Toward a New Framework of Recommender Memory Based System for MOOCs

El Alami Taha[1], El Kadiri Kamal Eddine[2], Chrayah Mohamed[3]
[1,2]LIROSA Laboratory, AbdelMaled Essaadi University, Morocco
[3]TIMS Research Unit, LIROSA Laboratory AbdelMaled Essaadi University, Morocco

Article Info

Article history:

Received Jun 9, 2016
Revised Nov 20, 2016
Accepted Dec 11, 2016

Keyword:

Big data
Content base filtering
Distant learning
Moocs

ABSTRACT

Moocs is the new wave of remote learning that has revolutionized it since its apparition, offering the possibility to teach a very big group of student, at the same time, in the same course, within all disciplines and without even gathering them in the same geographic location, or at the same time; Allowing the sharing of all type of media and document and providing tools to assessing student performance. To benefit from all this advantages, big universities are investing in Moocs platforms to valorize their approach, which makes MOOC available in a multitude of languages and variety of disciplines. Elite universities have open their doors to student around the world without requesting tuition or claiming a college degree, however even with the major effort reaching to maximize students visits and hooking visitors to the platform, using recommending systems propose content likely to please learners, the dropout rate still very high and the number of users completing a course remains very low compared to those who have quit. In this paper we propose an architecture aiming to maximize users visits by exploiting users big data and combining it with data available from social networks.

Corresponding Author:

El Alami Taha,
LIROSA Laboratory,
AbdelMalek Essaadi University,
BP. 2121 M'Hannech II ,
93030 Tétouan Morocco.
Email: elalami.taha@gmail.com

1. INTRODUCTION

This article is part of a research work that falls within the research field of "Technology Enhanced Learning" (TEL) that studies the evolution of online learning, this research work aim to improve, perfecting methods and online learning platforms, especially the Massive Open Online Courses (MOOC) [1].

With MOOCs expansion, learners are exposed to various challenges and the traditional problem in TEL: “recommending the best learning resources” and “proposing it in the adequate timing” is more than ever up to date. In this article we are trying to respond to the question: how to hook student on Mooc platform? And how to lower the dropout rate in Mooc platforms?

Since nineteen's distance learning has seduced many students around the word and with the expansion of communication tools especially with Internet popularization, student are now well served considering the large possibilities allowed. Many student are tempt by distant learning experience and by the new born platform of Moocs. However the drop-out rate is still very high [2], studies mentioning a rate over 80% dropout [3]. To attract and retain user, suggestion through the platform or by e-mail, proposing material and courses that may interest users can be very useful, even more efficient if considering users state of mind that can be deduced from social networks [4].

Actually if we consider the sentimental factor as a coefficient to recommend courses, to web learners, we can provide content that will seem more close to students expectations, in the moment of time they are the most ready to learn, note that and unlike to existing researches this framework will have a big added value on the evolution of Moocs platforms, regarding utility, quality and efficiency in the way of reducing drop-out rate.

2. DISTANT LEARNING AND BIG DATA

2.1. Web 2.0

The word also used to refer to social software, it definition has always been a debate as it brings together the tools of production, communication and sharing, enabling the collaborator to contribute to the creation of content and the sharing of knowledge online [5]. Web 2.0 platforms are used in higher education for their ease of use, omnipresence, individual potentiality.

Since 90s websites designers proposed content that provoked users interaction, a Guest book section was a trend [6]. Web 2.0 tools are more likely interested in creating more connecting points between users, they are built to receive micro content and share interest; New trends appeared sharing photos using descriptive tags, users are collaborating authors in wiki pages, posting post and comment in blogs.

2.2. E-learning

E-learning came into use in the middle of 1990s along with development in the World Wide Web and interest in asynchronous discussion groups .The electronic learning “e-learning” is defined as instruction delivered on a digital device, such as computer or mobile device that is intended to support learning [7]. Many other definitions of e-learning have been proposed [8], Goodyear (2000) defined e-learning “as the systematic use of networked multimedia computer technologies to empower learners, improve learning, connect learners to people and resources supportive of their needs, and to integrate learning with performance and individual with organizational goals”.

