# Mobile Education: Towards Affective Bi-modal Interaction for Adaptivity

doi:10.3991/ijim.v3i2.774

E. Alepis[1], M. Virvou[1] and K. Kabassi[2]
[1] University of Piraeus, Piraeus, Greece
[2] TEI of the Ionian Islands, Zakynthos, Greece

***Abstract*—One important field where mobile technology can make significant contributions is education. However one criticism in mobile education is that students receive impersonal teaching. Affective computing may give a solution to this problem. In this paper we describe an affective bi-modal educational system for mobile devices. In our research we describe a novel approach of combining information from two modalities namely the keyboard and the microphone through a multi-criteria decision making theory.**

***Index Terms*—Educational systems, affective computing, mobile systems, adaptive systems .**

## I. INTRODUCTION

In the last decade, there is a growing interest in mobile technology and mobile networks. As a result, a great number of services are offered to the users of mobile phones including education. In the fast pace of modern life, students and instructors would appreciate using constructively some spare time that would otherwise be wasted [12]. For example, when they are travelling or even when they are waiting in queue. Students and instructors may have to work on lessons at any place, even when away from offices, classrooms and labs where computers are usually located. Assets of mobile interaction include device independence as well as more independence with respect to time and place in comparison with web-based education using standard PCs.

Mobile education may be quite impersonal since the presence of a human instructor and human co-students may not be available. A remedy to this kind of problem may be given by providing affective interaction based on the user's emotional state. The recognition of emotions can lead to affective user interfaces that take into account the users' feelings and can adapt their behavior according to these feelings. Regardless of the various emotional paradigms, neurologists/psychologists have made progress in demonstrating that emotions play an important role in the process of decision making and action deciding [1]. Moreover, the way people feel may play an important role in their cognitive processes [2]. Recently, significant research effort has been put in the recognition of emotions of users while they interact with software applications. Picard points out that one of the major challenges in affective computing is to try to improve the accuracy of recognizing people's emotions [3].

Improving the accuracy of emotion recognition may imply the combination of many modalities in user interfaces. Indeed, human emotions are usually expressed in several ways. Human faces, people's voices, or people's actions may all show emotions.

Ideally, evidence from many modes of interaction should be combined by a computer system so that it can generate as valid hypotheses as possible about users' emotions [4]. It is hoped that the multimodal approach may provide not only better performance, but also more robustness [5]. As it is stated in [6], although the benefit of fusion (i.e., audio-visual fusion, linguistic and paralinguistic fusion, multi-visual-cue fusion from face, head and body gestures) for affect recognition is expected from engineering and psychological perspectives, our knowledge of how humans achieve this fusion is extremely limited.

In previous work, the authors of this paper have implemented and evaluated with quite satisfactory results emotion recognition systems, incorporated in educational applications ([7], [8]). As a next step we have extended our affective educational system by providing mobile interaction between the users and the system. In many situations this means that learning may take place at home or some other site, supervised remotely and asynchronously by a human instructor but away from the settings of a real class.

The main characteristic of the proposed mobile system is that it combines evidence from two modes, namely the mobile device's microphone and the keyboard, in order to identify the users' emotions. The results of the two modes are combined through a multi-criteria decision making method. More specifically, the system uses Simple Additive Weighting (SAW) [9] for evaluating different emotions, taking into account the input of the two different modes and selects the one that seems more likely to have been felt by the user. In this respect, emotion recognition is based on several criteria that a human tutor would have used in order to perform emotion recognition of his/her students during the teaching course.

In view of the above, in this paper we describe a novel mobile educational system that incorporates bi-modal emotion recognition through a multi-criteria theory. The two modes of interaction are the interaction through the mobile device's keyboard and the interaction through the users' voice. Users may use their mobile device in order to read parts of the theory about a particular lesson, as well as to take tests. In the case of the sort examinations about particular lessons, users may write their answers directly to their mobile device through the keyboard, or in other cases they may use the mobile device's microphone as a mode of interaction. The proposed system collects evidence from the two modes of interaction and analyses

it in terms of criteria for emotion recognition. The main focus of the paper is on presenting the empirical studies that are essential for the application of a multi-criteria model, which is used for making final assumptions about the recognition of one or more than one emotional states.

