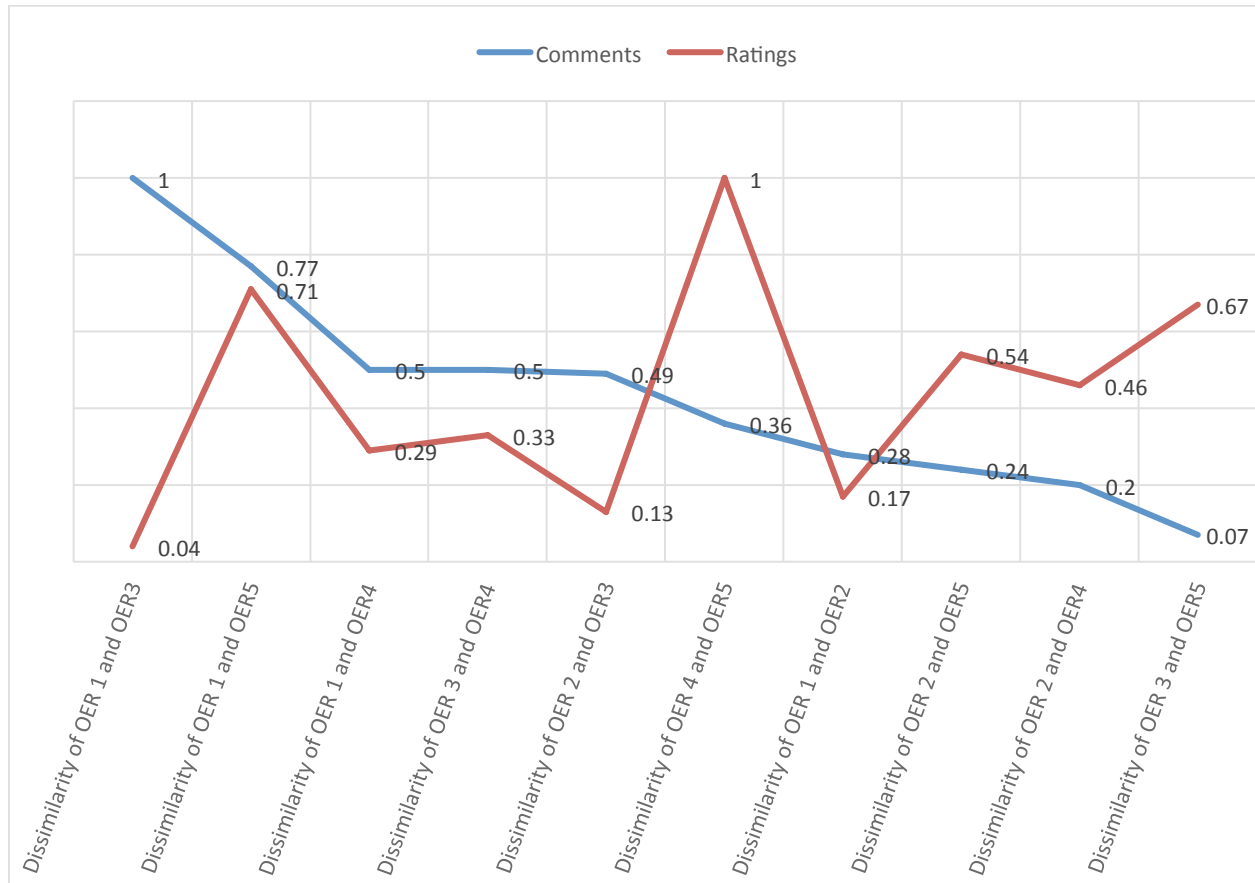| Dissimilarity	Comments	Ratings
Dissimilarity of OER 1 and OER3	1	0.04
Dissimilarity of OER 1 and OER5	0.77	0.71
Dissimilarity of OER 1 and OER4	0.5	0.29
Dissimilarity of OER 3 and OER4	0.5	0.33
Dissimilarity of OER 2 and OER3	0.49	0.13
Dissimilarity of OER 4 and OER5	0.36	1
Dissimilarity of OER 1 and OER2	0.28	0.17
Dissimilarity of OER 2 and OER5	0.24	0.54
Dissimilarity of OER 2 and OER4	0.2	0.46
Dissimilarity of OER 3 and OER5	0.07	0.67

Figure 3: Chart indicating the dissimilarity of comments and dissimilarity of ratings in the selected corpus of 112 comments with 5 OERs.

From Table 6 and Figure 3 we realize that OER1 has no similarity with OER3 in term of the comments, however, they are very close in terms of the rating with a 96% similarity. In contrast, OER4 has less similarity with OER5 in terms of ratings with 64% similarity in comments. These results could be verified by comparing the overall user ratings listed in Table 2.

2.6 Cluster Formation

As there is no training sample, then the process is defined as an unsupervised learning (Elmasri and Sham 2016). Therefore, we cluster the user comments using k-means assuming k = 3 using a python program. Then, we manually investigate the data associated with the lists in Table 2 and cluster the comments and ratings to compare results. For the comments, OER2 and OER5 are in a cluster as both has all terms, OER3 is in its own cluster as it only has the one term and OER1 and OER4 are in another cluster with few words in common. For the ratings, OER1, OER2 and OER3 are all in one cluster and OER4 and OER5 are each in its own clusters. There is a difference in the clustering by comments and by ratings.