In the same perspective, (Rosenberg, 2006; Sambrook, 2003) described E-learning as reference to the use of computer network technology, especially Internet, to hand information and instructions to learners. Thanks to its flexibility of access and its just-in-time delivery, e-learning is emerging as a popular approach for Before the expansion of internet, distance education was a one way communication, we all remember the educational radio and television programs, Magnetic tapes, mostly used for learning foreign languages, these methods doesn’t allow direct interaction between students and the institutions who are providing the teaching, so student questions remained unanswered, the main form was print or print+ broadcasting-based correspondence education.

The term learning in organizations or workplace settings.

The student is the center of e-learning system [9], he can present many benefits for learners:

a. Offers complete flexibility vis-à-vis the time and location constraints. « just in time - any time approach (Rosemberg, 2001);
b. Allows use of multimedia content, e-learning enhance comprehension and improve learning
c. Offers a personalized learning[10]
d. Offers the possibility to create learners community that allows dialogues and information exchange between learners or between learners-teachers[11]

2.3. Moocs

Massive open online courses (MOOCs) is the new phenomenon that has revolutionized remote learning (Figure 1), the term emerged in 2008 but have received large mainstream media coverage since 2012, "massive" refers to both students and courses, not only courses must be available for a significant number of student but it must also to propose a large materiel choices (project, Multiple choice, assessments, videos..) which can guarantee a good learning experience to all participants. The term "Open" stand out that no monetary cost requisite to participate in a course. "Online" used to notify that the course's element are hosted online, even if the majority of Moocs encourage students to form study group on social media or to set up meeting at a geographical locations as a manner to have a direct contact in addition to the virtual [12]. "Courses" defines in addition to designated time period over which the course progress, the necessity of registration to an instructional group.

MOOCs are capable of providing several ten thousands of learners with access to courses over the web (McAuly& al, 2010).Showing a great potential Moocs has gained so much attention and has attracted so many investments, Harvard and MIT developed a partnership through edX to improve online teaching and learning and build a global community of online learners, they jointly invested $60 million ($30 million each) to create a learning platform that will be presented as an open-source software so other universities and organizations be able to host the platform themselves and help Edx by proposing new adds and improvements.

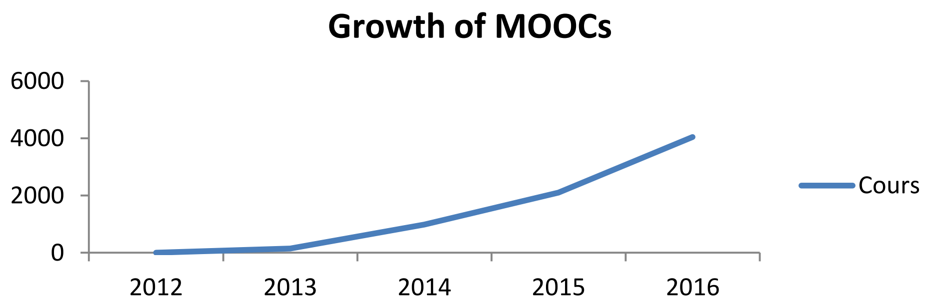

Figure 1. Shows the fast growth of MOOCs in the world (European MOOCs, 2016)

2.4. Big Data

Stocking and analyzing information was an always preoccupation, archeologist discovered in 1960 in Uganda The Ishango Bone, dated to the Upper Paleolithic era, they thought that it was the way used by Paleolithic tribes people to keep track of trading activity and supplies by marking notches into sticks or bones. the comparison of the notches on sticks using rudimentary calculations enabled them predictions such as how long their food supplies would last [13].

With the expansion of storage technologies and the emergence of web 2.0 each and every one of us is constantly producing and releasing data, according to IBM 90% of the data in the world today has been created in the last two years alone. This data come from everywhere: posts on social media, security cameras, cell phone GPS signal, commercial transactions, business data. The Table 1 below lists the main Big Data origin domains and targeted use or application, list not exhaustive, which illustrate that big data are present in every domain [14].

Table 1. Big data Origin and target use domains

Big data Origin	Big Data Target Use
1. Science	a. Scientific discovery
2. Telecom	b. New technologies
3. Industry	c. Manufacturing, process control, transport
4. Social media and networks	d. Personal services, campaigns
5. Living Environment, Cities	e. Living environment support
6. Healthcare	f. Healthcare support

Even if big data is the new buzz-word there is no exact definition, although scientific paper describe big data as having 5V properties:

2.4.1. Volume

Refers to the vast amount of data generated, new unit are being commonly used, the world is now talking about Petabytes and Zettabytes which are successively 10^{15} bytes and 10^{21} bytes [15], Facebook has 2.5 PB of user data (2009), eBay has 6.5 PB of user data (2009), fact that makes most datasets too large to store and analyze using traditional database technology. Distributed system dislocated around the world are used to store and analyze data [16].

