

An integrated approach towards intelligent educational content adaptation

Phivos Mylonas, Paraskevi Tzouveli and Stefanos Kollias
National Technical University of Athens
School of Electrical and Computer Engineering
Department of Computer Science
Image, Video and Multimedia Laboratory,
Zographou Campus, Athens, Greece, 157 73

Abstract. One of the major shortcomings of modern e-learning schemes is the fact that they significantly lack on user personalization and educational content representation issues. Semi- or fully automated extraction of user profiles based on users' usage history records forms a challenging problem, especially when used under the e-learning perspective. In this chapter we present the design and implementation of such a user profile-based framework, where educational content is matched against its environmental context, in order to be adapted to the end users' needs and qualifications. Our effort applies clustering techniques on an integrated e-learning system to provide efficient user profile extraction and results are promising.

Keywords. Online education, Personalized education, E-learning, Clustering-based user profiling

Introduction

In the current Internet-based world, new trends concerning online education and e-learning evolve. Users are constantly confronted with a series of technological improvements and developments; however, most of the proposed approaches do not tackle sufficiently the raised user personalization and educational content representation issues. Currently, traditional teaching techniques are finding themselves under revision and re-evaluation and new or sometimes radical ones do come into play. Above all, the Internet plays a significant role in all fields of education, contributing the most to the educational procedures [1]. As a result, Internet-oriented applications arise in the aid of educational needs, trying to close the gap between traditional educational techniques and technology-oriented education.

One of the most innovative ways to empower such a workforce with the skills and knowledge it needs is the utilization of e-learning; e-learning is currently propagating at quick rates and its impact on teaching and learning raises many questions with no clear answers. The impact of Information and Communications Technology (ICT) is a task that has become more and more apparent in learning and teaching at all levels of education [2]. Towards that goal, during the last years, e-learning systems were developed in terms of rather static software applications, lacking on educational

multimedia environments and personalized capabilities and without any interest given to the real users input and feedback [3], [4] and [5].

In the framework of the presented Leonardo SPERO project [6], we introduce a novel method for gathering information and estimating the ICT level of learners in all fields of education. This is achieved through a web-based interface that takes into account personalized, profile-based schemes. It has been designed to enable learners to gracefully increase their ICT knowledge and provide them with credible information and feedback, such as suitably selected e-courses and multimedia educational content.

The structure of this chapter is as follows: in section 1, the overall architecture design of the SPERO system is introduced, including its basic corresponding groups and components as well as its e-questionnaires. Afterwards, a short reference to the IEEE e-learning model is presented, current approach's adaptation is analyzed and the additional features provided by it are explained. In section 2, we begin by tackling the problem of the learner profile creation, followed by issues concerning the initial static profile extraction procedures. All of the above are used as the main feedback source for the forthcoming intelligent clustering profiling procedure, which is presented in section 3. In the same section a description of the utilized clustering algorithm is provided, together with experimental results on the clustering-based profiling scheme. Next section 4 describes the general context for this work, dealing with the educational content offering of the system and briefly presenting its categorized e-courses and operational examples of use. Finally, in section 5, we present our concluding remarks and some ideas for potential future efforts on the subject.

1. Framework's overview

The first step to consider towards the establishment of an efficient, integrated e-learning framework is the definition of its basic architecture, which in our case is shown in Figure 1:

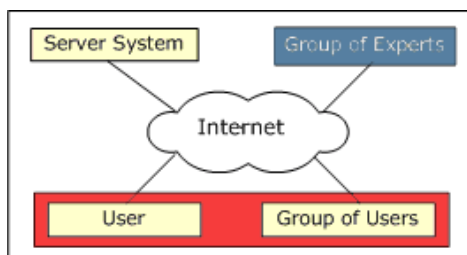


Figure 1. SPERO System Architecture

Three main networked components can be identified which are:

1. the group of the *system's users*
2. the group of *system's experts*, who play a key role in the initialization of the personalization process; this group includes teachers, experts in e-learning, data analysts, psychologists and software engineers
3. the actual *server system*, which includes all hardware and software needed to establish a 2-tier system core [7].

The first group includes every kind of teacher working either for general education or for the Special Education sector. Eventually, expanding the system's architecture, a user could be identified as any learner, whether a student, teacher or employee. The second group's role is crucial in the personalization process of the SPERO system, since it defines the initial set of specifications and limitations of the end-users' profiles, which justifies the variety of people comprising it. The third group includes all hardware and software that enables a web-server to be active, as well as efficient and robust. All distinct SPERO web applications together with an underlying Relational Database Management System to support profiling and user information are included in this configuration.

Moreover, one of the basic interaction components of SPERO and the mean of communication with its end-users is formed by the so called *e-questionnaire*. SPERO experts have designed and illustrated two groups of e-questionnaires: The first group contains questions about school units in order to collect general details about them. The second group contains questions about teachers' ICT background. The questions, which are addressed to the teachers, are intended to collect information about teachers' educational background, as well as their background in ICT. In addition, information concerning teachers' opinions about pedagogical utilization of ICT and the amount of using ICT in teaching procedure is also extracted. As already mentioned, learners could either be teachers or students, however both of them are in great need of ICT: on the one hand, teachers mainly because their role is continuously evolving and demanding new formation and on the other hand students, because of their need to have distance e-courses in the field of ICT. Each questionnaire is divided into several subsections, a portion of which is depicted in Figure 2. The presented sub-questionnaire collects information about general teachers' educational background, as well as their background in ICT.

