Taxonomy of Recommender Systems for Educational Data Mining (EDM) Techniques: A Systematic Review

Mosima A. Masethe, Sunday O. Ojo, and Solomon A Odunaike, *Member, IAENG*

***Abstract*— The educational management decision-makers (EMDMs) have no clear understanding of the available EDM techniques and variables to consider in selecting EDM techniques appropriate for their decision making needs. This research proposes taxonomy of EDM techniques through utilisation of recommender systems (RSs) to address EDM technique selection challenges. The RS approach addresses the need for a decision support system guide process of selecting appropriate EDM technique. Furthermore, the research presents systematic review of different approaches in recommender system such as content-based, collaborative filtering, hybrid, and knowledge-based recommender system. The research study looks at current existing challenges of different recommender system including proposed technique to solve the problems. Knowledge-based recommender system seems to solve problems encountered by content-based recommender system and collaborative filtering recommender system. The research further discus how knowledge-based RS notable employs case-based reasoning and ontology-based engineering to overcome challenges such as cold-start, scalability, sparseness, grey sheep, contend limitations, overspecialization, and inflexible information.**

***Index Terms*—Educational Data Mining (EDM), Recommender System (RS), Case-Based (CBR) RS, Knowledge- Based (KB) RS**

I. Introduction

THE diversity and complexity of different Educational Data Mining (EDM) techniques pose a huge challenge to educational management decision-makers. The velocity and volume of advances in EDM techniques coupled with a lack of common terminologies, make the selection of EDM techniques and appropriate variables more intractable. The supreme difficult task in EDM process is to choose the correct technique and the decision requires technical expertise, since there are numerous existing techniques for a scientist wishing to discover a model from the data. This diversity poses a serious challenge to a non-expert user who has no clear idea of what techniques are available to solve existing domain problem [1]. The classification of EDM techniques will simplify the understanding of the existing techniques. Yet again, a number of EDM systems do not have intelligent assistance for addressing EDM process, but instead provide a conceptual map.

Therefore, the proposed taxonomy of Recommendation Systems (RS) illustrates the identified gap in the literature reflecting the challenges, limitations and proposed methods to overcome them. The researchers propose the taxonomy of RS through systematic review of the literature.

This taxonomy summarizes Recommender Systems (RS) used in solving the problems of admission process, prediction of student performance and student retention. The proposed taxonomy outlines problems in each RS approach, and techniques used to solve the problem. However, most techniques do not entirely resolve the problems.

II. Related Works

A. *Recommendation System (RS)*

Recommender Systems (RS) is a software tool or technique that gives recommendations to help identify a set of elements that will be of interest and relevant to users. It is also defined as a system that chooses items suitable for a specific user [2];[3]. The central theme of RS is grounded upon similarity measures or data mining (DM) techniques [2]; [4];[5]

As an application of data mining techniques, EDM addresses educational matters in relation to a certain set of data from educational institutions (EIs). The results of previous research in EDM encourage us to create a recommender system for educational data mining techniques for institutions of higher learning as an expert system for automated decision making to solve problems of diversity and complexity of different EDM techniques[2]; [3]; [6]; [7].

Some of RS approaches include main categories such as:

- Content-based recommendation system
- Collaborative recommendation system
- Hybrid recommendation system
- Knowledge-based recommendation system

Many researchers identified the techniques of RSs like

Manuscript received July 10, 2019; revised July 16, 2019. This work was supported by the Tshwane University of Technology. The authors, M.A Masethe is with the Department of Computer Science at Tshwane University of Technology, eMalahleni Campus, 1035, South Africa (phone: +27 84-888-6624; e-mail: masethema@tut.ac.za).

S.S Ojo is a research professor with the Department of Computer Science at Tshwane University of Technology, Soshanguve Campus, Pretoria 0001, South Africa (phone: +27 382 9689; fax: +27 866-214-011; e-mail: ojoso@tut.ac.za)

S.A Odunaike is with the Department of Computer Science at Tshwane University of Technology, Soshanguve Campus, Pretoria 0001, South Africa (phone: +27 382 9151; e-mail: odunaikesa@tut.ac.za).

