Available online at www.sciencedirect.com
ScienceDirect
Computers & Education xxx (2007) xxx–xxx
COMPUTERS & EDUCATION
www.elsevier.com/locate/compedu

An intelligent e-learning system based on learner profiling and learning resources adaptation

Paraskevi Tzouveli *, Phivos Mylonas, Stefanos Kollias

National Technical University of Athens, School of Electrical and Computer Engineering, Athens, Greece

Received 7 February 2007; received in revised form 4 May 2007; accepted 4 May 2007

Abstract

Taking advantage of the continuously improving, web-based learning systems plays an important role for self-learning, especially in the case of working people. Nevertheless, learning systems do not generally adapt to learners' profiles. Learners have to spend a lot of time before reaching the learning goal that is compatible with their knowledge background. To overcome such difficulties, an e-learning schema is introduced that adapts to the learners' ICT (Information and Communication Technologies) knowledge level. The IEEE Reference Model (WG 1) defined by the Learning Technology Standards Committee (LTSA) is extended and used for this purpose. The proposed approach is based on the usage of electronic questionnaires (e-questionnaires) designed by a group of experts. Through the automatic analysis of the learners' responses to the questionnaires, all learners are assigned to different learner profiles. According to these profiles they are served with learning material that best matches their educational needs. We have implemented our approach in five European countries and the overall case study illustrates very promising results.

Keywords: E-learning; Learners' profiles; E-questionnaire; Personalization; Educational resources adaptation

1. Introduction

It is a common fact that technological developments cause continuous changes to every sector of the modern society. Education itself could not remain passive and unconcerned; all traditional teaching techniques are revised and re-evaluated and the new ones are introduced. Internet-oriented applications try to satisfy current educational needs (Brusilovsky, 2001) by closing the gap between traditional educational techniques and future trends in technology-blended education. E-learning (Rosenberg, 2002) forms the revolutionary and new way to empower a workforce with the necessary skills and knowledge (Dagger, Wade, & Conlan, 2003; Karpouzis, Caridakis, Fotinea, & Efthimiou, 2005).

Towards this goal, various e-learning systems have been developed during the last years, however, most of them form static old-fashioned applications, missing functionalities like educational multimedia environ-

* Corresponding author.
E-mail address: tpar@image.ntua.gr (P. Tzouveli).

0360-1315/$ - see front matter
doi:10.1016/j.compedu.2007.05.005

ments, personalized capabilities and tracking of learners' input and relevance feedback (Brusilovsky, 2001). Personalized support for learners becomes even more important, when e-learning takes place within open and dynamic learning and information networks. However, the majority of e-learning systems (Dagger et al., 2003; Henze & Nejdl, 2001; Weber, Kuhl, & Weibelzahl, 2001), model the learner as an entity accompanied by a static predefined set of interests and options, without giving the appropriate attention to their needs. In addition, personalized learning using distributed information in dynamic learning environments is still an unsolved problem in e-learning research. Several approaches in the direction of adaptive web-based educational systems (Henze & Nejdl, 2001; Weber et al., 2001), are currently investigated, trying to offer personalized access and presentation facilities to learning resources for specific application domains. The lack of automatic evaluation of learner profile and the providing of appropriate learning resources is a common phenomenon in these systems. Additionally, none of these systems make a statistical analysis over the profiling in order to enhance their performance.

In our approach, the key point is determining the learners' interests through a pre-evaluation process, monitoring their progress and filtering the offered educational material according to those interests (Tzouveli & Kollias, 2004). Based on this approach, we propose an e-learning environment in the framework of SPERO (SPERO) that adapts to the learner's ICT level and knowledge. It is based on an e-learning schema, considering the IEEE Reference Model (WG 1) of the LTSA so as to model its architecture. Firstly, we use electronic questionnaires (e-questionnaires) designed by field experts aiming at detecting the learner's ICT level and learning preferences prior, during and after the learning process. In this way, the e-learning system gathers educational information and automatically estimates the ICT level of teachers and students through a visual, web-based, learner-friendly interface, as well as a personalized, profile-based scheme assisting it.

The personalization part has been designed to enable teachers and students to gracefully increase their ICT knowledge, providing them with appropriate e-courses, according to dynamically updated, profiling information. A novel mechanism is proposed, that creates updates and uses learners' profiles, extracted directly from their preferences and their usage history. Learners' preferences are utilized for personalization of the multimedia educational content offering and retrieval process, aiming at suitable content delivery through the integrated e-learning system. With the aim of having a more accurate and reliable profile extraction, a novel clustering algorithm is also implemented within the profiling procedure. Finally, taking into consideration learners' answers to the above mentioned questionnaires, statistical information is calculated and evaluated.

The structure of the paper is as follows. In Section 2 we present the system architecture, while in Section 3 we provide the profiling and evaluation procedure of the proposed system. In Section 4, the SPERO system is described as a case study of the proposed methodology, while in Section 5 results based on real-life experiments are provided. Finally, Section 6 gives our concluding remarks and describes briefly future work.

2. System architecture

2.1. General architecture of the system

The general architecture of the proposed e-learning system includes three main components that form standalone entities, while at the same time collaborate with each other. The first entity is the *Learner* or in most cases the *Group of Learners.* It includes single learners, participating in the learning process as individuals, as well as learners participating in the educational process in groups, e.g. collaborative learning.

The second entity is formed by the *Group of Experts*, which contains teachers both from general or special education, experts in the e-learning field, computer engineers, IT managers, data and statistical analysts, and psychologists. The group's goal is the design and continuous improvement of the system, with respect to computational integrity and educational efficiency. It is noticeable, that the Experts' Group is responsible for determining the learning topics that the e-learning schema can provide to the learners' group to which is addressed.

The last entity is formed by the *Server System*, including all hardware and software systems. Each subcomponent of this entity can be part of a centralized computer system or rely on separate hardware units scattered around some network (e.g. the Internet). It also includes the database of the system, which is considered to be the heart of the proposed system.

Please cite this article in press as: Tzouveli, P. et al, An intelligent e-learning system based on learner profiling ..., *Computers & Education* (2007), doi:10.1016/j.compedu.2007.05.005

2.2. Internal architecture of the system

The internal architecture of the proposed system is based on the IEEE Reference Model (WG 1) of the LTSA (Fig. 1). The IEEE-WG1-LTSA (P1484.1 Architecture & Reference Model) contains three types of entities: *processes*, *stores* and *flows*. In particular, processes (e-learner entity, evaluation, coach, delivery), treat the information received from the stores entities via flows. The stores (learner records, learning resources) implement inactive system components used as an information repository, while the flows (learning preferences, behaviors, assessment information, performance information, queries, catalogue info, locators, learning content, interaction context) denote the transfer of information (control or data) from one subsystem to another.