SPERO
Teacher Questionnaire

A. Pedagogical Utilisation of Information Technologies

Please complete the following:

A.1. Changes in the daily activities of the teaching staff due to the use of ICT.

A.1.1. The widespread use of ICT as a teaching tool in school will lead to a change in some aspects of daily teaching practice. Please, indicate which:

	Mostly	Rarely
A change of attitude of the teaching staff will be necessary	<input type="radio"/>	<input type="radio"/>
New learning activities with the students will have be designed and implemented	<input type="radio"/>	<input type="radio"/>
It will not affect other learning activities	<input type="radio"/>	<input type="radio"/>

Figure 2. Part of Teachers' Questionnaire

SPERO e-questionnaires are developed in the framework of conducting a European survey. Consequently, they are translated in eight European languages and these translations are stored within SPERO's Relational Database Management System. Software has been developed for automatic presentation of e-questionnaires in every one of these eight languages. Moreover, the e-questionnaire is used to estimate the ICT level of individual users, by using the calculated user's profile categorizations that are automatically extracted in the following by the SPERO software. More than one

learning resources (e-courses and educational material) are selected by experts, up to one for each of the distinct collaborative user profiles categories. The set of e-questionnaires is used for ICT level estimation in the framework of the distance-learning architecture that is presented in Figure 3:

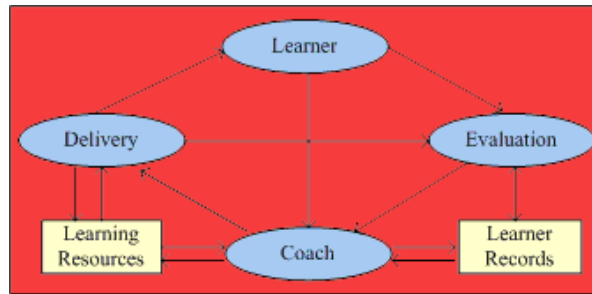


Figure 3. IEEE learning system entities

In this work we attempt to extract learner profiles through the evaluation entity of the above architecture, proposed by the IEEE Reference Model (WG1 LTSA) of the Learning Technology Standards Committee [8], [9]. This standard covers a wide range of systems, such as learning technology, education and training technology, computer-based training, computer assisted instruction, intelligent tutoring, and is pedagogically neutral, content-neutral, culturally neutral, and platform-neutral. However, in this generic approach to e-learning systems, a system's ability to adapt its operation to the user is not defined, although an evaluation process exists. Aiming at extracting learner profiles through this entity, we are proposing the replacement of the IEEE standard *Evaluation* entity, with the novel *Re-evaluation* entity, which, additionally, is strongly related to two new entities: the *E-survey* entity for gathering statistical information and the *Profile Database* entity dealing with all learners' profiles. A schematic diagram of the proposed replacement is presented in Figure 4, the replaced learning components are shown in Figure 5, whereas a detailed description of the above concepts can be found in [10].

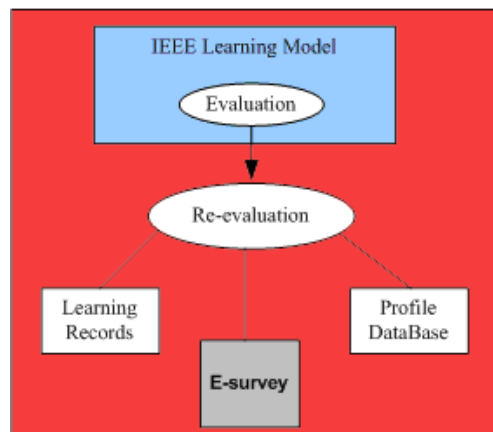


Figure 4. Proposed replacement of *Evaluation* entity

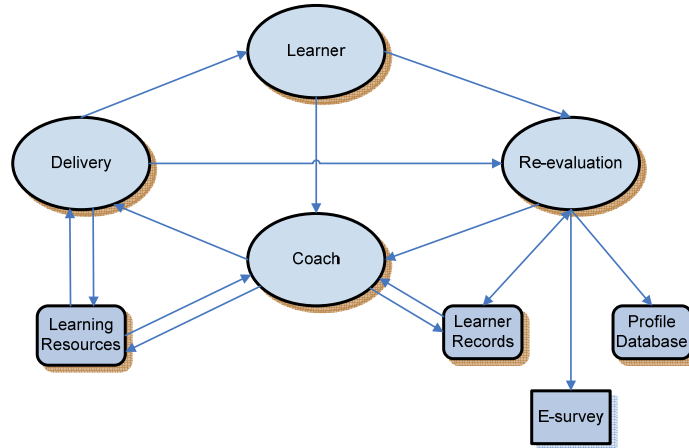


Figure 5. Proposed e-learning system after replacement of *Evaluation* entity

In order to assist the profiling process, the need for this re-evaluation step is essential; the usage of an appropriate e-questionnaire is considered necessary in order to collect user input data and build a large, reliable ground truth, basing profiling information on top of them. For this purpose, experts designed and illustrated an e-questionnaire which collects information about learners' ICT background, learners' opinions about pedagogical utilization of ICT and the amount of using ICT in teaching procedure. Additionally, software has been developed to allow e-surveys to be conducted based on users' answers.