Collaborative Filtering (CF), Content-based, Knowledge-based and the Hybrid recommendation systems whereby Content-based and Collaborative Filtering have been applied mostly in the retail field [8]. These systems are more classified as navigators to direct non-expert users in simplifying their decision-makings within new or unknown environment. The purpose is to preserve user interest or summary that contains users' preferences [8].

B. Collaborative Filtering RS (CFRS)

The assumption is that other users' views are hoarded more in order to deliver a reasonable prediction of the active user's preference. Instinctively, it is assumed that users who agreed on importance of an item, will possibly agree on other items[9]; [10]. The CF traditional RS is widely used in e-commerce and browsing document. The notion is in discovery of users in a community that share obligations [11]. CF approach depends on availability of user ratings information, for example, Peter likes items A, B and C but dislikes items E and F. It recommends targets for user based on items that comparable users have liked previously, without depending on items information.

CF applications use classification, clustering, association and sequential techniques to learn new and thought-provoking models that assist to suggest RSs. The RS is based on assumption that when users shared the same interests previously there is probability to have similar interest in the future exists [12]; [13].

Further studies reveals that have deployed web mining as a technique in data mining using various methods, patterns with collaborative filtering and content filtering applying clustering, classification, and association. This is a sequential pattern to solve identified e-learning problems such as prediction of performance of e-learners and registration, tracking assignment, analyzing e-learners feedback, and monitoring e-learners progress. Their system was tested using black box and white box methods [14].

Wang et al [9] further proposes a Tensor Factorization (TF) framework that is able to capture user's opinions on different aspects, since CFRS methods relied only upon users' overall ratings of items disregarding variability of opinions users may give towards aspect of items. Approaches closely related to this TF are Matrix Factorization (MF) and Maximum Margin Matrix Factorization (MMMF) that are based on the known entries in the model, but hardly scalable. TF extends CFRS to the N-dimensional case, giving autonomy to integrate opinion information. This solution addresses the data sparseness problem, which occurs in CFRS system when the ratio is too small to provide enough information for active predictions.

C. Content-Based Recommendation System (CBRS)

Content-Based RS (CBRS) allows the user to rate items while the system evaluates common characteristics among the past data items and recommends items with extraordinary grade of similarity to the user according to their favors and likings. This depends on accessibility of item descriptions and a summary that allocates importance to these characteristics. This type of RS concentrates on user profiles generated at the beginning and consists of a survey of users characteristic and their rated items [16]; [17]; 18].

In the RS process the engine compares items already positively rated by the existing user with items not yet rated and looks for similarities, and finally items most similar to the rated ones are recommended to the user [18]. The approach relies on the item description to produce recommendations from items that are similar to those the target user has liked previously and do not rely on other user preferences [19].

In contrast, Content-Based approach depends on the item descriptions to produce recommendations to the user from items that are comparable to those the target user has liked previously and do not depend on preferences of other users [19]. Other studies reveal that CBRS is applied in academic social networks to propose important items to members of online societies, hence making a substantial contribution to the user satisfaction [20]. Furthermore, it assists academics in finding proper content by determining clusters of similar users and inferring user's interest in a resource, hence enriching a lesson, enhancing knowledge about a topic of interest, thereby improving performance of learners in academia [20]. The CBRS simply endures in the CFRS manner. Like CF, it requires user's data from the past, but its characteristics bring limitations of interest to the users preferences [21].

Even though CBRS does not depend on other user's data to avoid the issue of cold start and sparseness, it relies on the requirements of recommended item's structure. It struggles in finding user's new items of interest and challenged by complex attributes. CBRS is suffering from over-specialization, limiting users to discover new and different recommendations [22].It directs users to recommendations that are already known to them. Therefore, due to massive data rising in the educational domain, with different and complex attributes contained there-in, CBRS gets disqualified in assisting this study to bring the solution for better selection of the educational data mining techniques to a non-expert user [21].

D. Hybrid Recommender System (Hybrid RS)

Hybrid RS is a cross method, where two or more of the RSs comes together to formulate one suitable RS. For example, we noted that CFRS mostly suffers from limitations such as Cold Start, Grey Sheep and sparseness problems, whereas CBRS also suffers from limitations such as Limited Content, Overspecialization, and inflexible information problems. it is appropriate to hybridize the two approaches CFRS and CBRS in order to obtain better performance and overcome limitations observed [23]; [24].