In the generic approach to e-learning systems outlined in the LTSA Draft Standard, a system's ability to adapt its operation to the learner is not defined, although an evaluation process exists. To handle this issue, we introduce in Fig. 1 an extension of the IEEE LTSA through an e-questionnaire, which achieves learner's adaptation to the proper learning topic and course. The extension of the IEEE LTSA draft (P1484.1 Architecture & Reference Model) is implemented by adding two stores, *Questionnaires Texts* and *Learner Profiles*, a process called *E-survey* and by enriching all process modules with new functionalities, on the purpose of achieving effective collaboration between them.

2.3. Data flow of the internal architecture

In this subsection we describe the internal data flow between the entities of the proposed system. Firstly, the group of experts defines the content and the forms of questionnaires providing questions which will extract the learners' knowledge level in a specific learning subject. More specifically, the store associated to the questionnaire texts can contain multiple questionnaires. Each one of these questionnaires has a predefined number of questions; these may be of various types (e.g. multiple choices). Additionally, the group of experts assign weights to the answers of each of question. Then, this final form of the questionnaire texts is stored in the questionnaire texts store; that is together with the weights of the answers.

In our system, the delivery process has been enriched, taking as input the texts stored into the questionnaire texts store. Then, it creates and dispatches the multilingual questionnaire presentation (in hypertext mode) to the learner entity. This presentation is illustrated to the learners through the multimedia flow (Fig. 1) on their first access to the system (SPERO questionnaires).

The learner process is considered as input to the system and forms the main entity of it. The learner initially receives the questionnaire from the delivery entity and then is called to answer. The learner entity's observable behavior and answers to the questionnaire are given as input to the evaluation process.

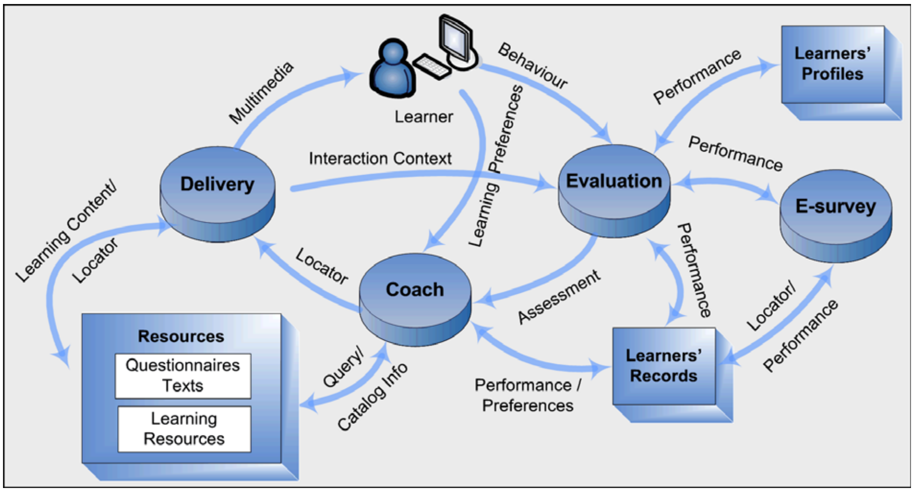

Fig. 1. Extension of LTSA IEEE learning system.

Please cite this article in press as: Tzouveli, P. et al, An intelligent e-learning system based on learner profiling ..., *Computers & Education* (2007), doi:10.1016/j.compedu.2007.05.005

The evaluation process produces assessment information and sends it to the coach process. In addition, the evaluation process creates a performance information flow to and from the learner records store. Performance information can derive from both the evaluation process (answers to the questionnaire, grades on lessons, learners' profile) and the coach process (certifications). The learner records hold information about the past (historical learner records), but can also hold information about the present (current assessments for suspending and resuming sessions) and the future (pedagogy, learner, or employer objectives).

In order to extract learner profiles, the evaluation process computes the values of the answers that each learner has given, taking into consideration the learners' profile store. This computation is necessary in order to find out the learner's level of knowledge in a specific subject and classify the learner's profile, which is then stored in the learner records. In addition, we design a new store, *learner profiles*, which contain the current profiles of the e-learning system.

In a general case, learner profiles are initially defined by the experts and are stored in the learner profiles store. Each one of them describes characteristics of learners, learning needs, and preferences. The answers to the questionnaires are used by the evaluation process to select the profile of a new learner, finding the existing learner profile that best matches the learner's ICT level. Thus, once a learner is assigned to a learner profile, the coach uses information in order to locate the learning material from the learning resources store that best match learner's profile.

A new learner profile can be created or existing ones can be adapted based on statistics extracted from the e-questionnaire database. New learner profiles or adapted versions of them are stored in the learner profiles store. Change of learners' profiles can be performed during their training, updating the learner records store. The entity coach can receive performance information from the learner records at any time. Performance information, such as assessment information, certifications, and preferences are stored in the learner records by the coach process. Based on this information, the coach process generates queries and forwards them to the learning resources store, in order to request learning materials that are appropriate for each learner.

The learning resources store is a database that represents knowledge, information, and other resources used in the learning experiences. The learning resources store replies to the coach with catalog info, which may be used by the delivery process to retrieve learning contents from the corresponding store. Finally, the delivery process transforms information obtained via learning content store into a presentation, transferred to the learner entity via a multimedia flow. Finally, the entities: questionnaires texts, learning resources, learners' records, and learners' profiles are stored in the database of the server.

The system allows an electronic survey to be conducted based on learners' answers to the questionnaires. The main role of this component is the presentation of statistical analysis results, conducted upon the already stored learner records. Conclusions derived from these e-surveys, as well as feedback from the use of the e-questionnaires, are fed back to the system, possibly modifying each learner's ICT level. The results of the system's e-survey are presented to any interested user or organization via the Internet (SPERO e-survey).

3. Profiling and evaluation procedure

One of the technical novelties introduced in the proposed e-learning system is the handling of its learners in a personalized manner, by building different profiles according to their ICT knowledge and expertise. The current section presents the design and implementation of a profile-based framework, which matches content to the context of interaction, so that the latter can be adapted to its learner's needs and capabilities. Our approach evolves in steps, tackling initially a static, basic version of the questionnaires and extending it to an automatic, intelligent information management system.

3.1. Static learner profiling scheme

The basic profile generation approach of the proposed system consists of a static profiling mechanism (Tzouveli & Kollias, 2004). The group of experts has defined a set of three – automatically generated – learner categories, forming the basis of the static profiling representation: *Beginner*, *Advanced* and *Expert* learners. Their basis is information provided by the learners' input data, obtained through their responses to e-questionnaires. In this manner, a continuous data collection, data processing, and profile updating are achieved, used mostly for the learners' training. The profiling mechanism creates and updates learners' profiles, mapping specific e-questionnaires question triggers, to particular static profiling. It is designed in a way that facilitates the utilization of the learners' preferences for profile creation and the recording of the learners' history and the learners' replies to the questionnaires.