As a result, the core of the system relies on this replaced *Evaluation* entity. The latter forms an independent personalization subsystem, where user profiling information is extracted, according to statistics gathered from the e-questionnaire database and the e-survey. Delivery of educational content is then possible, based on the results of the profiling procedure, providing personalized views to the system's end-users and taking into consideration their particular ICT levels of education and needs.

2. Initialization of user profiles

At this point, a brief presentation of the system's personalization subsystem is essential. It is the intended nature of SPERO that dictates utilization of two profiling approaches; both of them look very different in the beginning, but they are combined at a later phase. One approach is followed at the initial stage of constructing the profiling ground truth and is characterized by a static profiling mechanism, whereas the second approach exploits results provided by the previous one towards dynamic extraction of current and future SPERO users. First step is materialized through a fixed mapping of obtained user input to specific types of user profiles explained in the following within this section. Second step is acting on top of the first one and is founded on application of a clustering technique [11]. The latter is presented in the next section. Accordingly, the corresponding dataset was divided into two parts, one primary utilized by experts during the first step and another utilized by the automated clustering methodology.

Let us illustrate the functionality of the initial profile representation, which is based on the static profiling mechanism. Experts, based on experience and intuition, define a set of three user characterizations, forming a static profiling representation. This step forms a pre-processing task performed by hand, combined with statistical analysis and later cross-correlation selection of features, which indicates the most important ones that were considered in the following. It is considered to be extremely reliable and accurate, as it was the result of three years of intensive collaboration among experts from 8 European countries that examined this through questionnaires in national and cross-national analysis. These initial characterizations are also utilized at a later stage, during the intelligent profiling process, providing a rock-solid point of reference and although they are thought to be static, they are actually generated automatically from the system. Their basis is information provided by the users' input data, obtained from the e-questionnaires. In this case, personalization was needed in order to aid with the ultimate educational content offered by the system; this was successful, based on the electronic mining of knowledge gathered from the system's questionnaires subsystem.

The profiling mechanism creates updates and uses system's user profiles, matching specific e-questionnaires question triggers, to particular identified patterns. The profile model's design facilitates both the process of using user preferences in profile creation, as well as the process of preference tracking throughout the whole profiling procedure. Furthermore, it is designed in a way that allows for the automated extraction of user profiles, based on these preferences and the users' input history. This model forms an initial static version of the user profile denoted by *UserProfile* (Figure 6).

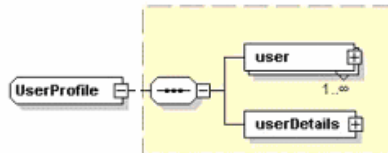


Figure 6. Structure of the *UserProfile*

As seen in the figure, the main abstract structure of *UserProfile* compound type contains two elements. The first one (i.e. *user*) stores information about the user's history, while the second (i.e. *userDetails*) stores the user preferences. As the initial profiling process instantiates, all user profiles are stored within a single, central mapping structure, whose abstract model is presented in Figure 7. The *UserProfile* is mapped against information retrieved either from the e-questionnaire itself, or directly from the input of the users. The first element, *QuestionId* holds all the information required for identifying the underlying e-questionnaire question, as well as its type, aiming at better understanding and fitting of the currently generated profile. The second element, *UserInput* contains user data related information, such as the user's answers. Both, the sequences of *QuestionId* and *UserInput* elements denote the existence of large amount of different system's e-questionnaires, questions and users' input data.

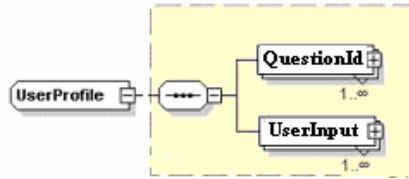


Figure 7. Mapping structure of *UserProfiles*

The core of this methodology is summarized in the following step of weights association, performed according to the following guidelines: Once a user answers a question of the input e-questionnaire, a relevance degree is associated to it and adjusted to her/his specific QuestionId element, and thus also propagated to the UserProfile element. As more and more answers from the end user enter the UserProfile structure, additional relevance degrees are registered to the corresponding QuestionID elements. Depending on the particular question, as well as the part of the e-questionnaire that this question belongs, different degrees are propagated. The latter is based on comparison of the provided numerical values with the range of values a-priori associated with the profiles. In order to better understand the underlying mapping structures, examples are presented in Figure 8, Figure 9 and Figure 10, derived from the group of experts directly from the e-questionnaires:

B.1.4.Teaching Experience	
Under 5 years	Beginner
5 to 10 years	Advanced
11 to 20 years	Advanced / Expert
Over 20 years	Expert

Figure 8. Static profile mapping example (1st part of e-questionnaire)

B.2.5.Which of the following tasks have you performed at least once ?	
Installation of software	a
Installation of a printer	b
Creation of backup	b
Installation of applications (e.g. MS Office)	b
Problem solving related to software	e
Problem solving related to the printer or Internet card	f
None of the above	g
Other, please indicate:	text