## II. AFFECTIVE INTERACTION IN MOBILE DEVICES

After a thorough investigation in the related scientific literature we found that there is a shortage of educational systems that incorporate multi-modal emotion recognition. Even less are the existing affective educational systems with mobile facilities. In [10] a mobile context-aware intelligent affective guide is described, that guides visitors touring an outdoor attraction. The authors of this system aim mainly at constructing a mobile guide that generates emotions. On the contrary our proposed system aims at recognizing the users' emotions through their interaction with a mobile device rather than generating emotions. As a second related approach we found that Yoon et al. [11] propose a speech emotion recognition agent for mobile communication service. This system tries to recognize five emotional states, namely neutral emotional state, happiness, sadness, anger, and annoyance from the speech captured by a cellular phone in real time and then it calculates the degree of affection such as love, truthfulness, weariness, trick, and friendship. In their approach only data from the mobile device's microphone are taken into consideration, while in our research we investigate a mobile bi-modal emotion recognition approach. Moreover, our proposed system is incorporated in an educational application and data pass through a linguistic and also a paralinguistic stage of analysis. This derives from the fact that in an educational application we should take into consideration how users say or type something (such as low or high voice, slow or quick typing speed), as well as what users say or type (such as correct answers or mistakes).

## III. OVERVIEW OF THE SYSTEM

The main architecture of the mobile bi-modal emotion recognition system is illustrated in figure 1. Participants were asked to use their mobile device and interact with a pre-installed educational application. Their interaction could be accomplished either orally (through the mobile device's microphone) or by using the mobile devices keyboard and of course by combining these two modes of interaction. All data are captured during the interaction of the users with the mobile device through the two modes of interaction and then transmitted wirelessly to the main server. All the input actions are used as trigger conditions for emotion recognition by the emotion detection server. Finally all input actions as well as the possible recognized emotional states are stored in the system's database.

The discrimination between the participants is done by the application that uses the main server's data base and for each different user a personal profile is created and stored in the data base. In order to accomplish that, user name and password is always required to gain access to the mobile educational application.

A snapshot of a mobile emulator, operated by a participant is illustrated in figure 2. Users may answer questions and take tests using the mobile system. They may write their answers through the mobile device's keyboard, or alternatively give their answers orally, using

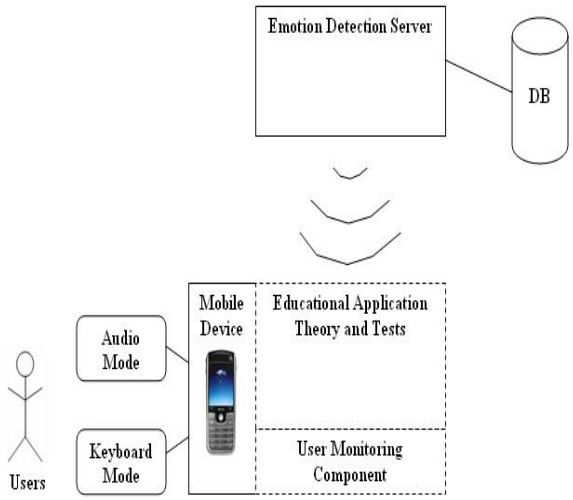


Figure 1. Architecture of the mobile emotion detection system.


Figure 2. A user is answering a question of a test either using the keyboard or orally through the mobile device's microphone.

their mobile device's microphone. In both cases, the data from the two possible modes of interaction are stored in the main system's database (emotion detection server), in order to be processed for emotion recognition purposes. When participants answer questions, the system tries to perform error diagnosis in cases where the participants' answers have been incorrect. Error diagnosis aims at giving an explanation about a participant's mistake taking into account the history record of the participant and the particular circumstances where the error has occurred.

## I. EMPIRICAL STUDY

The empirical study that we have conducted concerns the audio-lingual emotion recognition, as well as the recognition of emotions through keyboard evidence. The audio-lingual mode of interaction is based on using a mobile device's microphone as input device. The empirical study aimed at identifying common user reactions that express user feelings while they interact with mobile devices. As a next step, we associated these reactions with particular feelings.

