Integrating web usage mining for an automatic learner profile detection: A learning styles-based approach

Ouafae EL AISSAOUI
LSI, Polydisciplinary Faculty -Taza
University of Sidi Mohammed Ben Abdellah
Taza, Morocco
ouafae.elaissaouil@usmba.ac.ma

Yasser EL MADANI EL ALAMI
LIIAN laboratory, Computer science department
University of Sidi Mohammed Ben Abdellah
Fez, Morocco
yasser.elmadanielalami@usmba.ac.ma

Lahcen OUGHDIR
LSI, Polydisciplinary Faculty -Taza
University of Sidi Mohammed Ben Abdellah
Taza, Morocco
lahcen.oughdir@usmba.ac.ma

Youssouf EL ALLIOUI
LS3M, Polydisciplinary Faculty of Khouribga
Hassan First University
Khouribga, Morocco
youssouf.elallioui@uhp.ac.ma

***Abstract*— with the technological revolution of Internet and the information overload, adaptive E-learning has become the promising solution for educational institutions since it enhances students' learning process according to many factors such as their learning styles. Learning styles are a criteria of great import in E-learning environment because they can help the system to effectively personalize students' learning process. Generally, the traditional way of detecting students' learning style is based on asking students to fill out a questionnaire. However, using this static technique presents many problems. Some of these problems include the lack of self-awareness of students of their learning preferences. In addition, almost all students are bored when they are asked to fill out a questionnaire. Thus, in this work, we present an automatic approach for detecting students' learning style based on web usage mining. It consists in classifying students' log files according to a specific learning style model (Felder and Silverman model) using clustering algorithms (K-means algorithm). In order to test the efficiency of our work, we use a real-world dataset gathered from an E-learning system. Experimental results show that our approach provide promising results.**

Keywords—e-learning; learning style; learner profile; clustering algorithms; web usage mining.

I. INTRODUCTION

The advent of the Internet and Information technologies has opened a new era of learning principles, tools and techniques. In fact, E-learning is considered as a promising solution to educational institutions in order to enhance their traditional learning systems or to use an alternative way for virtual learning environment [1]. The main challenge of E-learning is to develop intelligent systems, which provide appropriate materials according to the learner's profile, what is called adaptive E-learning systems. This approach attempts to improve the effectiveness of e-learning platforms by providing customized environment. This can be done by building the user profile and using it as a primary entry to generate personalized learning process. Generally, user profiles consist of useful characteristics of each learner such as preferences, knowledge, skills, learning style and so on [2]. One of the main components of a learner profile is the learning styles. It consists of various factors which describe the way in which the learner interacts with the learning environment [3]. Learning styles are particularly important in E-learning environment since they may help the system to personalize the learning process according to characteristics of learners. In [4], authors argue that incorporating learning styles in E-learning platforms improve positively the learning process. Learning styles have been described in models in order to classify students according to various scales and dimensions. The challenge of improving adaptive learning process largely depends on correctly identifying the learning style of each learner. The classical method for detecting learning styles consists in using a static way. It is based on asking students for a test or to fill out a questionnaire. Despite the reliability of these traditional techniques, they suffer from many limitations that weaken the precision of identifying learning styles. Certain limitations does not address the problem set students' boredom to fill out questionnaires, since some of them require an extra effort and concentration. Secondly, questionnaires technique assumes that students are conscious of their learning styles and preferences, which may not always be the case.

To address this kind of problem, many works have been done in this field [5] [6] [4] [7]. However, identifying the appropriate learning style of a particular student is still a confusing task. In this work, we propose an automatic approach for detecting students' learning styles based on web usage mining techniques. The basic idea of our proposal relies on classifying log files of learners using a clustering algorithm according to a specific learning style model [8]. In our case, we use Felder and Silverman Model [9] since it is the most widely used in educational systems thanks to its ability to quantify students' learning style [5]. In addition, we use k-means algorithm as clustering method since it is easy to implement. In addition, in our case the number of clusters is known.

The rest of this paper is organized as follows: in section 2 we give a literature review of several related works. Section 3, presents our proposed approach. In section 4 we analyze and discuss our experimental results. Finally, we conclude with some perspectives for the future enhancements.

978-1-5386-4396-9/18/$31.00 ©2018 IEEE

II. PRELIMINARIES

In this section, we will introduce some concepts that adaptive learning approaches are based on.