Checked (a...f)>4 => "Expert"
2<Checked (a...f) <4 => "Advanced"
2 < Checked (a...f) OR Checked (g) => "Beginner"

Figure 9. Static profile mapping example (2nd part of e-questionnaire)

B.3.4.How many hours a week, on average, do you use the Internet or educational software with your students?	
General	if (General or/ and SEN teaching hours)>5 "=> Expert"
SEN	else if 2 < t. h. <5=>"Advanced" else if (t. h.)<2=>"Beginner"

Figure 10. Static profile mapping example (3rd part of e-questionnaire)

The e-questionnaire acts as an intermediate towards the information gathering process, and as the amount of the answered questions increases, the more entries are summed up in the "UserProfile" structure. Thus, the overall process results in an aggregated weighted mapping of the end user to the specified profile, which is different for each user's answers and depends on their particular nature. This mapping is temporarily preserved and as the completion of the e-questionnaire is carried out, the

above mentioned weighted mappings are aggregated. In that manner, they continuously and dynamically change every user's profiling, until a final equilibrium profile state is achieved. Test bed experimental results within the SPERO project indicate that after answering an approximately 50% of questions, it is possible for the system to balance to a solid, static, initial user profile with great confidence. As a result, the entire pre-processing task is governed by great accuracy and reliability, although in general it is difficult to handle in such cases, where data is characterized by numerous measurable features like answers to e-questionnaires. This is the case when multiple independent features characterize data, and thus more than one meaningful similarity or dissimilarity measures can be defined. A common approach to the problem is the lowering of input dimensions [11], which may be accomplished by ignoring some of the available features/answers, and is the one followed herein. Additionally, a statistical cross-correlation analysis of the importance of features was also performed. For the sake of space we omit detailed presentation of the feature selection process, however the basic principle to be followed throughout this chapter is that while we expect elements of a given meaningful set of e-questionnaires to have random distances from one another according to most features, we expect them to have small distances according to the features that relate them. We rely on this difference in distribution of distance values in order to identify the context of a set of elements, i.e. the subspace in which the set is best defined and provides the most meaningful results in terms of semantic clarity. The result of this process is a set of 44 meaningful e-questionnaire questions.

The final output of this process, following the application of the weights, is the extraction of a "1-1" profile-end user relation. In that manner, each end user is classified to an initial, static profile that characterizes his behaviour, his interests and his further treatment from the system. This particular profile characterization forms the basis of the following intelligent clustering procedure, which includes the notion of profile extraction and integration within this system. In Figure 11 we present an indicative sample of the end-users' static profiling, extracted by previously analyzed procedure within the SPERO system, according to the users' answers collected by the e-questionnaires.

	User ID	Professional Development	Personal ICT Background	Teaching use of ICT
1.	509	beginner	unspecified	beginner
2.	659	beginner	unspecified	beginner
3.	807	expert	expert	beginner
4.	808	unspecified	beginner	beginner
5.	809	expert	expert	advanced
6.	811	expert	expert	advanced
7.	813	expert	expert	advanced
8.	814	expert	expert	advanced

Figure 11. Initial static profiling mapping

3. Clustering-based profiling

At this point, experts are considered to have successfully labelled the first part of the dataset, provided initially as the ground truth of the profiling mechanism, by assigning profiles to each user according to the responses given to e-questionnaires. The next step to follow consists of applying intelligent clustering techniques to group

similar profiles in the current as well as second part of the dataset. Traditional classification schemes are not applicable in this case and an intelligent clustering methodology is favoured. This step is necessary for the unsupervised operation of SPERO, where no manual labelling is available. Of course, at this phase various predefined profile schemes may arise and results are not bound to the so far manual characterization of system's experts. In this manner, more than three profiles may be obtained and categorization to the predominant ones needs to be applied. The latter is necessary for us, in order to be able to compare results between the two phases and derive meaningful conclusions. After a small theoretical analysis of the proposed clustering technique, we present in detail the steps of the clustering algorithm for the problem at hand.

The core of the clustering data concept is to identify homogeneous groups of objects based on the values of their attributes. It is in general a difficult problem to tackle and is unquestionably related to various scientific and applied fields, especially when clustering is applied to user modelling [12], [13]. The problem gets more and more challenging, as input space dimensions become larger and feature scales are different from each other, as is the case in our system. In particular, a consideration of the original set of questions of the e-questionnaires as input space, results into a large number of 176 unique features to be taken into consideration when performing clustering on the user answers. The best way to go in this direction is to use a hierarchical clustering algorithm, which is able to tackle such a large scale of features [11], [14]. Although such a method does not demand the number of clusters as input, still it does not provide a satisfactory framework for extracting meaningful results. This is mainly due to the "*curse of dimensionality*" that dominates such an approach, as well as the inevitable initial error propagation and complexity along with data set size issues.

In order to increase the robustness and reliability of the whole clustering step of our system, the use of an unsupervised extension to hierarchical clustering in the means of feature selection was evident [11]. Using the results of the application of this clustering to 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 step of fixed clustering, using the predefined profile characterizations as a priori label information.

The general structure of such hierarchical clustering algorithms, which forms the structure of SPERO's clustering approach as well, is summarized in the following steps and presented analytically in [11]:

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 value of the distance.