2.4.2. Velocity

The velocity characteristic, refers to the speed of which new data is created, and the speed of which data is delivered. Now-a-days flow of data is enormous and continuous, fortunately actual technology do allow analyzing data while it's created, called it in-memory analytics, in order to save the time that will take the process of storing it into databases. NoSQL (Not Only SQL) is an eclectic and increasingly familiar group of non-relational data management systems, plays a major role in taking care of the velocity related challenges with big data.

2.4.3. Variety

Data are comes from a high variety of sources and in different formats and can contain multidimensional data field, so we can distingue different types : Structured, Semi-Structured and Unstructured Data.

a. -Structured data is the traditional Data-base that consists of Rows and Columns and resides in fixed fields in a file.
b. Unstructured data does not have and does not adhere to a pre-defined structure like E-mails, video and audio
c. -Semi-Structured Data does not confirm to a specific arrangement but consists of Ttags to separate the data elements.

2.4.4. Veracity

Veracity is the specification that define the quality of data and the level of trust in various data sources, it refers to data's messiness or trustworthiness. With the high velocity of big data the quality and accuracy are less controllable, the most visible and concrete example is the posts on social networks hash tags, colloquial speech and informational posts do not guarantee any quality of information and are not trustworthy sources.

2.4.5. Value

The last V refers to the most important characteristic of big data, which is the value that can be extracted from, many companies are starting to generate economic benefits from their big data and look for business value creation in terms of new products or services. In the last two American general elections, the capability of Barack Obama's campaign team to effectively wield big data analytics was seen a factor in his victory over his rivals.

3. RELATED WORK

Moocs are a big opportunity for researchers interested in learning methods, the platforms constitute a big datasets of students interactions. Considering the huge number of participants, large sets of students data can be analyzed to decipher users online behaviors and their commitment patterns (Coffrin et al. 2014).

Kizilcec et al. (2013) have used three Moocs on Coursera platform to used to understand students engagement and disengagement behaviors. Collected data like student watching video commitment or submission of assessment enabled the classification of engagement into four types labeled "on track", "behind", "auditing" and "out" that represent successively the "assignment was set on time", "delayed", "just seen videos and took tests bur did not the assignment" and "the students that did not participate in the course at all".

Anderson et al. (2014) have been interested by the relation between student engagement/activity and the finale grade. They note that the main characteristic of high achievement is the fact of watching so many videos on the platform. Same statement was mentioned by Karpicke and Blunt (2011) when they showed that the more the learners watch videos the highest learning performance he can get.

Santos et al. (2014) have showed that the students who are participation more on courses activities and are frequently communicating, discussing and collaborating with others have better chances passing the course.

To maximize performance recommending courses and videos is very important, most of recommender systems on Mooc platform are based on traditional recommendation approaches: collaborative-based ,content-based or the fusion of those two approaches (hybrid-based filtering), techniques have been shown to be useful in Mooc platform. Collaborative filtering use community data (feedback, ratings …) to make recommendations. Content-based study items that user has expressed his satisfaction over it and propose item with similar content. The main idea is to propose either items that satisfied similar users or to propose items similar to the ones that pleased the user. In some cases, educational recommender systems aim to support students at a specific time period; The presence of a time parameter affects the recommendation being retrieved.

Therefore, with our framework, we aim to attend a new level of recommendation where the automatic recommender system will get a little human sense and consider human state of mind as an influential attribute that may improve quality of recommended content.

4. RECOMMENDER SYSTEM

First webmasters problem is to attract users to their websites by presenting pleasant content and creating a brand image for their website so it can be visible on the web, the second problem is to keep users hooked up on the website and attaching them by offering content that pleases the user and the most close to their preferences.

Like YouTube or Amazon many other websites uses recommender system so they can keep surfers hooked on the website, so they can maximize their benefit, by suggestion item that may please to client for the case of e-commerce company as Amazon or EBay.

Many Methods are used to recommend content, the most popular are content based methods that focus on the content (item's description for commercial article, video metadata …) and collaborative filtering methods that focus on the other users feedback and interaction.

