International Journal of Latest Trends in Engineering and Technology
Vol.(8)Issue(2), pp. 212-219
DOI: http://dx.doi.org/10.21172/1.82.029
e-ISSN:2278-621X

CLASSIFYING LEARNERS' COGNITIVE ENGAGEMENT FROM ONLINE DISCUSSION USING TEXT MINING

Hind HAYATI[1], Mohammed KHALIDI IDRISSI[2] and Samir BENNANI[3]

Abstract- This paper proposes a learners' classification system according to their levels of cognitive engagement based on their activity in online discussion forums. To achieve this classification, Text mining is applied to two types of data: the posted threads and learners' participation in forums. In order to increase the efficiency and the response time of the proposed system, a semantic classification of the threads has been made in advance to filter those that are important and relevant to the context chosen by the tutor. This step is performed by combining two methods, namely the Ontology OWL and the LSA method.

Keywords – Online discussion forum, E-learning, Cognitive engagement, Text mining, Ontologie OWL, SVM, Classification, LSA

I. INTRODUCTION

Over the last decades, technological advances have accelerated the emergence of online education and have made it more successful. Several forms of online education - such as Mobile Learning, Moocs and SeriousGames - have emerged to meet the needs of a new generation; the objective being to change the traditional setting for distant learning and to make necessary improvements to the teaching methods, that are now challenged by space-time constraints, in order to ensure a better learning. Consequently, e-learning is now a complex learning environment in which the learner feels autonomous and responsible for his / her educational path [1]. That said, the learner has to show enough commitment and motivation to succeed in an online learning environment.

The analysis of the scientific literature revealed that abandonment constitutes one of the challenges of on-line training. This is often a recurring fact that prompts researchers to analyze the factors influencing learning: demotivation [2], lack of time [3], lack of course prerequisite knowledge [4], the absence of direct contact within e-learning platforms [5], etc. In this article, the factor addressed is learners' engagement in e-learning: a factor whose importance has attracted the researchers' attention given its impact on the success and academic performance of learners [6]. Several studies [7] [8] [9] have shown the correlation between engagement and academic progress, in that engaged learners are more likely to succeed than give up [10].

The learner's engagement within the course or in a specific activity may be sentimental, behavioral or even cognitive. In this article, we address the cognitive dimension of engagement as it can express the learners' reflections and mental efforts. Furthermore, we propose a learners' classification system according to their levels of cognitive engagement based on their participation in online discussion forums. On the one hand, discussion forums can provide information about the learner from his learner-to-learner and learner-to-tutor interactions. On the other hand, it is an activity that allows learners to express themselves, present their ideas and have a deep reflection on the discussed course.

The objective of our contribution is to analyze the threads in the discussion forums in order to help the tutor assess the level of each learner's cognitive engagement. The latter may eventually take the necessary steps to ensure that the

[1] *RIME Team- Networking, Modeling and E-learning LRIE Laboratory- Research in Computer Science and Education Laboratory EMI- Mohammadia School of Engineering, Mohammed V University in Rabat, Rabat, Morocco*
[2] *RIME Team- Networking, Modeling and E-learning LRIE Laboratory- Research in Computer Science and Education Laboratory EMI- Mohammadia School of Engineering, Mohammed V University in Rabat, Rabat, Morocco*
[3] *RIME Team- Networking, Modeling and E-learning LRIE Laboratory- Research in Computer Science and Education Laboratory EMI- Mohammadia School of Engineering, Mohammed V University in Rabat, Rabat, Morocco*

learning process is carried out more efficiently, by giving learners good support and helping them to become more committed.

This article is organized as follows: in the first section, work related to the engagement of learners will be presented, then to address the issue of learners' engagement and its different types in on-line learning environment, as well as the benefit of using discussion forums as data source. The third section will illustrate the methods used to classify learners according to their engagement level. A detailed description of the proposed system will follow. Finally, the conclusion of this work and the perspectives of research will be presented.