The basic version of the learners' profile is represented by the *LearnerProfile*. The *LearnerProfile* component contains three elements. The first, *LearnerPreferences* stores the preferences of the learner (useful and unusable I/O devices, learning styles, physical limitations). The second one, *LearnerHistory* stores information about the learner's browsing history, the learner's performance information (grades, interim reports, log) and the learner's portfolio information (accomplishments and works).

The third element, *LearnerReplies* stores the learner's replies to the questionnaire itself or directly from the input of the learner and contains two sub-elements. The first sub-element QuestionID holds all the degrees of relevance to the three basic profile types. The second sub-element LearnerInput contains the learner's answers to the questionnaire. The *LearnerReplies* holds the total grade that the learner receives after answering the e-questionnaires. Depending on the particular question, as well as the part of the e-questionnaire that this question belongs, different degrees are counted. According to the possible answers, appropriate mapping to an available static profile is performed by the system, characterizing the learner (for the specific questionnaire) as Beginner, Advanced or Expert learner. The core of the methodology is based on comparison of the provided numerical values with the range of values a priori associated with the profiles. This degree is stored in the LearnerInput element, associated with the specific QuestionID part of it.

3.2. Adaptation of learning material

The overall system is able to continuously adapt to its learners preferences, usage history, and specific needs as they evolve during its usage. The developed software is a web-based learning portal, designed and implemented according to well-known learner-friendly solutions and flexible e-learning software application. When the learners visit the web site, they are subsequently called to fill in the e-questionnaires in order to automatically create their profile. This profile is stored in the database as soon as learners provide answers to the entire questionnaire. Subsequently, a registration form is available in order to declare learners login name and password for the site.

Validated learners are able to start studying e-courses within the portal, as soon as they complete the initial e-questionnaires. A need to rate existing e-courses and educational material according to the required level of ICT knowledge is, therefore, necessary. The selected e-courses do correspond to the available learner profiles. So, learners who receive an e-course are monitored throughout the system, regarding their selections, and thus their selection behavior is observed. In order to estimate whether the suggested e-course covers their level of knowledge and is adapted to the specific requirements of each learner, an evaluation procedure is required and therefore performed. For the evaluation of the e-courses selection, the behavior of the learner when attending the course is recorded, framed in the appropriate context. This behavior information is hold, providing the main source of transparent feedback for the learners' future e-courses selections.

3.3. Clustering-based learner profiling scheme

The proposed learner profiling scheme consists of two profiling approaches; both of them start on different grounds, but are merged and their results combined at a later phase. The dataset obtained from the e-questionnaires responses, is divided into two parts, one primary utilized by experts during the first step and another utilized by the automated clustering methodology. The first approach (already described in Section 3.1) is used at the initial stage of constructing the profiling ground truth and is characterized by a static profiling mechanism. The second approach exploits results provided by the previous one towards dynamic extraction of current and future learners' profiles and it is founded on the application of a data clustering technique (Mitchell, 1997), whose main goal is to identify homogeneous groups of objects based on the values of their attributes. When the clustering problem is applied to user modelling (Paliouras, Papatheodorou, Karkaletsis, & Spyropoulos, 2002; Smith, 2001), it gets more challenging, as input space dimensions become larger and feature scales are different from each other, as it is the case presented herein. In particular, a consideration of the original set of questions of the e-questionnaires as input space, results into a huge number of unique features to be taken into consideration when performing clustering on the learner answers.

To tackle such a large scale of features (Duda, Hart, & Stork, 2001; Mitchell, 1997), we use a hierarchical clustering algorithm, which does not demand the number of clusters as input a priori. In order to increase the robustness and reliability of our clustering step, the use of an unsupervised extension to hierarchical clustering in the means of feature selection is essential (Miyamoto, 1990). The results of the application of this clustering to only a portion of the system's dataset in question are then refined and extended to the whole dataset. The performance of the proposed methodology is finally compared to the previous static step, using the predefined profile characterizations as label information.

The general structure of the proposed system's clustering approach is summarized in the following steps:

1. Turn each input element into a singleton, i.e. into a cluster of a single element.
2. For each pair of clusters c_1, c_2 calculate their distance $d(c_1, c_2)$.
3. Merge the pair of clusters that have the smallest distance.
4. Continue at step 2, until the termination criterion is satisfied. The termination criterion most commonly used is thresholding of the distance value.

In our case, where the input space dimensions are large, the Euclidean distance is the best distance measure used (Yager, 2000). Still, this is not always the case, due to the nature of the individual features; consequently a selection of meaningful features needs to be performed, prior to calculating the distance $d(c_1, c_2)$ (Wallace & Kollias, 2003). Moreover, one feature might be more important than others, while all of the features are useful, each one to its own degree. In this work we tackle weighting of features based on the following principles:

(i) We expect elements of a given meaningful set to have random distances from one another according to most features, but we expect them to have small distances according to the features that relate them.
(ii) We select meaningful features based on the nature of the specific questions of the e-questionnaires. In particular, system experts perform an initial selection of meaningful questions, restricting the input space dimensions.
(iii) We further perform a second level filtering of the input data, based on the type of the input, leaving out answers – and thus questions – of arbitrary dimensions, such as free text input boxes of the e-questionnaires. Information collected from such answers fails out of the scope of clustering data and identifying user profiling information, being more useful for plain statistical approaches.

One of the key elements of the above algorithm is the ability to define a unique distance among any pair of clusters, given the input space and the clustering features. This clustering approach creates crisp clusters and does not allow for overlapping among the detected clusters. Thus, it forms the basic procedure, with the aid of which learners are automatically categorized to a distinct profile class that characterizes their behavior and their future interests and choices within the system. Additionally, each identified cluster is related to replies and specific comments of the questionnaires. We used these clustering results to classify the responses of learners to specific parts of the questionnaire, deriving information based on the learners' profiles as described in Section 4.2. The automatic extraction of clusters provided experts with the ability to distinguish the responses of the clustered learners with respect to more precise profiles. According to the cluster to which learners belong, educational content, appropriately selected by the system's experts, is offered to them.

4. SPERO: a case study

The methodology followed in this work and its implementation presented herein form a standalone, real-life framework, SPERO (SPERO), which has been developed for the extraction and adaptation of statistical information towards the automated generation of learner profiles, as well as educational content delivery to the learners. The ICT level of teachers and students is gathered through a visual, web-based, learner-friendly interface and appropriate e-courses are then provided to them according to their profiling information. Based on the adaptation of the IEEE fundamental e-learning model, the system provides faster learning at reduced costs, increased access to learning information and clear accountability for all participants in the learning process, thus indicates an efficient approach to the learning process via simple to use web interfaces.