It is worth noticing, though, that in our case, where the input space dimensions are large, the Euclidean distance is thought to be the best distance measure used [15]. 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 [16]. 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:

- a) 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,
- b) 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 and
- c) 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 falls out of the scope of clustering data and identifying user profiling information, being more useful for plain statistical approaches.

More formally, let c_1 and c_2 be two clusters of elements. Let also r_i , $i \in N_F$ be the metric that compares the i -th feature, and F the overall count of features (the dimension of the input space). A distance measure between the two clusters, when considering just the i -th feature, is given by:

$$f_i(c_1, c_2) = \sqrt[\kappa]{\frac{\sum_{a \in c_1, b \in c_2} r_i(a, b)^\kappa}{|c_1||c_2|}} \quad (1)$$

where e_i is the i -th feature of element e , $|c|$ is the cardinality of cluster c and κ is a constant. The overall distance between c_1 and c_2 is calculated as:

$$d(c_1, c_2) = \sum_{i \in N_F} x_i(c_1, c_2)^\lambda f_i(c_1, c_2) \quad (2)$$

where x_i is the degree to which i , and therefore f_i , is included in the soft selection of features, $i \in N_F$ and λ is a constant. Based on the principle presented above, values of vector x are selected through the minimization of distance d , i.e.:

$$x_i(c_1, c_2) = \frac{1}{\sum_i \left[\frac{f_i(c_1, c_2)}{f_i(c_1, c_2)} \right]^{\frac{1}{\lambda-1}}} \quad (3)$$

$$x_i(c_1, c_2) = x_1 \left[\frac{f_i(c_1, c_2)}{f_1(c_1, c_2)} \right]^{\frac{1}{\lambda-1}} \quad (4)$$

Of course, when $\lambda=1$ the solution is trivial and the feature that produces the smallest distance is the only one selected. The degree to which it is selected is 1. A more detailed approach on the issue can be found in [17].

In the following, we present the proposed algorithm implementation with our system's data set, using the Euclidean distance as the distance measure. The clustering algorithm has been applied to a small portion of the dataset, namely a 10% of the overall system's users; it contained 100 elements/users, characterized by 44 meaningful features/questions. The overall dataset consisted of answers provided by 1000 users

from 8 European countries and in 7 European languages, shown in Figure 12, on the same set of 44 selected features.



Figure 12. SPERO dataset languages

Statistical analysis and cross-correlation selection of features indicated the most important ones that were considered in the following. Although meaningful features formed a merely 25%, this proved to be accurate and efficient in the process. They correspond to a set of questionnaire questions together with their possible answering options that are summarized in the following Table 1 and have been considered indicative of the profiling extraction process. Features (column 1) are grouped by the corresponding question id (column 2) of the e-questionnaire. For each verbal question description presented in the third column of Table 1, the number of associated question ids and features used in the clustering procedure are shown in the first and second columns of Table 1 respectively.

<i>Feature</i>	<i>QuestionID</i>	<i>Description</i>
1	125	Are you a teacher dedicated to or working in Special Education Needs?
2, 3, 4	126, 127, 128	Qualification/training in Special Education Needs?
5	130	Teaching Experience
6	133	Do you have a computer at home?
7	134	Do you have access to the Internet from your home?
8	139	How often do you personally use your Internet connection at home?
9, 10, 11, 12, 13	141, 142, 143, 144, 145	For which of the following did you use the computer at least once in the past month?
14, 15, 16, 17, 18, 19, 20	147, 148, 149, 150, 151, 152, 153	Which of the following tasks have you performed at least once, without any help?
21	155	Are there any computers in your work environment?
22, 23	300, 301	How often did you use the computer last week at the school?
24, 25	168, 169	Do you have access to the Internet or educational software in your work environment?
26, 27	174, 175	In your teaching, how many hours a week, on average, do you use the Internet or educational software with your students?
28, 29	176, 177	Do you use the Internet for search and retrieval of information relating to the needs and problems faced by SEN students?

30, 31	182, 183	Do you use the Internet from the school in order to find additional sources of educational material?
32, 33	400, 401	Do you use the Internet to connect with other schools?
34, 35, 36, 37	261, 262, 264, 265	Is your post permanent – temporary?
38, 39, 40, 41, 42, 43	267, 268, 270, 271, 273, 274	Age of your students
44	275	Area served by your school

Table 1. Feature selection

The above elements belonged to three fixed 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, 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 end users at a later stage. Results are shown in Table 2, Table 3 and Table 4, 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 computationally 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 identifiable clusters, optimal results have been obtained for a number of nine clusters, as indicated in the following Table 2, Table 3 and Table 4, where clustering results are presented for three variations of output clusters (3, 5 and 9):

Clusters	Elements	%
1 st	(2, 6, 9)	(11.77%, 35.29%, 52.94%)
2 nd	(11, 2, 25)	(28.95%, 5.26%, 65.79%)
3 rd	(14, 1, 30)	(31.11%, 2.22%, 66.67%)

Table 2. 100 users clustering results – 3 clusters

More specifically, Table 2 presents the clustering results of 100 users. The hierarchical clustering algorithm terminated by the time it reached a threshold of 3 clusters. The first cluster comprises of 17 users, namely 2 belonging to the *Experts* class, 6 to the *Beginners* class and 9 to the *Advanced* class. The corresponding percentage distribution clearly indicates that the 9 *Advanced* dominate the first cluster. The second cluster consists of 38 users: 11 *Experts*, 2 *Beginners* and 25 *Advanced*. *Advanced* are dominant in this cluster as well, whereas in terms of percentages their domination is confident (i.e. 65,79%). Finally, the third cluster contains 45 users, 14 of whom are *Experts*, 1 is *Beginner* and 30 are *Advanced*. Domination of *Advanced* is more indicative, since a 66,67% gives them a clear advantage.