II. RELATED WORKS

Numerous research in e-learning has shown that the participation of learners in online discussions is a factor of success, as it offers an additional learning channel [11] and represents an engaging activity in which learners can share and transfer information about the course, reflect deeply on its content and express their ideas freely. "Shukor & all" attempted to evaluate the learning quality of learners by their cognitive participation in discussion forum using k-mean clustering algorithm [1]. According to her, qualitative factors such as the number of views and posts are not enough to assess the cognitive level because even with an important number of discussion views and posts, some learners still have a low level of cognitive engagement.

"Wang & all" on his part finds that the cognitive behavior of learners in forums has not been treated as a factor impacting their learning process, especially since there is a significant bound between "the quality and quantity of participation in the discussion forum" and "the learning results" [9], but manually coding learners' behavior during online discussions can result in some errors and decrease the reliability of the results. On the other side Kovanovic & all uses LIWC and Coh-Matrix to automate the coding of messages in order to deduce the cognitive presence [12], this method shows its efficiency in classification but only for a small number of thread-based context Features which can cause generalization issues.

SVM (Support Vector Machines) was used to analyze the learners' behavior in discussion forums. Nan Li & all wanted to analyze the emotional dimension by developing an algorithm that automatically identifies the emotional polarity of a text and then combines it with two Text mining techniques, namely k-means clustering and SVM [13]. This approach is based on the concept of a hotspot, which is used as a cluster center and is determined from the hotspot history. But there are quite a few hotspots that are not related to the history and could therefore impair the results if the method is generalized.

III. BACKGROUND

In this section, the problematic of learner engagement will be approached according to the following three dimensions: emotional, behavioral and cognitive. As well as the benefit of using online discussion forums as data sources.

A. Engagement

Engagement is a vague concept that is generally defined by the determination to achieve a given task. in education, this concept is represented by the learner's interest in the educational program, the course or even associated activities such as forums, tests, surveys, etc. Scientists propose other definitions: according to Bouvier, engagement is defined as the will to have emotions and thoughts directed towards and determined by the mediated activity [14]. While Little-Wiles perceives that learners' engagement is not solely limited to the learner's involvement or learning, but also to the learning environment development and design [15].

There are three types of engagement: emotional, behavioral, and cognitive.

- **Emotional engagement** is linked to the sentimental reactions of the learner, which may be feelings of joy, boredom, unhappiness or interest in the course followed, in fact, the learner is emotionally engaged if the course pleases him and peaks his interest.

- **Behavioral engagement** is defined as the learner's commitment to following the rules and standards set by the tutor during the educational program in addition to his / her interest in the activities, efforts, persistence and contribution to the discussions.

- **Cognitive engagement** is accomplished when the learner makes mental efforts to engage with the learning materials. This type of engagement remains important in e-learning environment because of the learners' autonomy and they feel responsible for their learning. That is why their cognitive engagement level can influence their learning

and motivation. Zhu defines cognitive engagement by [16] "... attention to related readings and effort in analyzing and synthesizing readings demonstrated in discussion messages. Cognitive commitment, as defined, involves seeking, interpreting, analyzing, and summarizing information; Critiquing and reasoning through various opinions and arguments; and making decisions. "

In our work, we will focus on cognitive engagement as it is tightly related to the learner's mental efforts and can better express his level of understanding and learning. One can find, for example, a committed but not cognitively engaged learner: this is the case for learners who work hard but are still unable to achieve good results [17]. Given the lake of face to face contact between the tutor and the learner, and the nature of the cognitive process, this type of engagement remains difficult to observe and determine.

B. Discussion forums: asynchronous communication tool

Thanks to the development of technologies, communication between members of online training has become increasingly easy by removing the temporal and geographical barriers that separate them and ensuring better collaboration between them. There are two types of communication tools: synchronous communication tools that enable real-time communication and require simultaneous connection between participants such as chat, videoconferencing, audio conferencing, etc. And asynchronous communication tools that offer flexibility in their use such as e-mails, mailing lists, FAQs, discussion forums, etc.