A. Learner Model

The learner model is the most important component in an adaptive e-learning system [10], because of its ability to represent the characteristics of the learner according to which the educational system provides recommendations [11]. The characteristics that a learner model can contains are; the personal information, the knowledge level, the preferences and the learning styles [11].

B. Learning Style & Learning Style Model

The learning style refers to the preferred way in which the learner grasp and treat information , and it is considered as one of the main components of the learner model since it can describe the learner's behavior when he interacts with the e-learning environment [12] [13].

According to the behavior of the learners during the learning process we can classify them into many learning style categories; these categories can be used to create a learning style model. However, there are many learning style model such as , Kolb, 4MAT, VAK... and Felder-Silverman which is considered as the most popular one thanks to its ability to quantify students' learning style [5]. It proposes to categorize each learner in several categories according to his capacities, abilities and preferences.

C. Felder-Silverman Model

According to Felder-Silverman learning style model there are four dimensions where each dimension contains two opposites learning styles, the four dimensions [9] are:

Active/reflective: where the active category refers to the learners who prefer to work in group and to do physical experiment, while in the reflective category the learners prefer to work individually and to reflect about things before doing something.

Visual/verbal: Where the visual category refers to the learners who prefer to learn by seeing information as: schema, picture, diagram. While in the verbal category the learners prefer to read and hear information.

Sensing/intuitive: Where the sensing category refers to learners who prefer practical and procedural information, while in the intuitive category the learners prefer theatrical and conceptual information.

Sequential/Verbal: Where the sequential category refers to the learners who prefer to learn in small steps from the details to the big picture, while in the verbal category the learners prefer to see the big picture and then the details

D. Learning objects (LO)

We can define a learning objects as a small units of learning which construct a learning content. The learning object can take various digital resources forms as: figure, video, etc. and it is characterized by its ability to be reusable, interoperable, durable, and accessible.

The reuse of learning objects helps the educational institutes, on one side, to decrease the cost of creating them and on the other side to find useful contents by looking for them in dedicated repositories which contain collected learning objects.

To be easily found, a learning object has to be tagged and described with metadata .the most pertinent standards that are used to describe a learning object are: MPEG-7, Dublin Core, and LOM [14][15].

E. K-Means Clustering Algorithm

K-means algorithm is an unsupervised techniques with the goal of clustering similar objects, which classifies a set of data points into a predefined number of classes. The procedure follows a simple and easy way to classify a given data points. It is based on minimizing an objective function, which is defined as the sum of the squared distances from all points in a cluster from the centroid cluster. Formally, the K-means clustering algorithm is defined as follows:

$$J = \sum_{i=1}^{n}\sum_{j=1}^{c} d_{ij}^{2} \text{ ,} where\ d_{ij} = \|F_i - C_j\| \qquad (1)$$

d_{ij} is an indicator of the c data points distance from their respective cluster centers. It represents the distance between a data point F_i and the centroid cluster C_j.

By applying the k-means algorithm, the first step will be to arbitrarily initial several cluster centroids. Then, iteratively reassigns the data to clusters based on the objects' proximity to each center of the K clusters. We will stop the algorithm execution when a convergence criterion is attained.

III. RELATED WORK

Many research works have studied the positive impact of learning style to enhance the learning process in online education systems. In [4], authors have studied how learning styles influence learners' performance in e-learning systems. They studied the impact of using learning style data to build recommendations for learners, instructors, and contents of online courses.

In the light of this fact, many works have been conducted to find the appropriate learning style of a particular learner. In [5], the authors studied current directions made in the field of automatic learning styles detection. The authors reviewed, analyzed and summarized main findings from the studied works. They discussed works limitations, involvements and research challenges that can be considered as useful insights to find new methods and approaches in learning styles detection Area.

Authors in [4], Naïve-Bayes Tree as classification method to classify the learning style of learners as per FSLSM. As first step, learning styles are detected using the questionnaire index of Learning Styles[1] (ILS). This questionnaire is developed by Felder and Silverman. It consists of 44 questions for identifying the learning styles. However, it is too difficult to identify the

[1] https://www.webtools.ncsu.edu/learningstyles/

learning style of learners relying only on Felder and Silverman questionnaire. In fact, additional factors should be taken into account to find the appropriate learning style such as online usage behavior and learning objects' types.