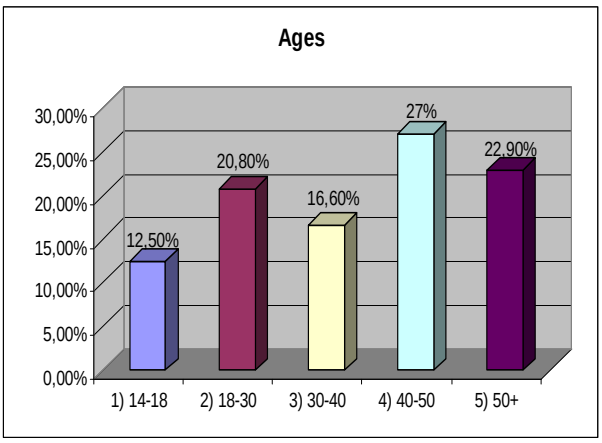


Figure 3. Distribution of the ages of the participants

The empirical study involved a total number of 200 male and female users of various educational backgrounds, ages and levels of familiarity with computers. Individuals' behavior while doing something may be affected by several factors related to their personality, age, experience, etc. Figure 3 illustrates the distribution of participants in the empirical study in terms of their age. In particular there were 12,5 % of participants under the age of 18, approximately 20% of participants between the ages of 18 and 30. A considerable percentage of our participants was over the age of 40.

The participants were asked to use a mobile educational application which incorporated a user monitoring component. The user monitoring component that we have used can be incorporated in any application, since it works in the background, recording silently each user's input actions. Our aim was not to test the participants' knowledge skills, but to record their oral and keyboard behavior. Thus, the educational application incorporated the monitoring module that was running unnoticeably in the background. Moreover, users were also videotaped while they interacted with the mobile application.

After completing using the educational application, participants were asked to watch the video clips concerning exclusively their personal interaction and to determine in which situations they were experiencing changes in their emotional state. Then, they associated each change in their emotional state with one of the six basic emotion states in our study and the data was recorded and time-stamped.

As the next step, the collected transcripts were given to 20 human expert-observers who were asked to perform audio and keyboard emotion recognition with regard to the six emotional states, namely happiness, sadness, surprise, anger, disgust and neutral. All human expert-observers possessed a first and/or higher degree in Psychology. These experts where first asked to analyze the data corresponding to the audio-lingual input only. To this end, they were asked to listen to the video tapes without seeing them. They were also given what the user had said in printed form from the computer audio recorder. The human expert-observers were asked to justify the recognition of an emotion by indicating the weights of the criteria that they had used in terms of specific words and exclamations, pitch of voice and changes in the volume of speech.

In the case of the keyboard mode, human expert-observers analyzed the data corresponding to the keyboard input separately. Thus, they were asked to watch the video tape and were also given a print out of what the user had written as well as the exact time of each event as it was captured by the user monitoring component. Human observers were asked to justify the recognition of an emotion by indicating the weights of the criteria that they had used in terms of specific changes in the speed of typing, in the absence of typing as well as to the written context. Correct answers, wrong answers as well as the consequence of these events were analyzed as long as these events involved only the keyboard mode of interaction. Frequent use of backspace and other basic keyboard buttons was also recorded and associated with specific emotional states.

Finally after processing the data from both the human experts and the monitoring component we came up with statistical results that associated user input action through the mobile keyboard and microphone with emotional states. More specifically, considering the keyboard we have the following categories of user actions: a) user types normally b) user types quickly (speed higher than the usual speed of the particular user) c) user types slowly (speed lower than the usual speed of the particular user) d) user uses the "delete" key of his/her mobile device often e) user presses unrelated keys on the keyboard f) user does not use the keyboard. These actions are considered as criteria for the evaluation of emotion with respect to the user's action in the keyboard.

Considering the users' basic input actions through the mobile device's microphone we have 7 cases: a) user speaks using strong language b) users uses exclamations c) user speaks with a high voice volume (higher than the average recorded level) d) user speaks with a low voice volume (low than the average recorded level) e) user speaks in a normal voice volume f) user speaks words from a specific list of words showing an emotion g) user does not say anything. These seven actions are considered as criteria for the evaluation of emotion with respect to what the user says.

Concerning the combination of the two modes in terms of emotion recognition we came to the conclusion that the two modes are complementary to each other to a high extent. In many cases the human experts stated that they could generate a hypothesis about the emotional state of the user with a higher degree of certainty if they had taken into account evidence from the combination of the two modes rather than one mode. For example, when the rate of pressing the mobile device's "deletion key" of a user increases, this may mean that the user makes more mistakes due to a negative feeling. However this hypothesis can be reinforced by evidence from speech if the user says something bad that expresses negative feelings.