From an analytical perspective, forums are a place for analyzing data both as information and as interaction – e.g. as behavior. In our work, we are interested in asynchronous communication tools and more specifically the discussion forums since they present more freedom for its users and provides them a better structuring of their knowledge as they take more reflection time.

A discussion forum is a tool that allows the communication between different participants at any time while keeping the traces of the various mediated exchanges. Given the degree of autonomy of a learner during an online training, Larkin-Hein [18] sees that the discussion forums represent a promising way to both achieve affective attachment and acquire an active role within the program. Althaus [19] adds that learners learn best through the use of online discussions because these latter places them in an intellectual environment that encourages active participation, reflection, and provides equality among all learners.

IV. METHODS

The proposed solution relies on two main methods: Ontology and Text mining. This section illustrates how these methods contribute to the development of the learner classification system according to their cognitive engagement levels.

A.Ontology

An ontology is a cartography that formally defines a common set of terms used to describe and represent a domain [20]. According to Saadia LGARCH, an ontology is a formal and explicit specification of a shared conceptualization [21]. in knowledge engineering, it is a formal representation of a domain knowledge that is modeling, reuse and sharing oriented. The formal aspect ensures the understanding of the ontology by machines, while the modeling via the ontologies makes it possible to represent a domain comprehensible by the human actors and the software agents. Reuse allows the ontology to be exploited by several applications in several contexts. Finally, the sharing aspect indicates that ontology supports consensus knowledge, and is not restricted to a certain individual, yet it is instead accepted and shared by a group or community.

The main role of these ontologies is to represent knowledge and apply reasoning to it. As part of our RIME team (Computer Networks, Modeling and E-learning) work and in order to ensure the continuity of the accomplished works, we will adopt Saadia’s OWL Ontology [21] which allows semantic classification of messages in two categories: Interesting and not interesting in relation to a given context specified by tutor.

B.Text mining

The diversity of and easy access to information sources today can harm many users as they feel disoriented and lost in a huge data stream, which does not always present interesting material. Data mining is therefore introduced as an Analysis process for knowledge discovery in a database. However, datamining methods only process data with a well-structured format, which limits its use for some data sources like ours.

Online discussion forums are considered to be unstructured data sources. That is why Text mining has been chosen as a method of knowledge extraction since it focuses on the exploration of structured data and the extraction of useful information from unstructured text data collections. [22] Text mining, also known as Intelligent Text Analysis, Text Data Mining or even Knowledge Discovery in Text (KDT), offers several classification algorithms such as Bayesian classifier, Decision Tree, K-nearest neighbor (KNN), Support Vector Machines (SVMs) and Neural Networks.

Among these algorithms, SVM has recently attracted more attention, and several studies have shown that SVM as a classification method is remarkable compared to others, with its effective and surpassed classification [23] [24] [25]. Support vector machines (SVMs) is based on the principle of Structural Risk Minimization (SRM) and promotes linear separation, it is recognized as one of the most accurate discriminant classification methods [26]. In fact other researchers have implemented and measured the performance of the main supervised and unsupervised multilingual categorization approaches and demonstrated that SVM was the leader of supervised methods [27].

We have also chosen the SVM as a classification method for the proposed system because of its importance in relation to the textual data. The latter are large in size and represent many characteristics of significant importance, but they tend to be correlated with each other and are generally organized into linearly separated categories.

V. PRESENTATION OF THE CLASSIFICATION SYSTEM FOR COGNITIVE ENGAGEMENT

As part of our work we propose a classification system for learners' cognitive engagement in the pedagogical process according to their participation in online discussion forums. In this section the architecture of the system will be presented as well as its operation.