The SPERO framework uses an electronic questionnaire collecting information about the ICT usage in the modern schools and providing reference material to all participants according to their profiles. The group of learners of the SPERO system consists of teachers whose have prepared a set of questionnaires. This set of questionnaires, contains questions about teachers' professional development, ICT background and use of ICT in educational activities, as analyzed in the following.

4.1. SPERO questionnaires

When teachers/learners access the SPERO web site, they are asked to select the language in which the questionnaires will be displayed. Having selected a language, the Teacher Questionnaire, divided into three different sections, is displayed (SPERO questionnaires). Questionnaire A, refers to the pedagogical utilization of information technologies asking for the teacher's opinion about the use of ICT in educational activities. Questionnaire B is further separated into three parts, *B1: Professional Development*, which investigates teacher's professional development status, *B2: Personal ICT Background*, which evaluates the teacher's familiarity with ICT and *B3: Teaching use of ICT*, in which the usage of ICT in teaching is evaluated. Finally, Questionnaire *C* provides a free text section, where teachers have the opportunity to express their ideas, complaints or comments.

The learner's profile is automatically extracted after completion of the teacher questionnaires (Section 3). Learners can then access the SPERO learning material related to their own profiles. Let us examine in the following some questions of the SPERO questionnaire in order to clarify the questions' weighting procedure.

4.2. SPERO learners' profiling

The e-questionnaire acts as a mediator towards information gathering; as the amount of answered questions increases, the new entries are summed up in the *LearnerReplies* structure. As already mentioned, the overall process results in a weighted mapping of the learner to the specified profiles. The final output of this process, is an extraction of a 1–1 learner-profile relation. In this way, learners are classified to a static profile that characterizes their behaviors, their interests and defines their further treatment by the system (SPERO profiles). This profile characterization forms the basis for the advanced profiling mechanism based on the already presented novel clustering procedure. Next, we shall examine the application and results of the static profiling mechanism, which is set at the first time that a learner accesses the SPERO site, as well as the advanced profiling mechanism based on the proposed novel clustering procedure.

4.2.1. Static profiling mechanism

In this section we explain the static profiling mechanism considering that part B1 of SPERO questionnaire is constituted of four questions. The scope of these questions is the detection of learner's knowledge about usage of computers and internet services. After completion of several questions of this part of the questionnaire, the system uses the learners' answers in order to count their knowledge level in internet services and provide the learner with a proper tutorial.

This questionnaire contains three groups of questions. Each possible answer to these questions is considered as a feature to be used by the clustering algorithm. The type of the first question ("Do you have computer at home") and the second question ("Do you have access to the Internet from your home?") is a radio button, having as possible answers yes or no. Similarly, the type of the third question ("How often do you personally use your Internet connection at home?") is also a radio button with possible answers: (a) daily, (b) weekly, (c) seldom, and (d) never. This question corresponds to feature 8 (Table 1) and the possible answers to it correspond to values: "daily" with a value of 100%, "weekly" with 60%, "seldom" with 30% and "never" with 0%. Assuming that the learner chooses answer "*(b) weekly*" and according to the rating, the learner is characterized as quite familiar with basic usages of Internet services.

The fourth question ("For which of the following did you use the computer at least once in the past month?") is a checkbox so the learner can give more that one answer. The possible answers for the fourth

Please cite this article in press as: Tzouveli, P. et al, An intelligent e-learning system based on learner profiling ..., *Computers & Education* (2007), doi:10.1016/j.compedu.2007.05.005

Table 1
Feature selection

Description	QuestionID	Feature
Do you have a computer at home?	133	6
Do you have access to the Internet from your home?	134	7
How often do you personally use your Internet connection at home?	139	8
For which of the following did you use the computer at least once in the past month?	141, 142, 143, 144, 145	9, 10, 11, 12, 13

question are: (a) Internet access, (b) Receiving/sending email, (c) Games, (d) Word processing, and (e) Educational software. Consider now that the learner has chosen answers: *(a) Internet access* and *(b) Receiving/sending e-mails*. Having chosen this combination of answers, the system considers that the learner is familiar with basic Internet services. Consequently, the learner is considered to have the basic level on Internet services.

The learner continues with the rest of questions which this part contains. When completing this part of questionnaire, the system automatically estimates the learner's level in ICT. Then, the system categorizes the learner's level in Internet services, say, higher than basic; it then provides the learner with the Internet services tutorial for advanced learners, stored in the learning resources.

4.2.2. Advanced profiling mechanism

In this section, we examine the implementation of the proposed clustering algorithm using SPERO system's data set and the Euclidean distance measure. The clustering algorithm has been applied to a small portion of the data set, namely a 10% of the overall system's learners; it contained 100 elements (learners), characterized by 44 meaningful features and answers. These features correspond to a set of questions which have been considered appropriate for the profiling extraction process by a group of experts.

Let us consider that the advanced profiling mechanism takes as input the four aforementioned questions. The first column of the Table 1 presents the description of this question while the number of associated question ids and features used in the clustering procedure are shown in the second and third columns of Table 1, respectively.

The above elements belonged to three static profile classes, but this labeling information was not used during clustering; the labels were used, though, for the evaluation of the quality of the clustering procedure, as described in Mylonas, Wallace, and Kollias (2004), prior to projecting the results to the whole data set. More specifically, each detected cluster was then assigned to the experts' provided class that dominated it. In the general case, identified clusters define specific interests and profiles, which do not necessarily correspond to the a priori known classes that are utilized during the first phase. These clusters are useful in producing collaborative recommendations of the e-learning content to the learners at a later stage, as described in Wallace and Kollias (2003). The results are shown in Table 2, whereas the numbers inside parenthesis separated by commas denote the elements belonging to its one of the three profile classes in each step.