In [7], authors have proposed a method which incorporates Artificial Neural Network technique based on a Multi-Layer Perceptron to classify the learner's data. The proposed method combines a set of behavior characteristics extracted from the learning sessions to increase the precision of the learning styles prediction. The classification step relies on classifying learner's data based on the four categories (Active-Reflective and Sequential-Global) of learners provided by FSLSM.

Authors in [6] presented a learning style classification technique to classify and identify students' learning styles. The students were classified according to three learning styles: dilatory students, transitory students and persistent students. This classification is based on students' behavior and the time spent on learning objects. Authors combined genetic algorithms with an improved K-NN to classify students according to their learning styles.

However, almost all works have taken into account the specific features of learning styles based on a particular learning style model. In fact, learning style models influence greatly the produced results. In addition, despite the positive impact that learning style models can play in recommendation phase to enhance the learning process, they are often ignored.

IV. Our approach

After extraction of learners' behaviors making best use of web usage mining techniques, we will follow the following steps: 1) mapping the learning objects onto FSLSM, 2) modeling the learner profile and finally, we describe our methodology to automatically detect and identify the learning styles of learners.

A. Mapping of LOs as FSLSM

In this work, we have opted for FSLSM bill for a number of reasons. Firstly, it has a questionnaire and a rating mechanism in order to detect and identify the learning style of a given learner. Secondly, it was validated in many adaptive E-learning environments.

The mapping of LOs as FSLSM require that all LOs need to be labeled according to the closest learning style category. To do this, we offer a mapping approach of LOs as FSLSM's categories. this approach is based on several related works, especially those of [16][17]. The obtained results of mapping LOs onto FSLSM are presented in Table I. Learning styles corresponds to FSLS dimensions. The mapped LOs are considered as input parameters to our applied clustering algorithm. We are therefore talking about K-means algorithm.

Table I. LEARNING OBJECTS AS PER FSLSM

Categories	LO as Video	LO as PPT	LO as Demonstration	LO as Exercise	LO as Assignment	LO as PDF	LO as Announcement	LO as Reference	LO as Example	LO as Practical Material	LO as Forum	LO as Topic list	LO as Image	LO as Chart	LO as Email	LO as Sequential
Active	√	√	√	√	√											
Reflective	√	√				√	√	√								
Sensing	√					√			√	√						
Intuitive	√	√				√		√			√	√				
Visual	√							√					√	√		
Verbal	√					√	√								√	
Sequential				√	√			√								√
Global				√	√			√				√				

Once the LOs are mapped, it is possible to recommend contents according to learners learning style.

B. Learning profile modeling

In this work, we build an ontology-based learning profile that integrates e-learning knowledge, FSLSM and learner information. The focus is to capture learners' behavior and using it to identify learners' learning style and make accurate recommendations. The use of ontology-based learner profile (OBLP) is argued by its capability to increase the interoperability and reuse of concepts. Furthermore, OBLP is used to annotate learning objects in order to facilitate contents recommendation task.

To build the learners' behavior ontology, we adopt the On-to-knowledge ontology development methodology [18]. The learning profile ontology and its relationships with other components are depicted in Fig. 1.

The concept *LearningObject* is used to model an elementary resource. The concept *Group* serves as a category for specifying relationships between learners. The concept *Learner* describes information about learners. The concept *Session* aims to provide information about the period of learning a specific course. The concept *sequence* represents a sequence of learning objects used for learning during a continuous time interval. Finally, the concept *FSLSM_category* serves to describe FSLSM in order to annotate a particular *LearningObject* based on our proposed mapping learning objects presented in Table **I**. A particular learning object can be presented in a web page or as a specific file.

To apply a classification/clustering algorithms, we have opted for an approach based on data mining techniques to capture the behaviors of a learner. Then we modeled the captured data as an ontology of learner profile. The various learning objects used by a given learner in his sessions represent the learner's behavior. They are characterized by the time that a particular learner has spent on visualizing each file and/or page in each sequence. The Time spent on learning objects is considered in seconds.