A. Proposed system architecture.

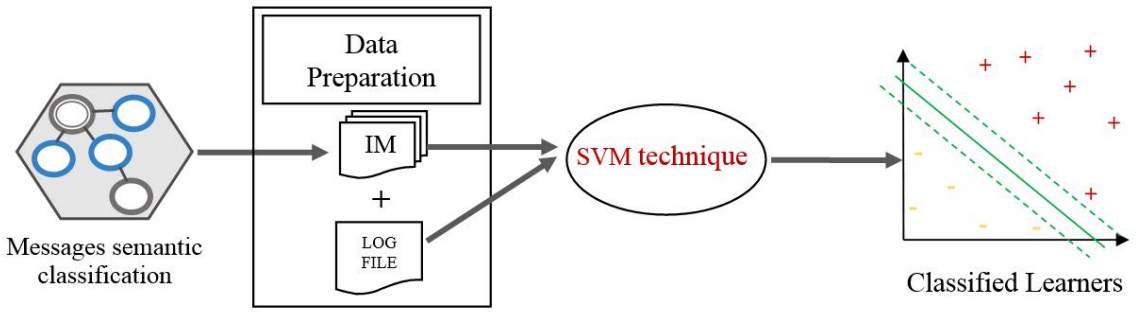
To ensure our objective, which is none other than classifying learners according to their level of engagement, the proposed system has two main stages.

Since on-line discussions generate a very large mass of messages, we will first propose a semantic classification of messages to eliminate noise. We will then switch to the system kernel consisting of a classification method using Text mining technology represented by the Support Vector Machine (SVM) algorithm.

Figure 1 represent the architecture of the proposed system.

Figure 1. Proposed system architecture.

B. System Operation



1) Semantic classification of messages

Asynchronous communication tools, and more specifically online discussion forums, allow the information exchange in a very flexible way. However, they generate a large mass of messages that vary in context and objective, making

this undesirable mixing may cause a blockage and a slowness of time. To solve this problem, we propose to semantically classify the forum messages according to the desired context in order to eliminate the generated noises. This will also make it possible to clearly identify the cognitive commitment of the learners in relation to a given domain. The majority of the classification methods propose a keyword search which allows to have search results dependent of and proportional to the relevance of the used keywords [28] for this reason we opted for a semantic classification of messages. To ensure the continuity of the work within our research team we adopted the semantic classification system of Saadia LGARCH [7].

Figure 2 shows the architecture of the semantic classification system based on the OWL ontology and the LSA (Latent Semantic Analysis).