Recommender system are also a solution to facilitate research and economize time consumed to find information in a web, growing exponentially, to permit personalizing website according to users interest so they will be able to get information needed with minimum effort.

4.1. Content-based Filtering

Also referred to as cognitive filtering, propose items based on correlation between user past item choices and the content of each item, so it recommend items, similar to those previously selected, that match probably the most to user preferences.

Content based filtering are based on creating connections between items in a collection, so when user manifest preference for a specific object, system trace the item connections, recommend items with maximum degree of semblance [17]. Pure content-based recommendations do not take in consideration other users preferences (Schein, Popescul, &Ungar, 2002).

4.2. Collaborative Filtering

Is the process of collecting user feedback on a content (rating, like/dislike ...) and by comparing similarities and differences among several users profiles, we will be able to determinate how to recommend a content.

Collaborative filtering is the most commonly used method in making recommendation system. The basic idea is identifying group of users with similar preferences and recommending favorite items of the group to each user. In other words, the evaluation given by user to respective items based on the statistics of value, are used to identify the user group with closer preference, it is those that recommend items to other users of the group. the evaluation values are manifested explicitly and implicitly. Explicit manifestations are those obtained by explicitly expressing the evaluation of the item by user (clicking on the like or dislike button, rating note ...). Implicit manifestations is to show appreciation and interest by taken action on the item (downloading, sharing, revisiting, recommending to a friend, purchasing ...).

4.3. Hybrid filtering

Is the hybrid approach that combine collaborative to content-based filtering, this approach can be implemented in several ways:

a. Combining result of content based and collaborative based filtering executed separately.
b. Giving content-based method the capabilities of the collaborative-based one or the inverse.
c. Unifying into one model the two approaches.

Studies that compared the results of the hybrid method to the collaborative and the content based methods have confirmed the lead of hybrid approach to the two other pure approaches vis-a-vis the accuracy of results.these method is also a solution to the problematic of cold start and sparsity problem.

5. PROPOSED TEMPLATE

As we know feelings has big influences on actions taken by individuals, actually digital world is empty of emotion, and don't take the importance of state of mind in consideration even if all choices and decisions are based on. In this context and considering given facts our researches on Mooc dropout problematic, we propose a recommender system that will keep student attached to Mooc platform by using a system that will combine hybrid filtering algorithm to data deduced from Big Data to suggest courses to user according to his state of mind, and then we can propose easy courses when student is tired or difficult quizzes when student is happy.

Figure 2 shows the proposed platform for recommender system.

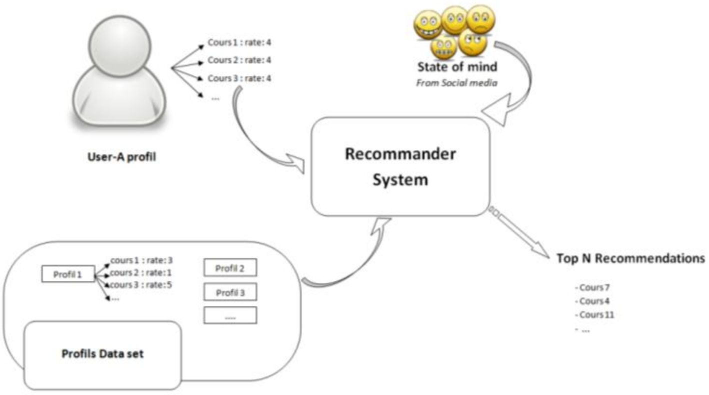

Figure 2. The proposed platform for recommender system

Our objective is not only proposing courses to student but proposing good courses that will be the most close to student taste, recommending the ones that users could not resist to their attraction and also to retain learner to the end of the course. To do so we have the delicate challenge of defining the most accurate standards and criteria's that can influencethe most and define users taste.