## II. THE MULTI-CRITERIA MODEL

The input actions that were identified by the human experts during the empirical study provided information for the emotional states that may occur while a user interacts with an educational system. These input actions are considered as criteria for evaluating all different emotions and selecting the one that seems more prevailing. More specifically, each emotion is evaluated

first using only the criteria (input actions) from the keyboard and then only the criteria (input actions) from the microphone. In cases where both modals (keyboard and microphone) indicate the same emotion then the probability that this emotion has occurred is increased significantly. Otherwise, the mean of the values that have occurred by the evaluation of each emotion is calculated and the one with the higher mean is selected.

For the evaluation of each alternative emotion the system uses SAW for a particular category of users. This particular category comprises of the young (under the age of 19) and novice users (in computer skills). The likelihood for a specific emotion (happiness, sadness, anger, surprise, neutral and disgust) to have occurred by a specific action is calculated using the formula below:

$$\frac{em_{1e_1 1} + em_{1e_1 2}}{2}$$ where

$$em_{1e_1 1} = w_{1e_1 k1}k_1 + w_{1e_1 k2}k_2 + w_{1e_1 k3}k_3 + w_{1e_1 k4}k_4 + w_{1e_1 k5}k_5 + w_{1e_1 k6}k_6 \quad (1)$$

$$em_{1e_1 2} = w_{1e_1 m1}m_1 + w_{1e_1 m2}m_2 + w_{1e_1 m3}m_3 + w_{1e_1 m4}m_4 + w_{1e_1 m5}m_5 + w_{1e_1 m6}m_6 + w_{1e_1 m7}m_7 \quad (2)$$

$em_{1e_1 1}$ is the probability that an emotion has occurred based on the mobile keyboard actions and $em_{1e_1 2}$ is the probability that refers to an emotional state using the users' input from the mobile device's microphone. These probabilities result from the application of the decision making model of SAW and are presented in Eqs.(1) and (2), respectively. $em_{1e_1 1}$ and $em_{1e_1 2}$ take their values in [0,1].

In Eq.(1) the k's from k1 to k6 refer to the six basic input actions that correspond to the keyboard. In Eq.(2) the m's from m1 to m7 refer to the seven basic input actions that correspond to the microphone. These variables are Boolean. In each moment the system takes data from the bi-modal interface and translates them in terms of keyboard and microphone actions. If an action has occurred the corresponding criterion takes the value 1, otherwise its value is set to 0. The w's represent the weights. These weights correspond to a specific emotion and to a specific input action and are acquired by the constructed database.

In order to identify the emotion of the user interacting with the mobile system, the mean of the values that have occurred using Eqs.(1) and (2) for that emotion is estimated. The system compares the values from all the different emotions and determines whether an emotion is taking effect during the interaction. As an example we give the two formulae with their weights for the two modes of interaction that correspond to the emotion of happiness when a user (under the age of 19) gives the correct answer in a test of our educational application. In case of $em_{1e_1 1}$ considering the keyboard we have:

$$em_{1e_1 1} = 0.4k_1 + 0.4k_2 + 0.1k_3 + 0.05k_4 + 0.05k_5 + 0k_6$$

In this formula, which corresponds to the emotion of happiness, we can observe that the higher weight values correspond to the normal and quickly way of typing. Slow typing, often use of the backspace key and use of unrelated keys are actions with lower values of stereotypic weights. Absence of typing is unlikely to take place. Concerning the second mode (microphone) we have:

$$em_{1e_1 2} = 0.06m_1 + 0.18m_2 + 0.15m_3 + 0.02m_4 + 0.14m_5 + 0.3m_6 + 0.15m_7$$

In the second formula, which also corresponds to the emotion of happiness, we can see that the highest weight corresponds to m6 which refers to the 'speaking of a word from a specific list of words showing an emotion' action. The empirical study gave us strong evidence for a specific list of words. In the case of words that express happiness, these words are more likely to occur in a situation where a novice young user gives a correct answer to the system. Quite high are also the weights for variables m2 and m3 that correspond to the use of exclamations by the user and to the raising of the user's voice volume. In our example the user may do something orally or by using the keyboard or by a combination of the two modes. The absence or presence of an action in both modes will give the Boolean values to the variables k1…k6 and m1…m7.