They are a conglomeration of two or more approaches to eradicate weaknesses of singular system, evade boundaries and increase performance of the system [13].

E. Knowledge-Based Recommender System (KB-RS)

The Knowledge Based RS (KB-RS) method applies knowledge about the users and items to generate a recommendation according to the user's requirements [24]. KB-RS does not need rating dataset to perform a recommendation but performs separately of user ratings.

ISBN: 978-988-14048-7-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

However, its weakness is upon the need for knowledge design [25]. Most KB-RS techniques frequently used case-based reasoning (CBR), hence the study recommends CBR technique since it has the ability to study from its prior experience to solve problems and also resembles reasoning model of human beings which enhances the accuracy of the recommender solutions [26].

F. *Case Based Reasoning RS (CBR-RS)*

The CBR-RS has its own framework, which consists of six (6) steps recommended for problem solving cycle and is very effective in assisting users make better decisions timeously by solving problems and influencing the decision-making by learning from the past [27];[28];[29]. CBR-RS is a problem solving technique and theory of reasoning grounded on the approach humans think, reason and solve problems, further encompass four main steps [29]:

- Retrieve phase: a new problem is compared to cases in the case base library and similar cases are retrieved.
- Reuse phase: results of the retrieved cases are reused for the new problems and the achievement evaluated
- Revise phase: if suggested solution does not please the new problem, adjustment occurs
- Retain phase: the amended solution and its conforming problem are retained in the case base library for future reference.

III. RESEARCH METHODOLOGY

The research study adopts Systematic Review (SR) Methodological Analysis to critically review previous research and experiences as recommended by [30]. SR Methodological Analysis is chosen to synthesize evidence from published papers, and explore literature relevant to RS in academic journals, books and conference proceedings [31][32]. [33] embraces search engines such as: Association for Computing Machinery (ACM), EbscoHost Premier Package, Emerald Management Xtra, IEEE Xplore Digital Library, IOPscience and National Research Foundation (NRF) Databases to obtain access to recommendation system information.

The researchers review the literature to see previous work in relation to what worked as well as what did not work. This may further assist this research work to identify better solution to the highlighted problem of simplification of EDM technique selection for enhanced predication accuracy in terms of student academic performances for academic decision makers. Furthermore, to identify data sources used by other researchers and contribute to the research field whilst demonstrating the understanding and ability to critically evaluate the research. All sources will be from different peer reviewed and published work, such as conference papers, accredited journals and books.

IV. RESEARCH RESULTS

A. *Systematic Review Results*

The research use systematic review framework composed of five important steps [17] [18][19]: which are: 1. Define research question, 2. Ascertain important studies, 3. Choose articles, 4. Graph the data, and 5. Organize, Encapsulate and report the research outcomes.

1. Define research question

The research question has been defined as:

Can educational data mining techniques be easily selected using a taxonomy of recommendation system?

2. Ascertain important studies

We searched electronic databases such as ACM Digital Library (http://portal.acm.org), IEEE Explore (http://ieeexplore.ieee.org) and Google Scholar (http://scholar.google.co.in) using search terms as identified by research team, and information specialists as inputs.