Our platform must also to exploit evaluations recorded from users experience on some given Moocs, to recommend the most suitable courses to other future users. To do so we will place a measure (usefulness) to every course Mooc $_j$, m $_{i,j}$[1,5] and to calculate it to a given user U $_i$. The usefulness matrix is the following Table 2:

Table 2. Usefulness matrix(user/Mooc)

F=

	$MOOC_1$	$MOOC_2$	....	$MOOC_m$
U_1	1			
U_2		5		3
U_3	3			2
....		4		1
U_n	4			

Our goal is to designate users with the high level of similarities to our student object of the case study, in order to recommend him a course that he is not likely to drop out from.

$$R_{u,Moocj} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Sim}\,(U, U_i)\,(U_{i,j} - \bar{U}_j) + \beta_u$$

With:

n: Number of users of the platform

u: User objet of case study

$Mooc_j$: course available on the platform

$U_{i,j}$: usefulness value given to the course $Mooc_j$by the user U i

$\bar{U}_j$: average of usefulness value given by users to $Mooc_j$

β_u: constant specific to user

$\mathrm{Sim}(U, U_i)$: function that defines users similarity

The similarity function is defined as follow:

$\mathrm{Sim}(U, U_i)$= $G(U, U_i)$+$ACL(U, U_i)$+ $STM(U, U_i)$¨+$J(U, U_i)$

where:

G: is gender

ACL: is the academic level
STM: is the state of mind deducted from users Big Data
J: is the professional activity

5.1. Gender

This variable that can have two value (Male/Female), it refers to sex's aspects related to identity and comportment, which goes beyond biological distinctness. Although since nineteen's and with the Globalization movement women are increasingly interested in and more opened to male-dominated field. However, the biological differences make men and women choices and preferences unsimilar. Within previous research, male students are more than female students, focused on specialties with technical and instrumental characteristics, female are more often focused on specialties with opportunities for relational aspects [18].

5.2. Academic Level

Next to gender, academic level may as well influence the way student is interested or not to register in a given MOOC, the selection of courses that will interest the student mastering a technique or a discipline will be different from the choice of a novice that aim to discover a field or a technique.

This variable may have 3 values: low (1), medium (1), high (1).

5.3. State of Min

All big decisions and choices in life are influenced by feeling and the state of mind. Within previous research, a large number of studies have already considered the impact of media sentiment and investor attention on financial markets these studies relate sentiment of the general population to the Dow Jones Industrial Average (DJIA) the New York exchange index, from what we can infer that the individual state of mind is a big factor influencing human choices.

5.4. Professional Activity

The professional activity is influential detail, because it determine how much and when the user may or may not have time for studying, a retired user is more likely willing to pursue a course than a company CEO.

5.5. Algorithm

We propose the following algorithm to make a list of the most accurate recommendations for a given user U:

Begin {algorithm}
-Fix n: number of users
-Fix m: number of Moocs

-Extract F[][]=$(U_{i,j})_{1\leq j\leq m}^{1\leq i\leq n}$
for $(1 \leq j \leq m)$ **do** Calculate $\overline{U}_j = \frac{1}{m}\sum_{i=1}^{m} U_{i,j}$
for $(1 \leq i \leq n)$ **do** Calculate $\mathrm{Sim}(U, U_i)$
for $(1 \leq i \leq n)$ Calculate β_u

for $(1 \leq i \leq n)$ **and** $(1 \leq j \leq m)$ Calculate $(R_{u,Moocj})$

Ascending Order$(R_{u,Moocj})_{1\leq i\leq n}^{1\leq j\leq m}$ vector

End

Our system is based on the memory and the history of the platform to define the algorithm parameters, the vector result of the algorithm form the order of recommendation that will be proposed to our user in accordance the Figure 3:

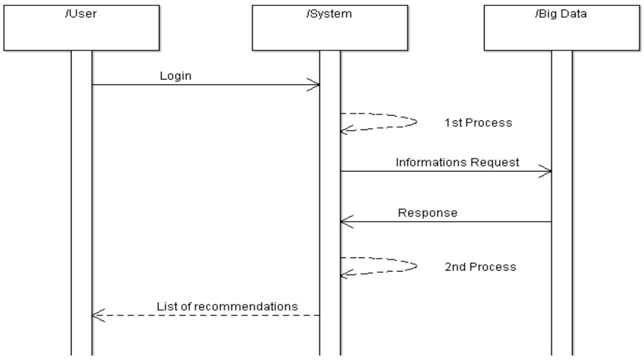

Figure 3. Sequence diagram of the proposed platform

6. CONCLUSION AND PERSPECTIVES

The proposed framework will take Moocs to a new perspective and will exploit students social media information especially the ones referring to student state of mind and take the sentimental side often forgotten as a significant factor to recommend right courses in the right time. the framework will also take other users judgments on content and match profiles using hybrid filtering algorithm to recommend quality courses in harmony with student aspirations at the right moment, when student feel ready for learning. As a perspective to this work we are working on improving the algorithm that will generate, based on user profile: his state of mind and the correlation between those two input, the suggestion that will please the most the users of our platform.