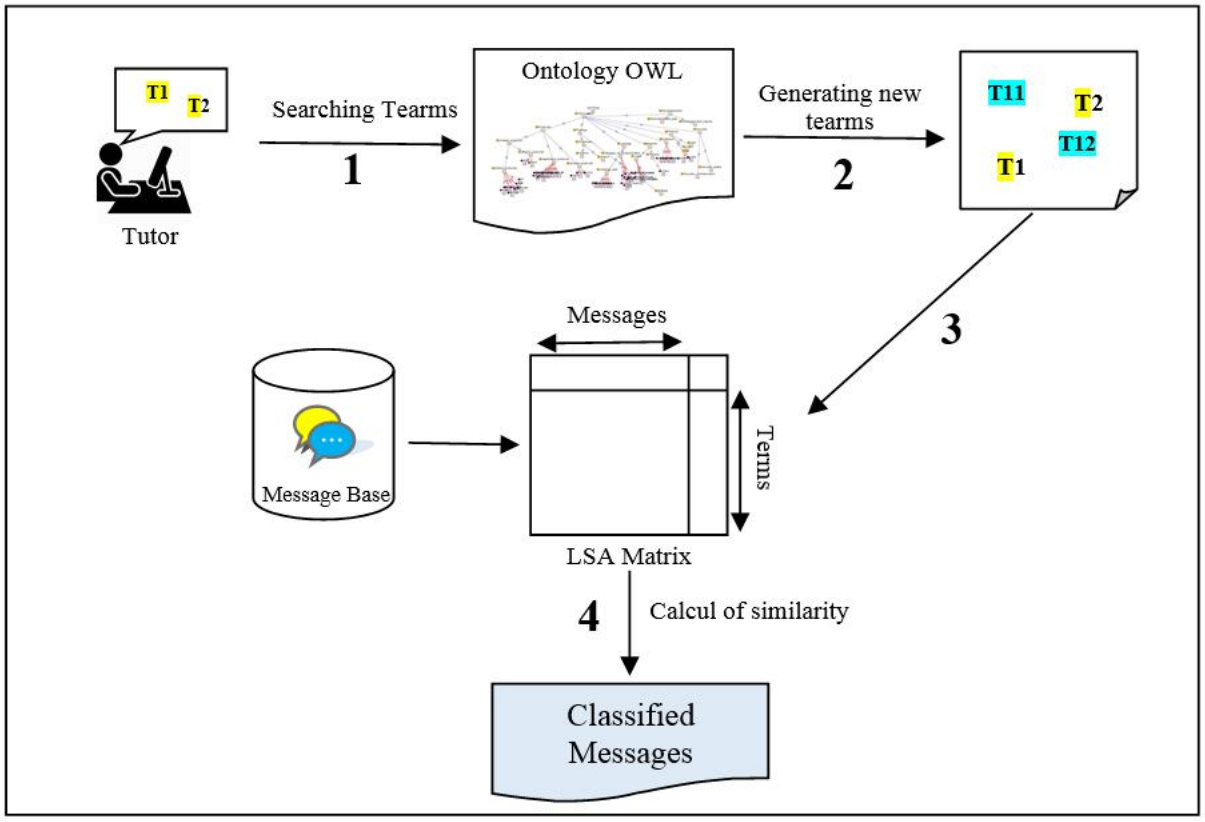


Figure 2. The architecture of the semantic classification system.

The system consists of four steps:

1. **Searching terms**: the user enters the terms that interest him, these terms will be the object of the inputs of the formal ontology OWL to build the knowledge base.

2. **Generating new terms**: after the deployment of the OWL ontology, new terms that are in semantic relation with the first entered terms are generated.

3. **LSA Matrix:** This step focuses on the construction of the LSA matrix with the new terms generated as rows of the matrix and the forum messages as columns.

4. **Calculation of similarity**: by applying the LSA method on the LSA matrix, this step makes it possible to find the similarity between the messages of the discussion forum so that they can be classified at the end.

2) Data preparation

We can identify two types regarding the data used within the system: qualitative and quantitative. The qualitative data are represented by the indicators generated from the log files of the platform but are always linked to the participation of the learners in the discussion forums; for example: the number of posts, number of views, number of answers, etc.

While the quantitative data are deduced from the written messages, in online discussions by analyzing the cognitive dimension.

3) SVM application

SVM is a supervised classification method that allows the definition of a linear separation hyperplane in order to maximize the margin. The margin being the distance between the separation boundary and the nearest sample. For our study the objective is to determine a hyperplane equation which can classify our samples into two classes. Thus, we must choose the hyperplane which presents the maximum margin with respect to the two classes. let's take an example to understand the operation of the SVM algorithm. Let the following linearly separable binary set be:

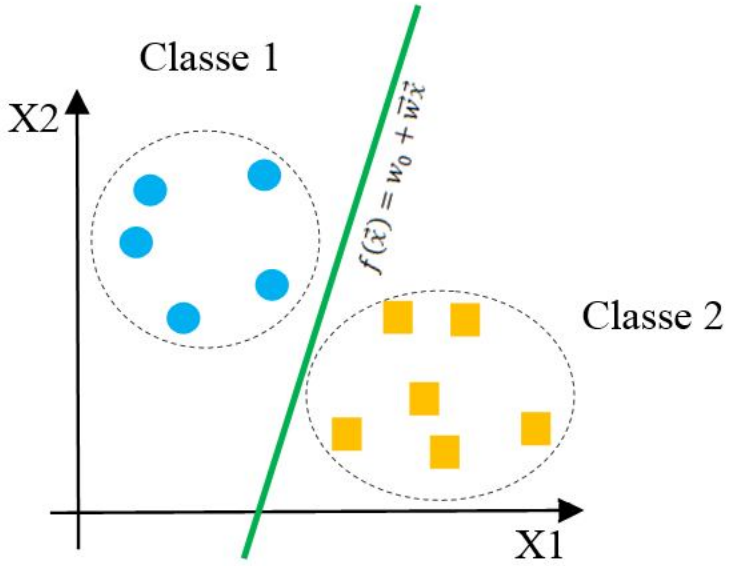


Figure 3. Lineary separable binary set.

Equation (1) represents the form of the equation of the hyperplane:

$$f(\vec{x}) = w_0 + \vec{w}\vec{x} \quad (1)$$

$$f(\vec{x}) \geq 1 \,, \forall \vec{x} \in classe\ 1$$

$$f(\vec{x}) \leq 1 \,, \forall \vec{x} \in classe\ 2$$

To maximize the margin, we must minimize the weight vector $\vec{w}$ which represents a nonlinear optimization solved by the Karush-Kuhn-Tucker (KKT) conditions using the Lagrange multipliers (2).

$$\sum_{i=0}^{N} \lambda_i y_i = 0$$

$$\vec{w} = \sum_{i=0}^{N} \lambda_i y_i \vec{x_i} \quad (2)$$

IV.CONCLUSION

The main objective of this paper is to classify learners according to their levels of cognitive engagement from their participation in online discussion forums, in order to help the tutor understand their cognitive behaviors and take the necessary steps to ensure better learning. By combining the OWL ontology, the LSA and the text mining, the level of

learners' cognitive engagement is analyzed. And since the undesirable mixing of messages from different contexts and objectives generates a blockage and a slowness of time, the proposed system begins with a semantic classification of forum messages in order to eliminate the noises according to the context chosen by the tutor. The formal ontology OWL and the LSA (Latent Semantic Analysis) method are combined to achieve this goal. All the messages obtained represent quantitative data and are combined with the traces of the learners within the e-learning platform in relation to their participation in the online discussion. These qualitative data are available to LMS log files, such as number of posts, number of views, number of responses, etc. In the second step, the SVM is applied, as a classification algorithm, to the two types of data to give two classes of learners according to their levels of cognitive engagement.

As perspective, we plan on combining the SVM with another algorithm to increase its performance since the latter requires a significant calculation time given the number of characteristics taken into account. We will also attempt to develop an oriented architecture SOA service to facilitate tutor intervention and ensure interoperability and reuse of the proposed system.

REFERENCES

[1] Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning.Procedia-Social and Behavioral Sciences, 116, 4844-4853.

[2] Clow D. (2013) «MOOC and the funnel of participation», 3rd International Conference on Learning Analytics & Knowledge (LAK'13). Leuven, Belgique, Avril 2013.

[3] Conole, G. G. (2015). MOOCs as disruptive technologies: strategies for enhancing the learner experience and quality of MOOCs. Revista de Educación a Distancia, (39).

[4] Onah, D. F., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: behavioural patterns. EDULEARN14 Proceedings, 5825-5834.

[5] Tahiri, J. S., Bennani, S., & Idrissi, M. K. (2016, September). An assessment system adapted to differentiated learning within Massive Open Online Courses using psychometric testing. In Information Technology Based Higher Education and Training (ITHET), 2016 15th International Conference on (pp. 1-7). IEEE.

[6] Greene, B. A., Miller, R. B., Crowson, H. M., Duke, B. L., & Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. Contemporary educational psychology, 29(4), 462-482.

[7] Bulger, M. E., Mayer, R. E., Almeroth, K. C., & Blau, S. D. (2008). Measuring learner engagement in computer-equipped college classrooms, Journal of Educational Multimedia and Hypermedia, 17(2), 129-143.