A possible situation where a user would use both the keyboard and the microphone could be the following: The specific user knows the correct answer and types in a speed higher than the normal speed of writing. The system confirms that the answer is correct and the user says a word like 'bravo' that is included in the specific list of the system for the emotion of happiness. The user also speaks in a higher voice volume. In that case the variables k1, m3 and m6 take the value 1 and all the others are zeroed. The above formulae then give us $em_{1e_1 1} = 0.4 * 1 = 0.4$ and

$$em_{1e_1 2} = 0.15 * 1 + 0.3 * 1 = 0.45 .$$

In the same way the system then calculates the corresponding values for all the other emotions using other formulae. For each basic action in the educational application and for each emotion the corresponding formula have different weights deriving from the stereotypical analysis of the empirical study. In our example in the final comparison of the values for the six basic emotions the system will accept the emotion of happiness as the most probable to occur.

## III. EVALUATION

The 200 participants that were involved in the first phase of the empirical study in section 3 were also used in the second phase of the empirical study for the evaluation of the multi-criteria emotion recognition mobile system. In this section we present and compare results of successful emotion recognition in mobile audio mode, mobile keyboard mode and the two modes combined. For the purposes of our study the whole interaction of all users with the mobile educational application was video recorded. Then the videos collected were presented to the users that participated to the experiment in order to perform emotion recognition for themselves with regard to the six emotional states, namely happiness, sadness, surprise, anger, disgust and the neutral emotional state. The participants as observers were asked to justify the recognition of an emotion by indicating the criteria that s/he had used in terms of the audio mode and keyboard

actions. Whenever a participant recognized an emotional state, the emotion was marked and stored as data in the system's database. Finally, after the completion of the empirical study, the data were compared with the systems' corresponding hypothesis in each case an emotion was detected.

TABLE I.
RECOGNITION OF EMOTIONS USING THE MULTI-CRITERIA THEORY.

| **Using Stereotypes and SAW** | | | |
|---|---|---|---|
| Emotions | Audio mode recognition | Recognition through keyboard | Multi-criteria bi-modal recognition |
| Neutral | 17% | 32% | 46% |
| Happiness | 52% | 39% | 64% |
| Sadness | 65% | 34% | 70% |
| Surprise | 44% | 8% | 45% |
| Anger | 68% | 42% | 70% |
| Disgust | 61% | 12% | 58% |

Table 1 illustrates the percentages of successful emotion recognition of each mode after the incorporation of modes' weights and the combination through the multi-criteria approach.

The results in table 1 indicate that the application of the multi-criteria model lead our system to noticeable improvements in its ability to recognize emotional states of users successfully.

## IV. CONCLUSIONS

In this paper, we described a mobile affective educational application that recognizes users' emotions based on two modes of interaction, namely the mobile device's microphone and the keyboard. The proposed system uses an innovative approach that combines evidence from the two modes of interaction based on a multi-criteria decision making theory to improve the system's accuracy in recognizing emotions. Real users participated in the first experiment as well as human experts and aimed at revealing the criteria that are taken into account for emotion recognition in mobile devices.

For the evaluation of the affective mobile system a second experimental study with the participation of human experts was conducted and its results revealed noticeable improvements in the proposed system's ability to recognize emotional states of users.

The results of these experiments also showed that the criteria used for emotion recognition in mobile devices were similar to the criteria that were identified by other experimental studies [7] for emotion recognition in personal computers. As a result, these findings may also be used by other systems that perform emotion recognition in other domains. However, what may differ in the application of these criteria in other systems is their weight of importance in emotion recognition. In such a case, the setting of the second experiment should be repeated for the new domain.

## REFERENCES

[1] E. Leon, G. Clarke, V. Gallaghan, F. Sepulveda, "A user-independent real-time emotion recognition system for software agents in domestic environments", Engineering applications of artificial intelligence, 20 (3): 337-345, 2007. (doi:10.1016/j.engappai.2006.06.001)

[2] D. Goleman, "Emotional Intelligence", Bantam Books, New York 1995.

[3] Picard, R.W., "Affective Computing: Challenges", Int. Journal of Human-Computer Studies, Vol. 59, Issues 1-2, 55-64, 2003.

[4] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information", In Proceedings of the 6th international Conference on Multimodal interfaces (State College, PA, USA, October 13 - 15, 2004). ICMI '04. ACM, New York, NY, 205-211, 2004.