3 CONCLUSION

By investigating the positive terms in user comments manually and comparing it with our findings, we were able to verify the quality of an OER. However, the methodology needs to be applied to the larger data set. This may enhance the outcome for the low user ratings, which we were not able to investigate due to the small number of comments associated with it. Further studies needed to directly relate the reaction to the OER and the subject area. Also, it would be helpful to investigate users' comments and find similarity among users on social media and the online repositories in which many educators and students engage and can develop a learning community.

In this paper, a method for classification of educational resources based on user comments is provided that uses TFIDF of the selected words and k-means clustering. In the highly ranked educational resources we found that for some resources sometimes there is a lack of ratings or lack of comments; sometimes even disagreement between comments and rating. Using this method to process comments and recommend educational resources in a specific subject area, will help students pursuing online learning or users engaging in educational social network environments.

REFERENCES

Almeshary, M., and A. Abhari, "A recommendation system for Twitter users in the same neighbourhood". *In Proceedings of the 16th Communications & Networking Simulation Symposium (CNS 2013)*. San Diego, California, The Society for Modeling and Simulation International (SCS).

Ahmed, L., and A. Abhari. 2014. "Agent-Based Simulation of Twitter for Building Effective Recommender System". *In Proceedings of the 17th Communications & Networking Simulation Symposium (CNS)*. San Diego, California, The Society for Modelling and Simulation International (SCS).

Ahmed, L., and A. Abhari. 2015. "Distributed Recommender System for Online Processing of Big Social Data". *In Proceedings of the Poster Session and Student Colloquium Symposium (Posters '15), pp. 15–16*. San Diego, California, The Society for Modeling and Simulation International (SCS).

Cechinel, C., S. da Silva Camargo, S. Sanchez-Alonso, and M.A. Sicilia. "Towards Automated Evaluation of Learning Resources Inside Repositories" In *Recommender Systems for Technology Enhanced Learning, edited by* H. Drachsler, N. Manouselis, O. C. Santos, K. Verbert. pp. 25-46. New York: Springer, 2014.

Clark, D.R. "Bloom's Taxonomy: The Affective Domain." http://www.nwlink.com/~ Donclark/hrd/Bloom/affective_domain.html. Accessed February 5, 2017.

Elmasri, R. and N. Sham. *Fundamentals of Database Systems*. 6th ed. (New Jersey: Pearson, 2016).

Karampiperis, P., A. Koukourikos and G. Stoitsis. "Collaborative Filtering Recommendation of Educational Content in Social Environments Utilizing Sentiment Analysis Techniques." In *Recommender Systems for Technology Enhanced Learning*, edited by H. Drachsler, N. Manouselis, O. C. Santos, K. Verbert. pp. 3-23. New York: Springer, 2014.

Lin, H. n.d. "Clustering". http://www.cs.cmu.edu/afs/andrew/course/15/381-f08/www/lectures/clustering.pdf . Accessed December 1, 2016.

Liu, T., and H. Qin. "Detecting and Tagging Users' Social Circles in Social Media." *Multimedia Systems,* 22, no. 4 (2016): 423-431. doi:10.1007/s00530-014-0435-4

J. Lopez-Vargas, N. Piedra, J. Chicaiza and E. Tovar, "OER Recommendation for Entrepreneurship Using a Framework Based on Social Network Analysis," in *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 10, no. 4, pp. 262-268, Nov. 2015. doi: 10.1109/RITA.2015.2486387

Perryman, L. and T. Coughlan. Using Open Educational Resources in Online Welfare Communities: Facebook as an Arena for Open Scholarship across the Commonwealth. http://oasis.col.org/bitstream/handle/11599/1849/2013_Perryman%26Coughlan_OEROnlineWelfare.pdf?sequence=1&isAllowed=y. Accessed November 11, 2016.

UNISCO.org. (2016). What are Open Educational Resources (OERs)?. http://ww2.unisco.org

AUTHOR BIOGRAPHIES

DALIA HANNA is a Ph.D. student in the Department of Computer Science at Ryerson University, Toronto, Canada. She holds a M.Sc. in Instructional Design and Technology with specialization in online learning form Walden University and a B.Sc. in Electronics and Communication Engineering with major in software programming from Cairo University. She is currently a Program Director in The Chang School of Continuing Education. Her research interests lie in teaching and learning in higher education as well as public safety especially in healthcare. Her email address is dhanna@ryerson.ca.

ABDELREZA ABHARI is a Professor in the Department of Computer Science at Ryerson University. He holds a Ph.D. in Computer Science from Carleton University. His research interests include data science, Web 2.0 social networking, web mining and information retrieval, data mining and database systems, big data analysis, sensor networks and distributed systems, soft computing and fuzzy logic, modeling and simulation. His email address is aabhari@ryerson.ca.

ALEX FERWORN is a Professor in the Department of Computer Science at Ryerson University. He holds a Ph.D. in System Design Engineering from University of Waterloo, Canada. His research interests include Computational public safety: Urban Search and Rescue (USAR) and Chemical, Biological, Radiological and Nuclear explosives (CBRNe) applications; mobile, autonomous and teleoperated robotics; artificial intelligence and network applications. His email address is aferworn@ryerson.ca.