Clusters	Elements	%
1 st	(3, 1, 7)	(27.27%, 9.09%, 63.64%)
2 nd	(5, 1, 8)	(35.72%, 7.14%, 57.14%)
3 rd	(5, 1, 13)	(26.32%, 5.26%, 68.42%)
4 th	(5, 9, 11)	(20.00%, 36.00%, 44.00%)
5 th	(11, 1, 19)	(35.48%, 3.23%, 61.29%)

Table 3. 100 users clustering results - 5 clusters

Table 3 presents the clustering results of the same 100 users, however a new threshold of 5 clusters terminates the clustering algorithm earlier. In this case, all clusters contain lesser users, in comparison to the prior case. The first cluster consists of 11 users, 3 *Experts*, 1 *Beginner* and 7 *Advanced*. The corresponding percentage distribution clearly indicates that the 7 *Advanced* users dominate the first cluster. The same applies to the second cluster with 14 users as well, where 5 users belong to the *Experts*, 1 user is *Beginner* and 8 users are in the intermediate state, i.e. *Advanced*. Third, fourth and fifth clusters are all resulting in the supremacy of *Advanced*; third cluster containing 19 users (distributed accordingly to 5 *Experts*, 1 *Beginner* and 13 *Advanced*), fourth cluster containing 25 users (5 *Experts*, 9 *Beginners* and 11 *Advanced*) and fifth cluster containing 31 users (11 *Experts*, 1 *Beginner* and 19 *Advanced*).

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

Table 4. 100 users clustering results – 9 clusters

Continuing in Table 4, the results of the clustering step demonstrate the clear trend underlying in the system's input data: users are characterized by intermediate ICT skills and expertise. This observation is extremely evident in the third column of Table 4, which indicates clearly that most users 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. Consequently, 5 out of 9 clusters (55.55%) are indicating a clear advantage of the *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. Clusters 8 and 9 are dominated by the *Beginner* profile class, whereas cluster 6 forms a solid representative of the *Experts*.

The above illustrated clustering approach forms the basic procedure, with the aid of which each SPERO end user is automatically categorized to a specific profile class that characterizes his behaviour and his future interests and choices within the system. Additionally, each identified cluster is related to replies and specific comments of the e-questionnaires. We used clustering results to classify the responses of users to specific parts of the e-questionnaires, deriving information based on the users' profiles. The automatic extraction of clusters during the SPERO project provided experts with the ability to distinguish the responses of the clustered users with respect to more specific profiles. More specifically, we combined profiling information from different parts with respect to the users input, a statistical analysis of which is provided in Figure 13, as well as analyzed results belonging to different parts of the e-questionnaires, as illustrated in Figure 14.