3. Choose articles

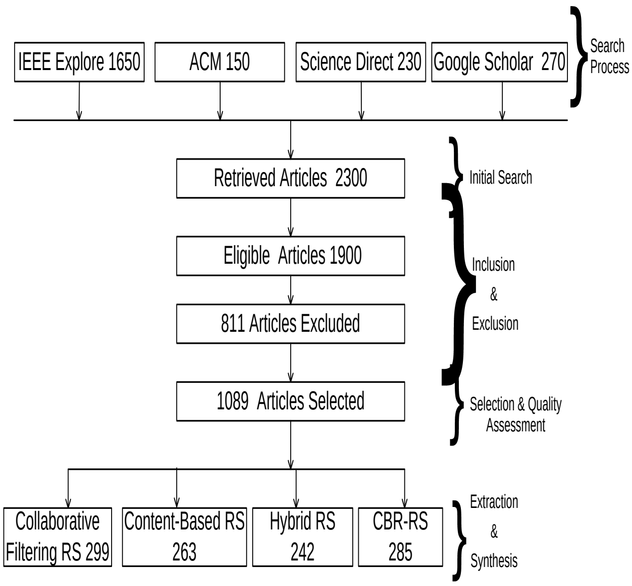

Fig. 1. Flow Chart of Search Result

The 2300 research articles were retrieved in 2013, as shown in Fig. 1 and selected for significance based on their titles and abstracts. Initial search brought the following number of articles through different search engines: Science Direct retrieved 230, IEEE Xplore retrieved 1650, ACM retrieved 150, and Google scholar retrieved 270. Research articles not published in English were excluded, including studies that avail abstracts only. Articles, which do not include EDM techniques or review in recommendation systems, were also excluded. The included articles were inspected to extract information about paradigms and outcomes from different perspectives, including experiences. Articles that met inclusion criteria were retained for systematic review.

ISBN: 978-988-14048-7-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

4. Graph the data, Organize and encapsulate and report the research results

Fig 2 shows frequency of RS and EDM techniques according to the number of research articles studied and their percentage. The figure also summarizes the popular techniques used in recommender systems. Bayesian network and decision trees seem to be popular techniques used by researchers in different papers followed by k-means, SVM, ANN and K-Nearest Neighbors (KNN) technique.

B. Proposed Taxonomy of EDM techniques

The proposed taxonomy of recommendation systems in Fig. 3 illustrates the identified gap in the literature reflecting