Performing the initial clustering on a mere 10% subset is not only more efficient computation-wise, it is also better in the means of quality and performance, when compared to the approach of applying the hierarchical process to the whole data set. Although clustering over this 10% of the data set resulted in different possible

Table 2
100 learners clustering results – 9 clusters

Clusters	Elements	%
1st	(1, 1, **4**)	(16.66%, 16.66%, **66.66%**)
2nd	(0, 1, **6**)	(0.00%, 14.28%, **85.71%**)
3rd	(4, 2, **5**)	(36.36%, 18.18%, **45.45%**)
4th	(3, 3, **6**)	(25.00%, 25.00%, **50.00%**)
5th	(4, 2, **5**)	(36.36%, 18.18%, **45.45%**)
6th	(**8**, 4, 5)	(**47.05%**, 23.52%, 29.41%)
7th	(**4**, 1, 4)	(**44.44%**, 11.11%, **44.44%**)
8th	(3, **10**, 6)	(15.78%, **52.63%**, 31.57%)
9th	(1, **4**, 3)	(12.50%, **50.00%**, 37.50%)

identifiable clusters, optimal results have been obtained for a number of nine clusters. Table 2 presents the clustering results over a dataset of 100 learners. The hierarchical clustering algorithm terminated by the time it reached a threshold of 9 clusters. The first cluster consists of 6 learners, 4 *Advanced, 1 Beginner* and 1 *Experts*. The corresponding percentage distribution clearly indicates that the 4 *Advanced* learners dominate the first cluster. The same applies to the second cluster with 7 learners as well, where 1 learner belongs to *Beginners* and 6 learners are in the intermediate state, i.e. *Advanced*. Third, fourth and fifth clusters are all resulting in the supremacy of *Advanced*. Moreover, cluster 7 acts as an intermediary between *Advanced* and *Experts*, as it illustrates a draw in the elements between those two profile classes whereas cluster 6 forms a solid representative of the *Experts*. Clusters 8 and 9 are dominated by the *Beginner* profile class. Consequently, 5 out of 9 clusters (55.55%) are indicating a clear advantage of the *Advanced*.

The clustering step results demonstrate the clear trend underlying in the system's input data: learners are characterized by intermediate ICT skills and expertise. This observation is extremely evident in the third column of Table 2, which indicates clearly that most learners of the system belong to the static, intermediate *Advanced* profile. The first two clusters identified by our algorithm are unambiguously dominated by the third profile class, i.e. *Advanced*. Additionally, clusters 3, 4, and 5 indicate a clear majority of the same third class in their elements as well.

4.2.3. System's evaluation and usage

The overall system developed in this approach was tested and evaluated by several hundreds of learners, spread all over Europe. After completing the questionnaire, learners access a list of e-courses that correspond to their single final profile. These learners are considered as certified learners of the SPERO system. The e-courses are presented in the language that learners have selected for displaying the questionnaire. The list of e-courses either points at links to official ICT sites or to e-courses that have been developed by the experts of the system and are hosted in the server of the proposed case study. The expert group have selected or created such proper learning material for each of the three static profiles.

The first category of e-courses which is corresponding to the first static profiles is for beginners. This category contains e-lessons with the following subjects: introduction to information technologies, usage of operating system, usage of text processor and work sheets, usage of internet. The second category of e-courses is for advanced learners and contains topics as usage of text processors and work sheets, usage of presentation programs, usage of databases, usage of operating systems, basic ideas of the programming languages (structure programming), and usage of programs for image processing. Finally, the last category classified experts and contains e-lessons such: scripting programming, basic structures, commands and applications, object-oriented programming languages, html tutorials, xml tutorials, and usage of programs for video processing.

5. Statistical interpretation/analysis

An important outcome of the SPERO study case is a statistical analysis of the learners' responses to the e-questionnaires; this provides useful information about teacher's status, level of expertise and usage of information, and communication technologies. Statistical analysis results are automatically created using the SPERO database and presented in the *Statistics* session of the main menu of the SPERO site. Two choices are available: (a) Statistics based on Learner's Profile and (b) Numerical Statistics. Choosing *Numerical Statistics*, a web page with four main choices is loaded. The choices are: SPERO participation statistics, School/Questionnaire statistics, Teacher Questionnaire statistics, and EU version statistics.

The choice Numerical Statistics classifies the teachers' responses to the e-questionnaires per country. The statistics, based on a dynamically updated data collection, are derived from responses coming from 1026 teachers throughout the countries which are participating in the SPERO initiative: Greece (450), UK (126), Denmark (114), Spain (183), Iceland(165). In this section, the results of the statistical analysis of the teachers' responses to the numerical fields of parts – B1: Professional Development, B2: Personal ICT Background, and B3: Teaching use of ICT, is discussed. A full analysis of the SPERO questionnaire is available in the SPERO statistics entity (SPERO e-survey).

The B1 part contains information about professional development of teachers participating in the SPERO electronic survey. In general, teachers hold a permanent full-time position at their school, except of Spanish teachers, a significant part of who (34.65%) hold a temporary full-time position. The majority of Greek teachers in special education are in segregated units with the rest working in a special unit of mainstream schools and very few providing learning support in mainstream schools. In contrary, the vast majority of the Icelandic teachers provide learning support. In Denmark there is a balance among the three categories. In UK and Spain two-thirds of the teachers belongs to the last category (Fig. 2).

The next question collects information about teachers' qualifications in Special Education. It is important to mention that 9 out of 10 Danish teachers have graduated from a teacher training college, while 20% of Icelandic teachers have a postgraduate degree in the field of Special Education Needs (SEN) training. Most of the UK teachers have a Diploma or equivalent in support for training in SEN (Fig. 3). In most countries, the majority of teachers have over 10 years of experience (in Iceland teaching experience is more balanced). About two-thirds are between 30 and 50 years old. Almost half of the participating teachers in Denmark are older than 50 years, which justifies some of their concerns with respect to ICT usage in learning (Fig. 4).

The main scope of the B1 part of the questionnaires is to extract information about the ICT background of teachers. The statistical analysis indicates that almost all teachers have computer and Internet access at home; Greek followed by Spanish teachers have the lowest access capabilities (Fig. 5). In the question about the teachers' usage of Internet at home, Greek and Spanish teachers have the lowest level of daily usage (32% for Greek teachers and 17% for Spanish teachers – Fig. 6).

The responses to the question about the usage of computer at home during the past month, the majority of teachers did, according to their replies, the following: word processing, internet access, receiving or sending an e-mail, with a significant portion looking for educational software (Fig. 7). Almost half of teachers replied that they have performed the following activities at least once without help: installation of software, installation of a printer, and creation of backup (Fig. 8).

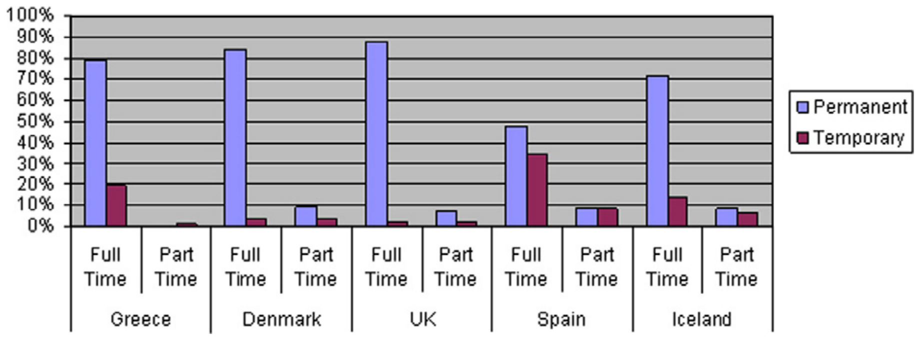

Fig. 2. Professional post.