[8] Hayati, H., Tahiri, J. S., Idrissi, M. K., & Bennani, S. (2016, October). Classification system of learners engagement within Massive Open Online Courses. In Information Science and Technology (CiSt), 2016 4th IEEE International Colloquium on (pp. 527-530). IEEE.

[9] Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating How Student's Cognitive Behavior in MOOC Discussion Forums Affect Learning Gains. International Educational Data Mining Society.

[10] Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of educational research, 74(1), 59-109.

[11] Wu, D., & Hiltz, S. R. (2004). Predicting learning from asynchronous online discussions. Journal of Asynchronous Learning Networks, 8(2), 139-152.

[12] Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016, April). Towards automated content analysis of discussion transcripts: a cognitive presence case. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 15-24). ACM.

[13] Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decision support systems, 48(2), 354-368.

[14] Bouvier, P., Sehaba, K., Lavoué, E., & George, S. (2013, July). Qualitative approach to identify and qualify players' engagement based on their traces of interaction. In IC-24th Francophone Knowledge Engineering Days (original language : french)

[15] Little-Wiles, J. M., Hundley, S. P., & Koehler, A. (2010, October). Work in progress—Maximizing student engagement in a learning management system. In Frontiers in Education Conference (FIE), 2010 IEEE (pp. T2C-1). IEEE.

[16] Zhu, E. (2006). Interaction and cognitive engagement: An analysis of four asynchronous online discussions. Instructional Science, 34(6), 451-480.

[17] H. Davis, J. Summers, and L. Miller, An Interpersonal Approach to Classroom Management: Strategies for Improving Student Engagement, Thousand Oaks, CA: Sage Publications, 2012, p. 23.

[18] Larkin-Hein, T. (2001). On-line discussions: a key to enhancing student motivation and understanding?. In Frontiers in Education Conference, 2001. 31st Annual (Vol. 2, pp. F2G-6). IEEE.

[19] Althaus, S. L. (1997). Computer-Mediated Communication in the University Classroom: An experiment with on-line discussions, Communication Education 46: 158-174.

[20] Hnida, M. E. R. I. E. M., Idrissi, M. K., & Bennani, S. A. M. I. R. (2014). A formalism of the competency-based approach in adaptive learning systems. WSEAS Transactions on Information Science and Applications, 11, 83-93.

[21] Lgarch, S., Idrissi, M. K., & Bennani, S. (2012). A Reusable and Interoperable Semantic Classification Tool Which Integrates Owl Ontology. International Journal of Computer Science Issues(IJCSI), 9(6).

[22] Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence, 1(1), 60-76.

[23] Heide Brücher, Gerhard Knolmayer, Marc-André Mittermayer; "Document Classification Methods for Organizing Explicit Knowledge", Research Group Information Engineering, Institute of Information Systems, University of Bern, Engehaldenstrasse 8, CH - 3012 Bern, Switzerland. 2002.

[24] Sahay, S. (2011). Support vector machines and document classification. URL: http://www-static. cc. gatech. edu/~ ssahay/sauravsahay7001-2. pdf.

[25] Soumen Chakrabarti, Shourya Roy, Mahesh V. Soundalgekar;, " Fast and Accurate Text Classification via Multiple Linear Discriminant Projection" , The International Journal on Very Large Data Bases (VLDB), pp. 170-185. 2003.

[26] Khan, A., Baharudin, B., Lee, L. H., & Khan, K. (2010). A review of machine learning algorithms for text-documents classification. Journal of advances in information technology, 1(1), 4-20.

[27] Chung-Hong Lee a,, Hsin-Chang Yang , "Construction of supervised and unsupervised learning systems for multilingual text categorization" , Expert Systems with Applications, pp. 2400–2410, 2009.

[28] Lgarch, S. A. A. D. I. A., Idrissi, M. K., & Bennani, S. A. M. I. R. (2010). A selection algorithm of terms from OWL ontology for semantically classify messages. International Journal of Engineering Science and Technology, 2(11), 6788-6800.