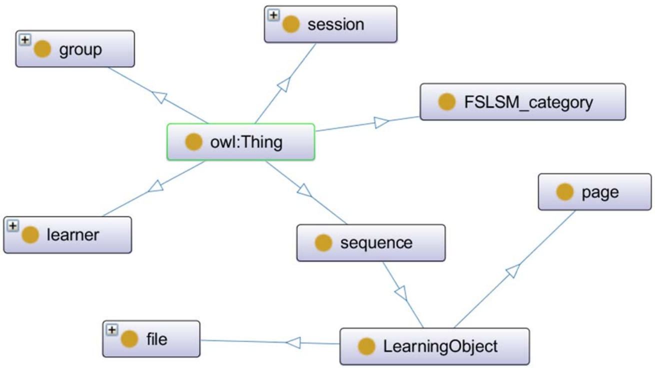

Fig. 1. Learning profile ontology

We bear in mind, that LOs have a very low level of granularity as resources. i.e., at level of text, diagram, drawing, figure, graphic, illustration, picture, etc. So, according to the preferred learning style of a learner, e-learning systems will produce personalized content.

About learning objects, which have higher level of granularity, they must be constructed based on LOs that have lower ones. e.g., a course is composed of a set of chapters, a chapter is divided on sections, while the section is composed of several little chunks and so on. Hence, using this method, we are capable of reusing LOs at all different levels.

In the next section, we will describe our methodology for clustering learners' sequences based on K-means algorithm in order to identify the appropriate learning style of each given learner.

C. Methodology

The main idea of our proposal relies on how to classify the profile of a particular learner according to learning style categories. Firstly, the learning behaviors have been captured and have been modeled as learning profile ontology. The learner's profile ontology contains labeled sequences based on FSLSM categories. Secondly, we propose to classify these captured sequences onto learning styles categories which represents the learning styles clusters. The proposed model is depicted in Fig. 2

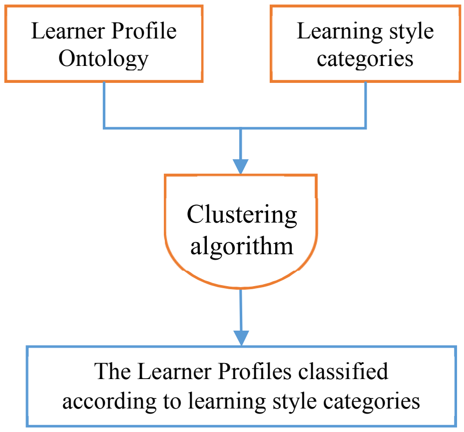

Fig. 2 Our Model for classifying Learner Profile according to Learning Style categories

In our work, we have utilized FSLSM as a learning style model since it is well known and widespread in adaptive E-learning systems. This is mainly due to its ability to estimate the dominant learning style of a given learner. In addition, FSLSM was validated in many educational environments. FSLSM proposes eight categories (Active, Reflective, Sensing, Intuitive, Visual, Verbal, Sequential and Global) to classify a particular learner.

To classify the learners' sequences, we used K-means clustering algorithm since it is uncomplicated, simple, easy to developed, and works fast in most situations. In addition, as the number of FSLS categories is known, it will be easier to run k-means algorithm. The k-means procedure is detailed in what follows:

Algorithm: K-means Clustering Algorithm

Initialize : Randomly select K sequences as the initial cluster centers $C = \{c_1, c_2, \ldots, c_k\}$

Input: a collection of captured sequences $S = \{s_1, s_2, \ldots, s_n\}$, number of clusters K

Output: an assignment matrix M of sequences to set of K clusters, M_{ij} = degree of member ship function x_i in cluster j

Repeat

Randomly choose $C = \{c_1, c_2, \ldots, c_k\}$ as initial centroids

Initialize M as zero

for all s_i in S do

let $j = argmin_{k \in \{1,2,\ldots,k\}} S(s_i, c_k)$

assign s_i to the custer j i.e. M_{ij}

end for

Update the cluster means as

$$C_j = \frac{\sum_{i=1}^{n} \sum_{j}^{k} M_{ij}^{\,n} s_i}{\sum_{i=1}^{n} \sum_{j}^{k} M_{ij}^{\,n}} \; for \; j = \{1, 2, \ldots, k\}$$

until meeting a given criterion function

V. Experiments and Results

To measure the effectiveness of our proposal, we conducted an experiment on students from computer science and Management school.

The learners' behavior data was collected from the E-learning platform of Sup'Management Group[2]. The online educational system was developed based on Moodle platform. Moodle[3] is a free Open Source software used to help educators and students to enhance the learning process.