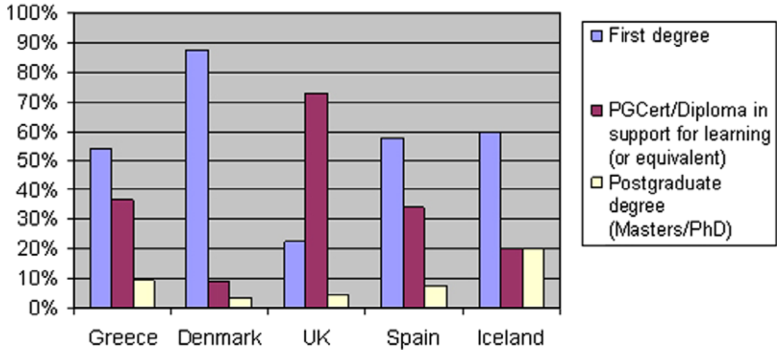

Fig. 3. Teachers' qualification/training in Special Education Needs.

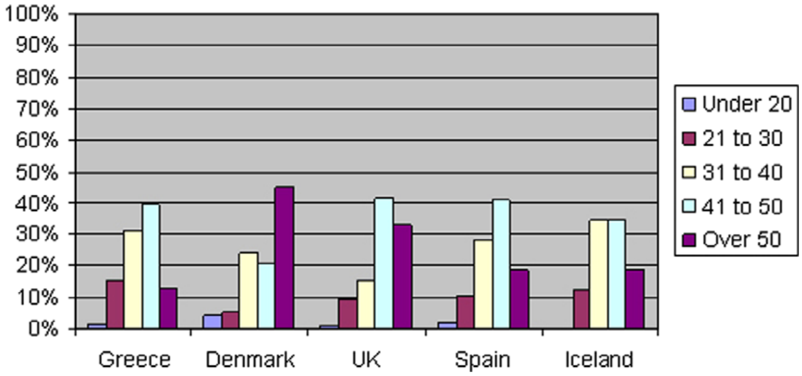

Fig. 4. Teachers' age.

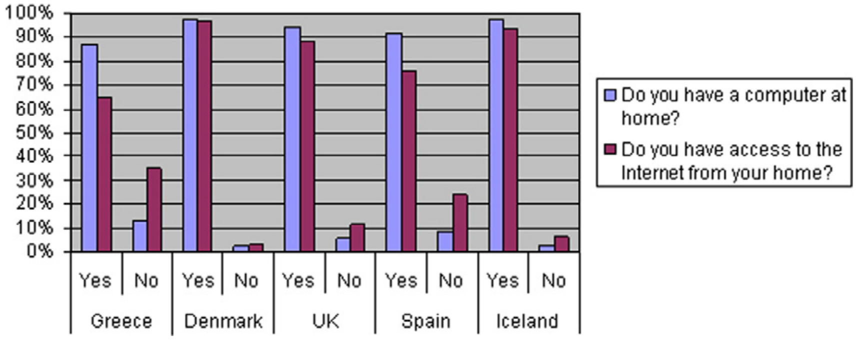

Fig. 5. Computers and Internet at home.

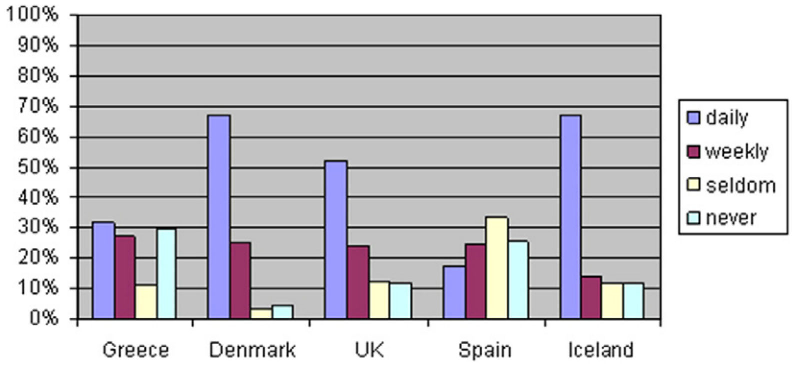

Fig. 6. Usage of Internet connection at home.

The main scope of the B3 part is to find out the usage of ICT in teaching activities. In all cases there exists at least one computer in each school (Fig. 9). However, in many cases computers are not used by teachers, as the responses to the next question indicate (Fig. 10).

In general, Danish and British teachers are using a computer at school more than teachers from the other countries while half of Greek teachers do not use a computer at all. The replies to the next question indicate that the majority of teachers have access to Internet or educational software at their work. At the same time, in Greece, there still exists a problem with special education schools (Fig. 11).

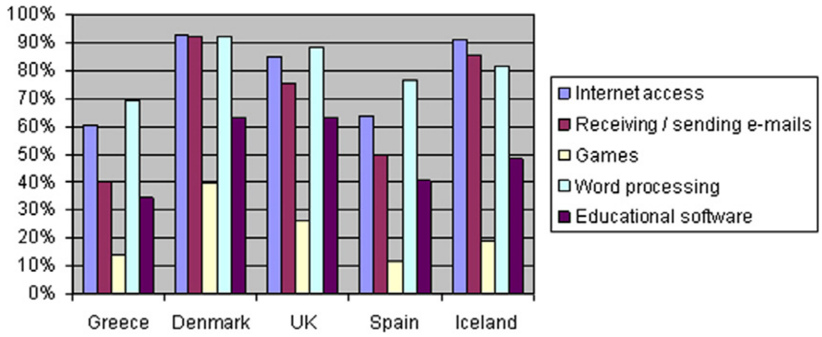

Fig. 7. Usage of computer the past month.

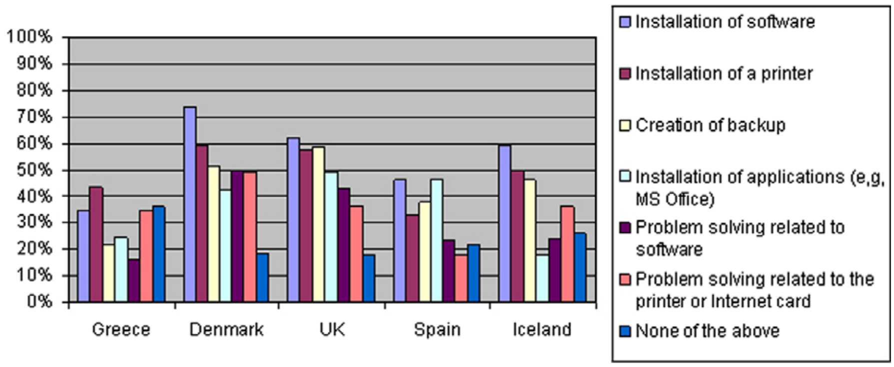

Fig. 8. Facing of software and hardware problems.