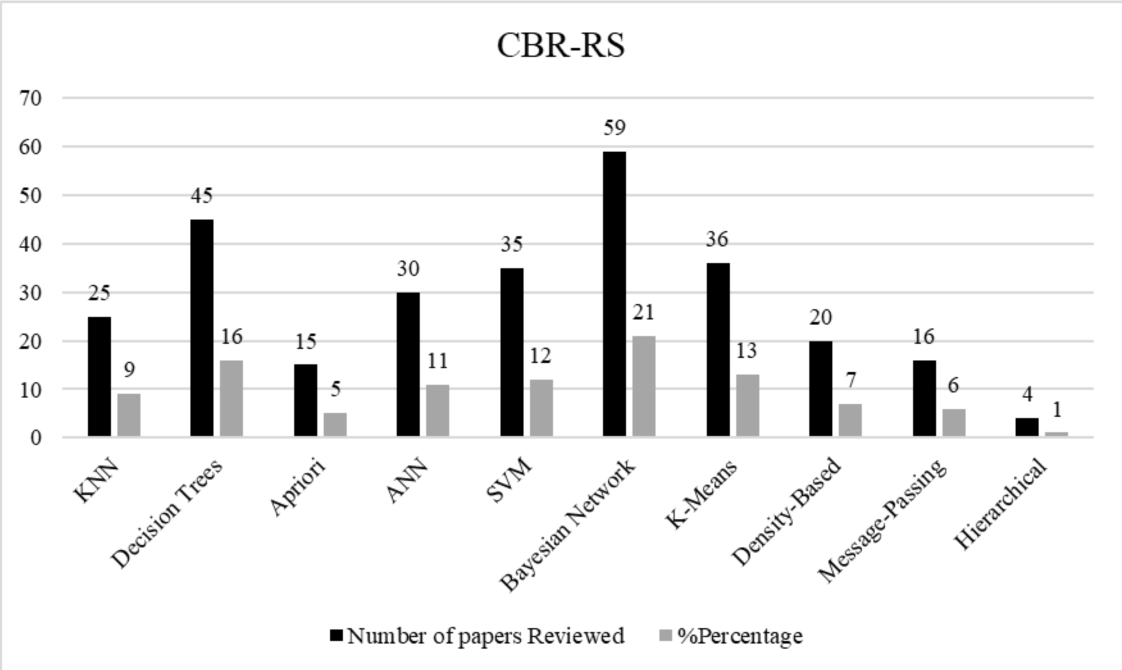

Fig. 2. CBR-RS

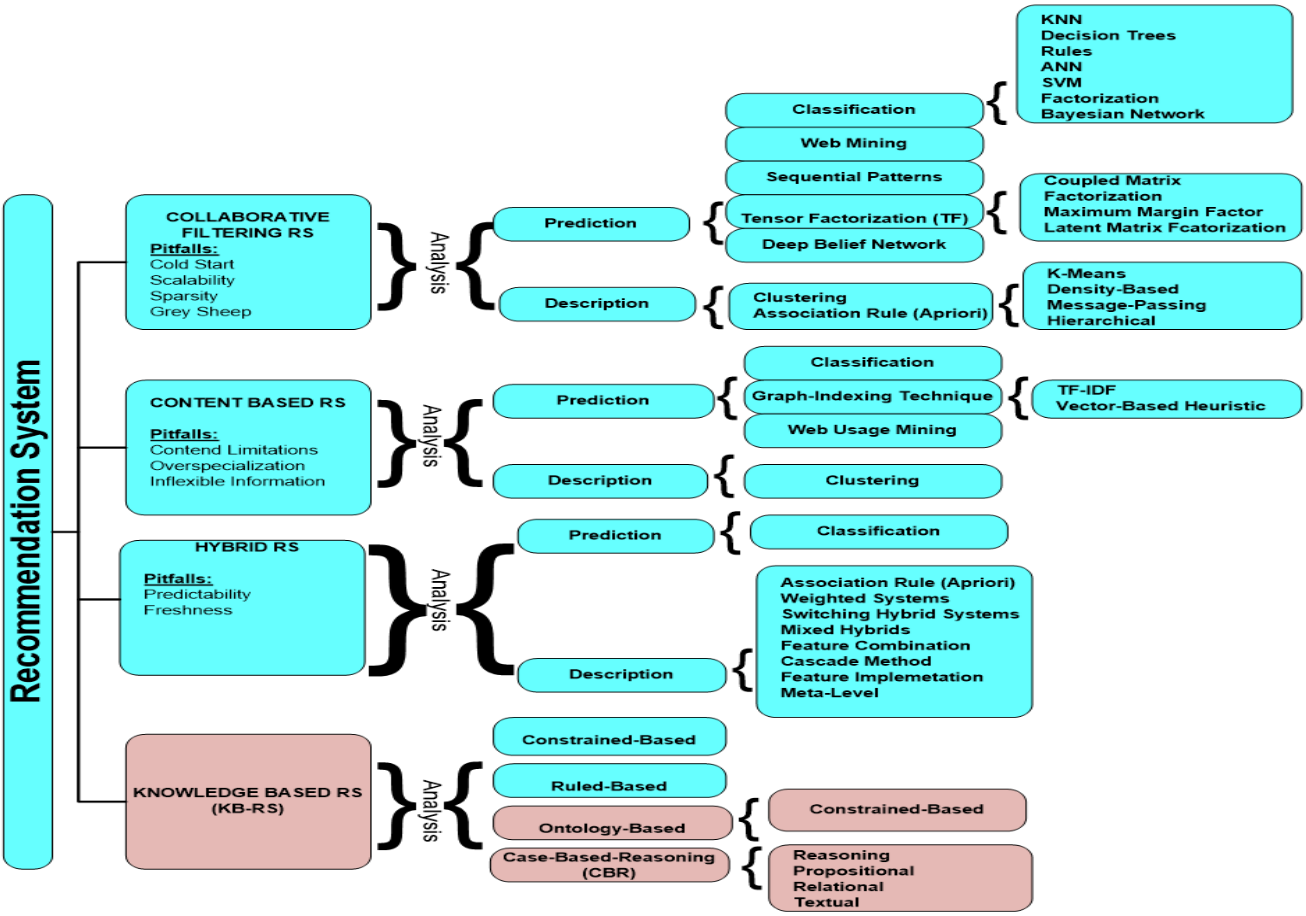

Fig. 3. Taxonomy of EDM Recommendation System (RS)

ISBN: 978-988-14048-7-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the challenges, limitations and proposed methods to overcome them. The researchers propose the taxonomy of RS in Fig.3 through systematic review of the literature to summarize recommender systems used in solving the problems of admission process, prediction of student performance and student retention. The proposed taxonomy outlines problems in each RS approach and techniques used to solve the problem. However, most techniques do not entirely resolve the problems.