A total of 126 students participated to this study. First, students filled in the questionnaire[4] based on the index of learning styles (ILS) developed by Felder-Silverman model. Second, we collected 1235 sequences from the E-learning platform, which represent the learners' behavior. The number of sequences varied from one to thirty-five sequences per students.

The sequences of learners constitute the input of the K-means clustering algorithm. The Center values for each cluster are computed based on the feature values based on mapping of learning objects to FSLSM categories. The result of the clustering is shown in Table II.

Table II RESULT OF K-MEANS CLUSTERING ALGORITHMS

Clusters	Active	Reflective	Sensing	Intuitive	Visual	Verbal	Sequential	Global
No. of Sequences clustered by K-means algorithm	169	134	127	162	256	118	144	125
Total	1235							

In order to compare the result of sequences clustering with the ILS questionnaire, we replaced each learner by its sequences. Fig 2 represents the classification of users' sequences according to k-means and ILSQ,

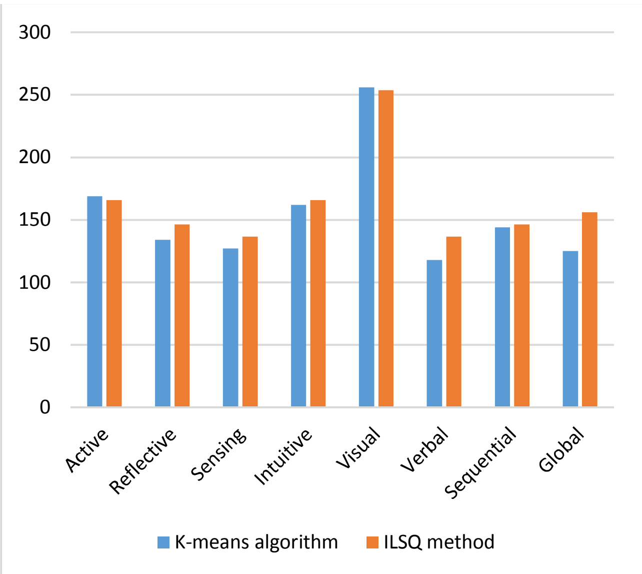

Fig. 3. Classification of users' sequences according to k-means and ILSQ methods

As we can see, the obtained results from the classification of k-means algorithm and the ILSQ method are approximatively similar.

A. Accuracy, Precision, Recall and F-Measure

In order to measure the goodness of our method, we use various validation metrics: Accuracy (A), Precision (P), Recall (R) and F-Measure (F1).

Precision (P): To find the number of input sequences which are correctly classified in a given class. Precision is calculated as in equation (2).

$$P = \frac{TP}{TP + FP} \qquad (2)$$

Recall (R): To find the number of sequences which are correctly identified in a given set of sequences. Recall is calculated as in equation (3).

$$R = \frac{TP}{TP + FN} \qquad (3)$$

Accuracy (A): To measure the performance by finding the ration of correctly predicted sequences. Accuracy is calculated as in equation (4).

$$A = \frac{TP + TN}{TP + TN + FP + FN} \qquad (4)$$

F-Measure also called F1 Score provides the weighted average of precision and recall. F1 is calculated as in equation (5):

$$F1 = \frac{2(R\ X\ P)}{(R + P)} \qquad (5)$$

$\mathbf{TP} \equiv \text{True} - \text{Positive} \equiv \text{sequences correctly identified}$

$\mathbf{TN} \equiv \text{True} - \text{Negative} \equiv \text{sequences correctly rejected}$

$\mathbf{FP} \equiv \text{False} - \text{Positive} \equiv \text{sequences incorrectly identified}$

$\mathbf{FN} \equiv \text{False} - \text{Negative} \equiv \text{sequences incorrectly rejected}$

The performance of our proposal can be found in Table **III**. We evaluate our approach using three metrics, precision recall and F1 measure. We have presented the obtained performance in each metric in the table. As we can see, our approach performed well in all metrics.