[5] M. Pantic, L.J.M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction", Vol. 91, Proceedings of the IEEE 1370-1390, 2003.

[6] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: audio, visual and spontaneous expressions", In Proceedings of the 9th international Conference on Multimodal interfaces (Nagoya, Aichi, Japan, November 12 - 15, 2007). ICMI '07. ACM, New York, NY, 126-133, 2007.

[7] E. Alepis, M. Virvou, and K. Kabassi, "Knowledge Engineering for Affective Bi-modal Human-Computer Interaction", Sigmap, 2007.

[8] E. Alepis, and M. Virvou, "Emotional Intelligence: Constructing user stereotypes for affective bi-modal interaction", Lecture Notes in Computer Science, Volume 4251 LNAI - I, Pages 435-442, 2006.

[9] P.C. Fishburn, "Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments", Operations Research, 1967.

[10] M.Y. Lim, and R. Aylett, "Feel the difference: A guide with attitude!", Lecture Notes in Computer Science, Volume 4722 LNCS, Pages 317-330, 2007.

[11] W.J. Yoon, Y.H. Cho, and K.S. Park, "A study of speech emotion recognition and its application to mobile services", Lecture Notes in Computer Science, Volume 4611 LNCS, Pages 758-766, 2007.

[12] M. Virvou, and E. Alepis, "Mobile educational features in authoring tools for personalised tutoring" Computers and Education, 44 (1), 2005, Pages 53-68, 2005.

## AUTHORS

**E. Alepis** is with the University of Piraeus, Department of Informatics, Piraeus 18534, Greece. He is a Ph.D student in the Department of Informatics. He received a first degree in Informatics in 1998. He has authored over 25 articles, which have been published in international conferences, books and journals. He has served as reviewer in international conferences. His research interests are in the areas of Affective Computing, E-learning, M-learning, User Modeling and Human Computer Interaction (e-mail: talepis@ unipi.gr).

**M. Virvou**, is with the University of Piraeus, Department of Informatics, Piraeus 18534, Greece. She is an Associate Professor in the Department of Informatics, University of Piraeus, Greece. She received a degree in Mathematics from the University of Athens, Greece (1986), a M.Sc. (Master of Science) in Computer Science from the University of London (University College London), U.K.(1987) and a D.Phil. from the School of Cognitive and Computing Sciences of the University of Sussex, U.K.(1993). She is the sole author 3 computer science books ("Object Oriented Software Engineering", "Object Oriented Analysis and Design" and "Introduction to Compilers"). She has authored or co-authored over 150 articles, which have been published in international journals, books and conferences. The international journals and conferences where she has published articles are in the areas of Knowledge based Software

Engineering, Web-based information systems, Computers and Education, Personalization Systems, Human Computer Interaction, Student/User Modeling. She has served as a member of Program Committees and/or reviewer of International journals and conferences. She has supervised or currently supervising 12 Ph.D.s in the areas of web-based information systems, knowledge-based human computer interaction, personalization systems, software engineering, e-learning, e-services and m-services. She has served or is serving as the project leader and/or project member in 15 R&D projects in the areas of e-learning, computer science and information systems. In the year 2003-2004 she received the first research award from the Research Center of her University as the member of faculty (among 200 colleagues of hers) having the highest number of research publications in high quality research journals. Many articles of hers have been among the top 5 most downloaded articles of the respective journals where they were published. She is a member of the IEEE. (e-mail: mvirvou@unipi.gr).

**K. Kabassi** is with the Technological Educational Institute of the Ionian Islands, Zakynthos, Greece. She is a Lecturer in the Department of Ecology and the Environment, Technological Educational Institute of the Ionian Islands. She received a first degree in Informatics (1999) and a D.Phil. (2003) from the Department of Information, University of Piraeus (Greece). She has authored over 40 articles, which have been published in international journals, books and conferences. She has served as a member of Program Committees and/or reviewer of International conferences. Her current research interests are in the areas of Knowledge based Software Engineering, Human Computer Interaction, Personalization Systems, Multi-Criteria Decision Making, User Modeling, Web-based Information Systems and Educational Software. (e-mail: kkabassi@teiion.gr).

The article was modified from a presentation at the ICDIM 2008 conference, November 13-16, 2008 hosted by the University of East London, London, UK. Manuscript received Manuscript received 16 December 2008. Published as submitted by the authors.