Systematic review in this section allows the researcher to form a strong argument that presents comprehensive and logical state for the taxonomy. Systematic review is a technique of identifying, evaluating and interpreting all research articles relevant to a research question or topic area of interest [34]. There exists a rapid growth of review techniques, each using diverse approaches with the intention to collect, assess and present research proof which includes systematic review, narrative review, conceptual review, rapid review, realistic review, scoping review and meta-analysis, in this research we adopt systematic literature review.

The systematic literature review in this study produces existing evidence of EDM techniques supported in fig 3 and also used in CBR-RS in a fair, rigorous, and open way.

V. Conclusion

The research discussed and analyzed the literature based on recommender system techniques. Further, it presented the different approaches in recommender system such as content-based, collaborative filtering, hybrid, and knowledge-based recommender system. The study looked at current existing challenges of different recommender system including proposed technique to solve the problems. Knowledge-based recommender system seems to solve problems encountered by content-based recommender system and collaborative filtering recommender system. The chapter also discussed how knowledge-based RS notable employs case-based reasoning and ontology-based engineering to overcome challenges such as cold-start, scalability, sparseness, grey sheep, contend limitations, overspecialization, and inflexible information.

References

[1] D. Swayne, W. Yang, A. Voinov, A. Rizzoli, and T. Filatova, "Choosing the Right Data Mining Technique: Classification of Methods and Intelligent Recommendation," in *2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting*, 2010, no. Kdnuggets 2006, pp. 1–9.

[2] H. Bydžovská, "Course Enrolment Recommender System," Faculty of Informatics, 2012.

[3] G. N. Rayasad, "ASSOCIATION RULE MINING IN EDUCATIONAL RECOMMENDER SYSTEMS Gurudatta Nanjund Rayasad," *Int. J. Adv. Res. Sci. Eng.*, vol. 6, no. 8, pp. 1285–1299, 2017.

[4] Y. Koren and R. Bell, "Advances in Collaborative Filtering," *Recomm. Syst. Handb.*, pp. 145–186, 2011.

[5] A. D. Kumar and V. Radhika, "A Survey on Predicting Student Performance," vol. 5, no. 5, pp. 6147–6149, 2014.

[6] F. M. Almutairi, N. D. Sidiropoulos, and G. Karypis, "Context-Aware Recommendation-Based Learning Analytics Using Tensor and Coupled Matrix Factorization," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 5, pp. 729–741, 2017.

[7] H. L. Thanh-Nhan, H. H. Nguyen, and N. Thai-Nghe, "Methods for building course recommendation systems," *Proc. - 2016 8th Int. Conf. Knowl. Syst. Eng. KSE 2016*, pp. 163–168, 2016.

[8] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, "Recommender Systems : An Introduction , by Dietmar," *Human-Computer Interact.*, no. October 2014, pp. 1–3, 2012.

[9] Y. Wang, Y. Liu, and X. Yu, "Collaborative Filtering with Aspect-based Opinion Mining : A Tensor Factorization Approach," in *12th International Conference on Data Mining*, 2012, pp. 1152–1157.

[10] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, "Collaborative Filtering Recommender Systems," *Found. Trends Human-Computer Interact.*, vol. 4321, no. 1, pp. 291–324, 2007.

[11] D. Asanov, "Algorithms and Methods in Recommender Systems," *Other Conf.*, 2011.

[12] "Recommendation in Higher Education Using Data Mining Techniques."

[13] A. M. K. Alsalama, "A Hybrid Recommendation System Based on Association Rules," *Int. J. Comput. Control. Quantum Inf. Eng.*, vol. 9, no. 1, pp. 55–62, 2015.

[14] D. Herath and L. Jayaratne, "A Personalized Web Content Recommendation System for E-Learners in E-Learning Environment," *Natl. Inf. Technol. Conf.*, pp. 89–95, 2017.

[15] L. C. Quispe and J. E. O. Luna, "A content-based recommendation system using TrueSkill," in *Proceedings - 14th Mexican International Conference on Artificial Intelligence: Advances in Artificial Intelligence, MICAI 2015*, 2016, pp. 203–207.