[2] http://www.supmanagement.ma/fc/
[3] https://moodle.org/
[4] https://goo.gl/forms/iVoTLzHa7Bp4n46x1

Table III THE PERFORMANCE OF OUR PROPOSAL

A (%)	P (%)	R (%)	F1 (%)
78.83	79.9	83.1	80.12

VI. CONCLUSION

In this work we have proposed an automatic approach for detecting students' learning style based on web usage mining. It consists in classifying students' log files according to the FSLSM as a specific learning style model. In our case, we used the K-means as a clustering algorithm. We captured the learners' behaviors using Web Usage Mining techniques. Then, we converted them into a set of sequences. These sequences are mapped to the eight categories of FSLSM and modeled as a learning profile ontology. The learner's profile ontology contains labeled sequences based on FSLSM categories. They form the input of the K-means clustering algorithm. In order to measure the effectiveness of our proposal, we use a dataset retrieved from the Sup'Management e-learning platform. We measure the performance of our work using various validation metrics: Accuracy, Precision, Recall and F1-Measure. The produced results show that our approach provides a very encouraging results.

REFERENCES

[1] S. Ninoriya, "CMS, LMS and LCMS For eLearning.," *Int. J. Comput. Sci. Issues*, vol. 8, no. March 2011, pp. 644–647, 2011.

[2] P. Brusilovsky and E. Millán, "User Models for Adaptive Hypermedia and Adaptive Educational Systems," *Adapt. Web*, pp. 3–53, 2007.

[3] J. W. Keefe, *Learning Style Theory and Practice*. National Association of Secondary School Principals, 1987.

[4] M. Abdullah, R. M. Bashmail, W. H. Daffa, M. Alzahrani, and M. Sadik, "The Impact of Learning Styles on Learner's Performance in E-Learning Environment," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 9, p. 8, 2015.

[5] J. Feldman, A. Monteserin, and A. Amandi, "Automatic detection of learning styles: state of the art," *Artif. Intell. Rev.*, vol. 44, no. 2, pp. 157–186, 2015.

[6] Y. C. Chang, W. Y. Kao, C. P. Chu, and C. H. Chiu, "A learning style classification mechanism for e-learning," *Comput. Educ.*, vol. 53, no. 2, pp. 273–285, 2009.

[7] A. Latham, K. Crockett, and D. Mclean, "Profiling Student Learning Styles with Multilayer Perceptron Neural Networks," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 2510–2515.

[8] J. B. MacQueen, "Kmeans Some Methods for classification and Analysis of Multivariate Observations," *5th Berkeley Symp. Math. Stat. Probab. 1967*, vol. 1, no. 233, pp. 281–297, 1967.

[9] R. M. Felder and L. K. Silverman, "LEARNING AND TEACHING STYLES," *J. Eng. Educ.*, vol. 78, pp. 674–681, 1988.

[10] H. A. E.-G. A. M. Riad, Hamdy K. El-Minir, "Review of e-learning systems convergence from traditional systems to services based adaptive and intelligent systems," *J. Converg. Inf. Technol.*, vol. 4, no. 2, p. 24, 2009.

[11] M. C. Murray and J. Pérez, "Informing and performing: A study comparing adaptive learning to traditional learning," *Informing Sci.*, vol. 18, no. 1, pp. 111–125, 2015.

[12] A. Y. Kolb and D. a Kolb, "The Kolb Learning Style Inventory — Version 3 . 1 2005 Technical Specifi cations," *LSI Tech. Man.*, pp. 1–72, 2005.

[13] L. CURRY, "Learning preferences in continuing medical education," *Can. Med. Assoc. J.*, vol. 124, no. 5, 1981.

[14] R. Keown, "Learning Objects: What Are They, and Why Should We Use Them in Distance Education?," *Distance Learn.*, vol. 4, no. 4, pp. 73–77, 2007.

[15] L. A. Mardis and C. Jo Ury, "Innovation – an LO library: reuse of learning objects," *Ref. Serv. Rev.*, vol. 36, no. 4, pp. 389–413, 2008.

[16] L. S. Mestre, "Matching up learning styles with learning objects: What's effective?," *J. Libr. Adm.*, vol. 50, no. 7, pp. 808–829, 2010.

[17] T. Chellatamilan, M. Ravichandran, R. M. Suresh, and G. Kulanthaivel, "Effect of mining educational data to improve adaptation of learning in e-learning system," in *International Conference on Sustainable Energy and Intelligent Systems (SEISCON 2011)*, 2011, pp. 922–927.

[18] R. (eds. . Staab, S.; Studer, *Handbooks on Ontologies*. 2007.