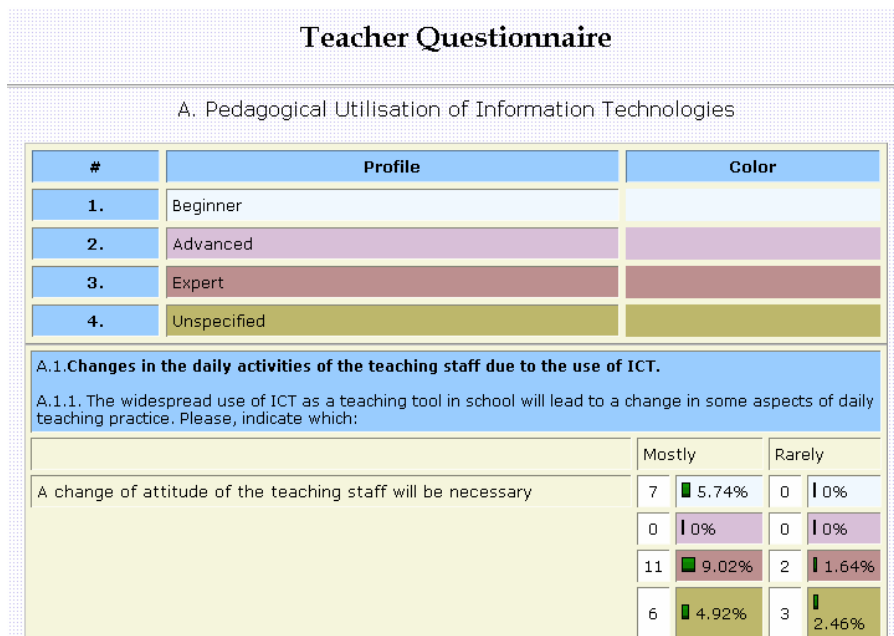


Figure 13. Learning resources linked to profile categories

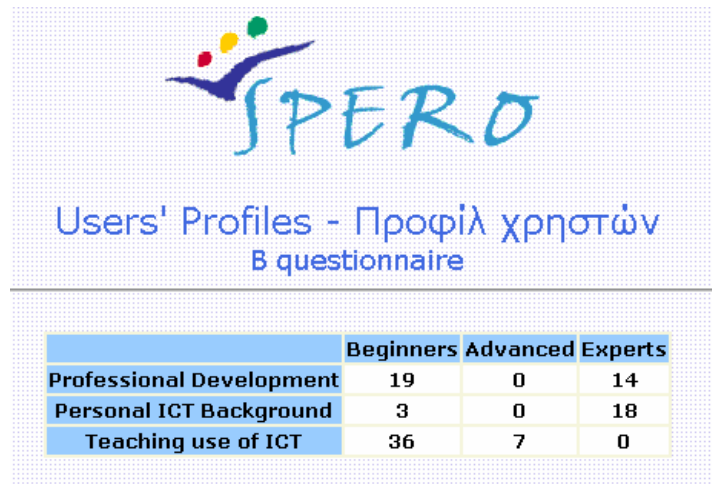



Figure 14. Combined profiling information based on clustering results

This research approach could also be combined with neural network machine learning techniques. For instance, another work of ours described in [17] uses clustering to initialize a three-layer neural network classifier that contains personalization information for extracting the local interest of users. However, this is not considered to be the main focus of the current chapter, which focuses on utilization of clustering towards providing content collaborative recommendations to the end users. According to the cluster to which each user belongs, educational content, appropriately selected by the system's experts, is offered to him. Because of flexibility and protection of crucial personal data reasons, the step of user characterization is only provided as an added value characteristic to the users that are willing to use it. Suitable verification procedures ensure that content offering filtering features are only enabled according to each end user's will.

4. Content adaptation and user tracking

The SPERO system software forms an integrated, web-based learning portal, designed and implemented according to well-known learner-friendly solutions and flexible e-learning software applications [18]. When system's users visit the SPERO portal, validation against the system user database is performed. Subsequently, they are called to answer the e-questionnaires in order to automatically establish their user profile based to the intelligent clustering techniques presented in the previous sections. This automatic profile extraction provides the extremely useful and fully personalized information needed. Learning resources have been linked up to each profile category that has been defined during the profile extraction process and are illustrated in Figure 15. The set of e-courses appropriate for each of the identified groups of user profiles is selected by the group of system's experts.




E-courses

1	ICT Material for Beginner User	The University of Strathclyde's Faculty of Education
2	ICT Material for Beginner User	The University of Edinburgh
3	ICT Material for Beginner User	The University of Glasgow
4	ICT Material for Beginner User	National Register of CPD Providers
1	ICT Material for Advanced User	The University of Strathclyde's Faculty of Education
2	ICT Material for Advanced User	The University of Edinburgh
3	ICT Material for Advanced User	The University of Glasgow
4	ICT Material for Advanced User	National Register of CPD Providers
1	ICT Material for Expert User	The University of Strathclyde's Faculty of Education
2	ICT Material for Expert User	The University of Edinburgh
3	ICT Material for Expert User	The University of Glasgow
4	ICT Material for Expert User	National Register of CPD Providers

Figure 15. Learning resources linked to profile categories

SPERO 's content offering contains links to educational content, separated into various sectors and providing services, like: Courses Catalogue, Announcement Service, Search Service, E-mail Service, Upload Files and Help Service. In particular, the main menu of the SPERO portal contains links to the following sectors/services:

- **Courses Catalogue:** It contains the titles, as well as a small textual description of one or more e-learning courses, that learners may take. An intelligent module takes over the selection of e-courses, according to user preferences and profiles, as well as their usage history. A small overview for each e-course is provided, demonstrating its main topics and concepts. A small notion of a selected e-course listing is presented in Figure 16.



Welcome Paraskevi!

[Courses](#)
[Announcements](#)
[Search](#)
[Email](#)
[Upload File](#)
[Help](#)

Course Details

Sessions	Main Topics	Last Viewed
Getting Started	Menus , Shortcut menus, Toolbars , Customizing toolbars	20.03.2004 12:34
Working With Files	Creating and opening documents , Saving documents, Renaming documents, Working on multiple documents, Close a document	20.03.2004 13:20
Working With Files	Typing and inserting text, Selecting text , Deleting text , Undo, Formatting toolbar , Format Painter	

Figure 16. Personalized e-course listing sample

- **Announcement Service:** This service provides a bulletin board where topics about e-courses or other educational subjects are published. Relative documents, regarding e-courses outlines and requirements are posted herein. Students' and teachers' messages are presented in a threaded view layout.
- **Search Service:** It provides a search environment to facilitate information and educational materials retrievals from SPERO site, e.g. members, school units, e-lessons, e-books, e-lectures, exercises, "live" educational content broadcasts, etc.
- **E-mail Service:** SPERO users are able to send and receive e-mails through the SPERO system.
- **Upload Files:** Learners have their own personal space where they can store their own material to which other learners may or may not have access to. Several levels of authorization access are implemented.
- **Help Service:** Analytical description of the usage and tasks of SPERO menu choices. It provides information about library links and online resources outside the SPERO system and answers general Frequently Asked Questions.