[16] T. Peng, W. Wang, X. Y. Gong, Y. Tian, X. G. Yang, and J. Ma, "A graph indexing approach for content-based recommendation system," in *2010 International Conference on MultiMedia and Information Technology, MMIT 2010*, 2010, vol. 1, pp. 93–97.

[17] S. Nadschläger, H. Kosorus, A. Bögl, and J. Küng, "Content-based recommendations within a QA system using the hierarchical structure of a domain-specific taxonomy," in *Proceedings - International Workshop on Database and Expert Systems Applications, DEXA*, 2012, pp. 88–92.

[18] D. Asanov, "Algorithms and Methods in Recommender Systems," *Other Conf.*, 2011.

[19] B. Smyth, "Case-Based Recommendation," *Adapt. Web, LNCS*, vol. 4321, pp. 342–376, 2007.

[20] V. A. Rohani, Z. M. Kasirun, and K. Ratnavelu, "An enhanced content-based recommender system for academic social networks," in *Proceedings - 4th IEEE International Conference on Big Data and Cloud Computing, BDCloud 2014 with the 7th IEEE International Conference on Social Computing and Networking, SocialCom 2014 and the 4th International Conference on Sustainable Computing and C*, 2015, pp. 424–431.

[21] H. Jun, Y. Fan, L. Jing, T. Jun, and C. Shuo, "A Comparative Study of the Recommendation Algorithm Applicate in the Network of Educational Resources Platform," no. Icetis, pp. 677–681, 2013.

[22] S. Jain, A. Grover, P. S. Thakur, and S. K. Choudhary, "Trends, problems and solutions of recommender system," *Int. Conf. Comput. Commun. Autom. ICCCA 2015*, pp. 955–958, 2015.

[23] G. Meryem, K. Douzi, and S. Chantit, "Toward an E-orientation Plateform," 2016.

[24] T. Zuva, O. O. Olugbara, S. O. Ojo, and S. M. Ngwira, "Kernel Density Feature Points Estimator for Content-Based Image Retrieval," 2012.

[25] M. G. Wonoseto and Y. Rosmansyah, "Knowledge Based Recommender System and Web 2.0 to Enhance Learning Model in Junior High School," in *International Conference on Information Technology Systems and Innovation (ICITSI)*, 2017, pp. 168–171.

[26] W. Husain and L. T. Pheng, "The development of personalized wellness therapy recommender system using hybrid case-based reasoning," in *ICCTD 2010 - 2010 2nd International Conference on Computer Technology and Development, Proceedings*, 2010, no. Icctd, pp. 85–89.

[27] A. Aamodt, "Knowledge Acquisition and Learning by Experience - The Role of Case-Specific Knowledge," 1995.

[28] F. Lorenzi and F. Ricci, "Case-based recommender systems: a unifying view," *Intell. Tech. Web Pers.*, vol. 3169 LNAI, pp. 89–113, 2005.

[29] S. Soltani, P. Martin, and K. Elgazzar, "QuARAM recommender:

Case-based reasoning for IaaS service selection," in *Proceedings - 2014 International Conference on Cloud and Autonomic Computing, ICCAC 2014*, 2015, pp. 220–226.

[30] M. Tsiknakis and M. Spanakis, "Adoption of innovative eHealth services in prehospital emergency management : a case study," in *10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB)*, 2010, pp. 1–5.

[31] S. C. Wangberg and C. Psychol, "Personalized technology for supporting health behaviors," in *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*, 2013, pp. 339–344.

[32] M. Fanti and W. Ukovich, "Discrete event systems models and methods for different problems in healthcare management," in *IEEE Emerging Technology and Factory Automation (ETFA)*, 2014, pp. 1–8.

[33] L. Fernandez-luque, R. Karlsen, T. Krogstad, T. M. Burkow, and L. K. Vognild, "Personalized Health Applications in the Web 2.0: the emergence of a new approach," in *32nd Annual International Conference of the IEEE EMBS*, 2010, pp. 1053–1056.

[34] J. M. Verner, O. P. Brereton, B. A. Kitchenham, M. Turner, and M. Niazi, "Systematic literature reviews in global software development: a tertiary study," in *16th International Conference on Evaluation & Assessment in Software Engineering (EASE 2012)*, 2012, pp. 2–11.

ISBN: 978-988-14048-7-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)