REFERENCES

[1] Bousbahi F, Chorfi H, "MOOC-Rec: A Case Based Recommender System for MOOCs", Procedia-Social and Behavioral Sciences, Volume 195, 3 July 2015, pp 1813-1822

[2] Anderson A, Huttenlocher D, Kleinberg J, Leskovec, *"Engaging with massive online courses"*, In Proceedings of the 23rd international conference on world wide web WWW '14 J. (2014) pp. 687–698.

[3] Jose Luis Santos, Erik Duval, David Gago, Luis Rodríguez, *"Success, activity and drop-outs in MOOCs an exploratory study on the UNED COMA courses"*, LAK '14 Proceedings of the Fourth International Conference on Learning Analytics And Knowledge March 24 - 28, 2014 pp 98-102

[4] Jolie Kennedy, "Characteristics of Massive Open Online Courses (MOOCs): A Research Review, 2009-2012", *Journal of Interactive Online Learning*, Volume 13, Number 1, Spring 2014

[5] A. Levine, B. Alexander, "Web 2.0 Storytelling: Emergence of a New Genre", *EDUCAUSE Review*, vol. 43, no. 6 (November/December 2008)

[6] Moe, Rolin, "The brief & expansive history (and future) of the MOOC: Why two divergent models share the same name", *Current Issues in Emerging eLearning*, Vol. 2: Iss. 1, Article2 2015.

[7] M. R. M. Veeramanickam, N. Radhika, "A Smart E-Learning System for Social Networking", *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 4, No. 3, June 2014, pp. 447-455

[8] B. Romdhane Emna, "Adoption of e-learning by learners: An empirical study by structural equations", 2016.

[9] Wang, M., Ran, W., Liao, J., Yang, S. J. H. "A Performance-Oriented Approach to E-Learning in the Workplace", Educational Technology & Society, 13 (4), 167–179.

[10] Mayer, R., Clark, R. "E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning, 3rd edn", Book 2012.

[11] R. Garriso, "E-Learning in the 21st Century: A Framework for Research and Practice", Book 2011.

[12] Kizilcec, R. F., Piech, C., & Schneider, E, "Deconstructing disengagement: Analyzing Learner Subpopulations in Massive open Online Courses", LAK '13 Proceedings of the Third International Conference on Learning Analytics and Knowledge Pages 170-179.

[13] Sachin Arun Thanekar, K. Subrahmanyam, A. B. Bagwan, "Big Data and MapReduce Challenges, Opportunities and Trends", *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 6, No. 6, December 2016, pp. 2911-2919

[14] S.F. Wambaa,S. Akterc, A. Edwardsd,G. Chopine, D. Gnanzouf, "How 'Big Data' can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study", *International Journal of Production Economics*, Volume 165, July 2015, Pages 234–246

[15] Ravi Narasimhan, Bhuvaneshwari T, "Big Data – A Brief Study", *International Journal of Scientific & Engineering Research*, Volume 5, Issue 9, September-2014.

[16] Mark Andrejevic, "The Big Data Divide", *International Journal of Communication 8 (2014)*, 1673–1689.

[17] Manouselis N, Drachsler H, Vuorikari R, Hummel H, Koper R, "Recommender Systems in Technology Enhanced Learning", Springer, Chapter Recommender Systems Handbook pp 387-415 October 2010

[18] Phil JM Heiligers, "Gender differences in Medical Students Motives and Career Choice", BMC Medical Education 2012 pp 12-82

BIOGRAPHIES OF AUTHORS

El Alami Taha, PhD Student at AbdelMalek Essaadi University, LIROSA Laboratory IT engineer from the National School of Applied sciences; Prefecture Mdiq-Fnideq IT Department chief - Ministry of Interior

El Kadiri Kamal Eddine, PhD doctor and professor of computer science at The Faculty of Sciences of Tétouan-Morocco. He is the Director of the ENSA School of engineers of Tetouan and the Director of LIROSA laboratory. He has published several articleson E-learning and Web 2.0. He is part of many boards of international journals and international conferences.

Chrayah Mohamed, PhD doctor and professor of computer science at ENSA School of engineers of Tétouan-Morocco. He has published many articles on E-learning and Web 2.0. He is part of many boards of international journals and international conferences.