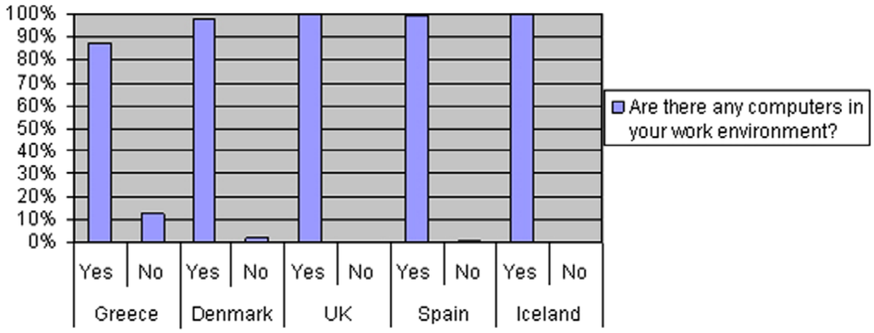

Fig. 9. Computers existence in work environment.

Almost half of the teachers reply that they are using Internet for search and retrieval of information related to the needs and problems faced by SEN students (Fig. 12).

Above ratings vary a little when referring to special education schools, but more attention needs to be paid to improving this capability. Nine out of ten Icelandic and about 8 out of 10 Danish and British teachers reply that they are using Internet at their school in order to find additional resources of educational material. In the same question less than half of Greek teachers' replies and about 60% of Spanish teachers' replies are positive (Fig. 13).

The replies to where students access Internet for learning purposes present large variations among countries. The computer lab is the leading place. It is worth mentioning the usage of classroom for Danish and

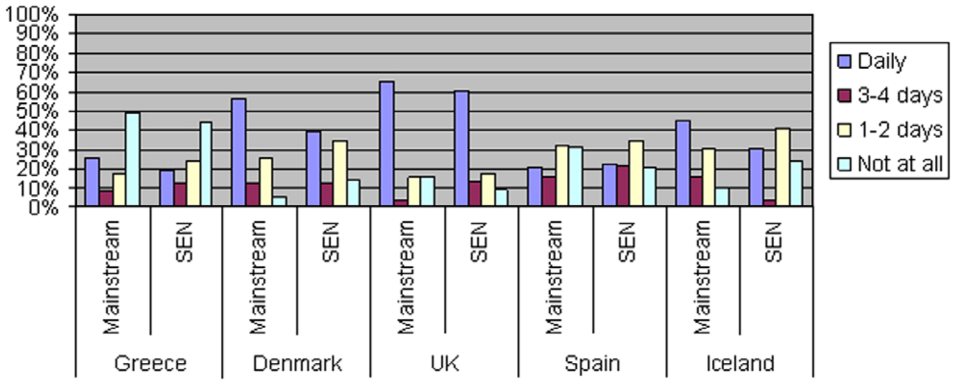

Fig. 10. Usage of computers in classroom.

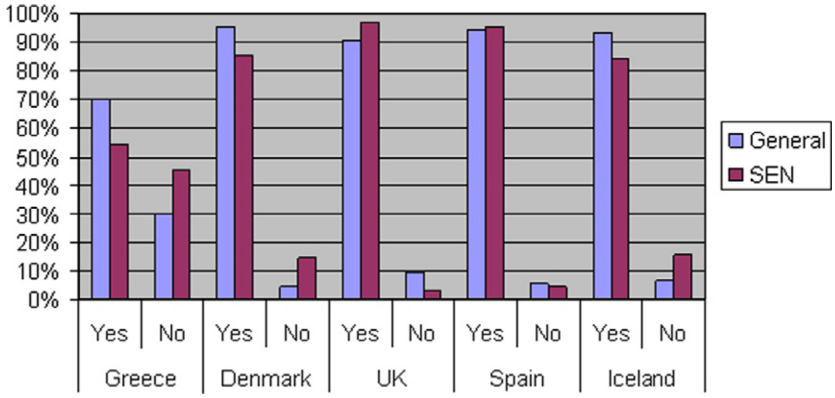

Fig. 11. Internet access and educational software in the classroom.

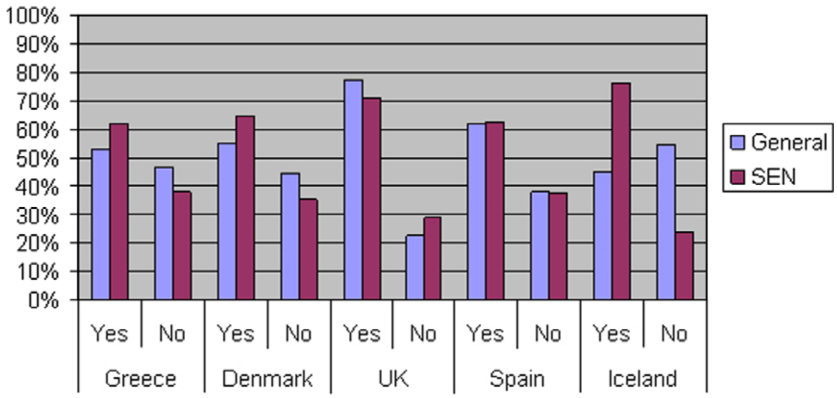

Fig. 12. Searching Internet for SEN issues.

UK students at more than one-third, and the no access of half of students in Greece (Fig. 14). Referring to the technical quality of educational software available in their classrooms, the majority of teachers believe that it is adequate (Fig. 15). The exception here is Greek SEN schools, where more than half of teachers do not think so. However, the majority of teachers in all countries think at availability of educational software in their classrooms is useful.

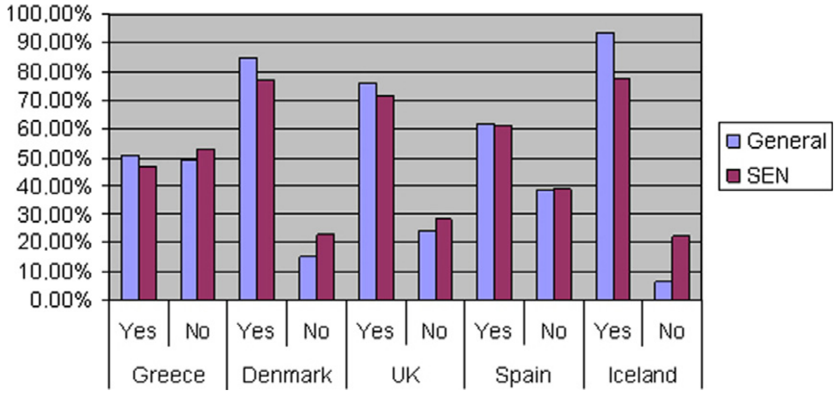

Fig. 13. Additional sources of educational material.

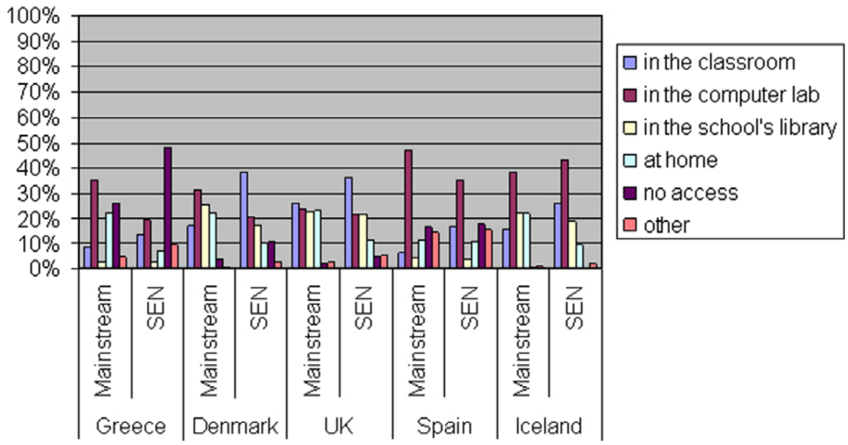

Fig. 14. Students access the Internet for learning purposes.

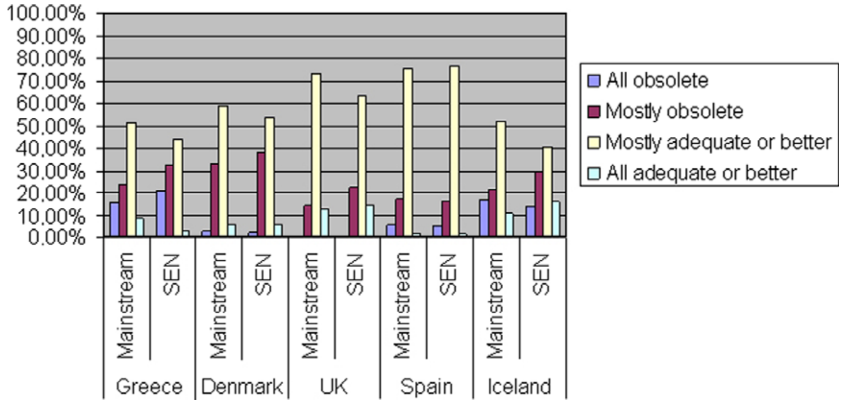

Fig. 15. Technical quality of educational software.

6. Conclusions and future work

In the proposed paper, we introduce a novel integrated e-learning application, being able to continuously adapt its services to its learners' preferences, usage history, and specific needs as they evolve during its usage.

The proposed e-learning system handles learners in a personalized manner, by building different learners' profiles according to their ICT knowledge and expertise. The usage of learners' responses to the e-questionnaires leads additionally to statistics analysis. A future meaningful extension of the proposed system is a connection between semantic and statistical information. Furthermore, utilization of a fuzzy relational knowledge representation model in the weight estimation process could also provide a new direction to the proposed method.

References

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction, 11*, 87–100.

Dagger, D., Wade, V., & Conlan, O. (2003). Towards "anytime, anywhere" learning: The role and realization of dynamic terminal personalization in adaptive e-learning. In *Proceedings of the world conference on educational multimedia, hypermedia and telecommunications*.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). Wiley.

Henze, N., & Nejdl, W. (2001). Adaptation in open corpus hypermedia [IJAIED Special Issue on Adaptive and Intelligent Web-Based Systems].

Karpouzis, K., Caridakis, G., Fotinea, S., & Efthimiou, E. (2005). Educational resources and implementation of a Greek sign language synthesis architecture [Special Issue in Web3D Technologies in Learning]. *Computers & Education, 1*, 1.

Mitchell, M. T. (1997). *Machine learning*. McGraw-Hill.

Miyamoto, S. (1990). *Fuzzy sets in information retrieval and cluster analysis*. Kluwer Academic Publishers.

Mylonas, Ph., Wallace, M., & Kollias, S. (2004). Using k-nearest neighbor and feature selection as an improvement to hierarchical clustering. In *Methods and applications of artificial intelligence*. *Lecture notes in computer science* (Vol. 3025). Springer.

Paliouras, G., Papatheodorou, C., Karkaletsis, V., & Spyropoulos, C. D. (2002). Discovering user communities on the internet using unsupervised machine learning techniques. *Interacting with Computers, 14*(6), 761–791.

P1484.1 Architecture and Reference Model, IEEE LTSC. <http://ieeeltsc.org>.

Rosenberg, M. J. (2002). *E-learning: Strategies for delivering knowledge in the digital age*. McGraw-Hill Inc., ISBN 007137809X.

Smith, A. S. (2001). Application of machine learning algorithms in adaptive web-based information systems. CS-00-01, School of Computing Science Technical Report Series. ISSN 1462-0871.

SPERO: Teleinformatics System for Continuous Collection, Processing, Diffusion of Material for Teacher Training in Special Education, Leonardo Da Vinci Programme. <http://www.image.ntua.gr/spero>.

SPERO e-survey. <http://www.image.ntua.gr/spero/statlist.php>.

SPERO profiles. <http://www.image.ntua.gr/spero/statsspero/profiles/>.

SPERO questionnaires. <http://www.image.ntua.gr/spero/questspero>.

Tzouveli, P., & Kollias, S. (2004). User adaptation in e-learning environments. In *Proceedings of IFIP world computer congress, technology enhanced learning*.

Wallace, M., & Kollias, S. (2003). Soft attribute selection for hierarchical clustering in high dimensions. In *Proceedings of the international fuzzy systems association world congress (IFSA)*.

Weber, G., Kuhl, H. -C., & Weibelzahl, S. (2001). Developing adaptive internet based courses with the authoring system. In *Proceedings of third workshop on adaptive hypermedia*.

Yager, R. R. (2000). Intelligent control of the hierarchical agglomerative clustering process. *IEEE Transactions on Systems, Man and Cybernetics, Part B, 30*(6), 835–845.

Please cite this article in press as: Tzouveli, P. et al, An intelligent e-learning system based on learner profiling ..., *Computers & Education* (2007), doi:10.1016/j.compedu.2007.05.005