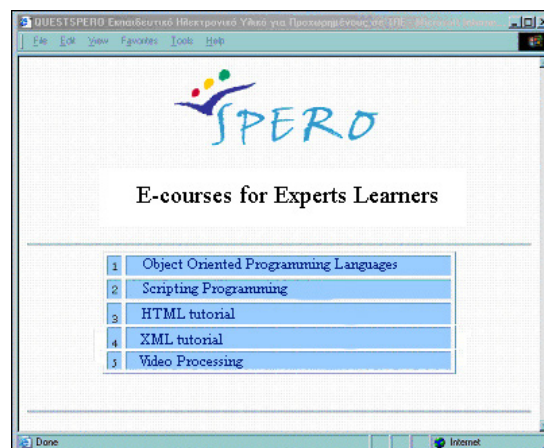


Figure 17. E-courses list for Experts Learners

In order to improve the ICT level of learners, different e-courses are also designed and implemented. Indicatively, groups of e-courses characterized by increasing difficulty and strong topic relativity are possible, such as the following chain of e-courses: *Introduction to Information's Technologies*, *Introduction to Operating Systems*, *Presentation and usage of Office and Educational software*, *Introduction and Usage of Internet*. Each group is characterized by the following aspects:

- **Group 1:** Introduction to Information Technologies (definition of data, bit, byte, presentation of hardware components, presentation of type of software).
- **Group 2:** Presentation and usage of operating systems.
- **Group 3:** Usage of text editor, software for work sheets, software for creation a presentation, educational software.
- **Group 4:** Usage of Internet (explorer in a browser, search machines, sending and receiving e-mails, access to a news group, access to a chat room).

The offered e-courses correspond to the available SPERO user profiles obtained at the previous step. For instance, e-courses for a learner, identified by the system as an

“Expert”, are depicted in Figure 17. This is ensured by an intelligent user tracking mechanism. This mechanism is based on each user individual session, the starts and stops of which are signaled by the time the user enters and leaves the SPERO portal respectively. Session information, along with validation and user access rights is stored in the “userDetails” part of the static user profile. It provides a robust and reliable method to ensure independency amongst the system’s users and efficiency of the whole user profiling-based content offering. As an illustrative example, consider that students who receive an e-course are tracked throughout the SPERO system and their behavior is observed and tracked internally. The overall procedure is transparent and provides the main source of feedback for the users' future system (e-courses and material) selections. Additionally, learners’ interaction is periodically monitored to allow dynamically change of their user profiles, according to their improvement of ICT skills and skillfulness. For instance, if a user studies, then it is foreseeable that he/she will improve his/her profile. Towards that scope, it is possible for the learner to re-enter the SPERO portal and provide new data input for the e-questionnaires. Then and based on his/hers unique id, the system monitors the user’s progress and changes his/hers ratings and corresponding content offering. This leaves an option for the learner that the system can provide and sets him/her on a specific personalized track.

5. Conclusions and future work

The methodology presented in this book chapter can be exploited towards the development of more intelligent, efficient and personalized educational content offering systems, thus enhancing the learning experience of their end-users. The proposed approach forms an integrated, state-of-the-art system that is able to identify its individual users. During this process, it extends work performed on precise, high level personalization algorithms, as it utilizes (transparently to the end-user) personalization techniques towards profiling extraction, introducing a novel conjunction of static and dynamic profiling mechanisms. It was also within our intentions in this book chapter to focus on acquired representative results of this work. For more e-learning type of results the reader is encouraged to visit SPERO’s online applications (<http://www.image.ntua.gr/spero>). Moreover, the main research effort of this chapter was to successfully introduce and apply clustering techniques in the process of user profiling and provide collaborative recommendations of e-learning content to the end users.

A major area of future research for this work is the utilization of a fuzzy relational knowledge representation model in the learners' profile weight estimation process. Our findings so far indicate, that such a combination between semantic and statistical information is possible and will have very interesting results, regarding the personalization of the educational content offered to the end-users. Additionally, the work proposed herein could also be enriched by combining it with neural network machine learning techniques. This work is part of our ongoing efforts in the field of designing and implementing an integrated, fully automated e-learning portal system. The main focus is given to the personalization aspects of the system’s user handling and educational content offering. Possible future work includes better selection of the clustering algorithm threshold criteria and possible increase of the static profiles

categories. Moreover, the system's e-questionnaires are susceptible to evaluation and improvements, as well as an increase in the number of participants in the e-surveys is viable. The overall proposed architecture of SPERO could be easily adapted to other e-learning schemes, mainly due to its robustness and entities clarity.

References

- [1] Rosenberg J. M., 2002. E-Learning: Strategies for Delivering Knowledge in the Digital Age. McGraw-Hill, Inc, ISBN:007137809X
- [2] Karagiannidis, C., Sampson, D.G., and Cardinali, F., 2002. An architecture for Web-based e-Learning promoting reusable adaptive educational e-content. *Educational Technology & Society*, 5(4), pp. 27-37.
- [3] Linser R., Waniganayake M. and Wilks S., 2004. Where is the Teacher? e-Learning Technology, Authority and Authorship in Teaching and Learning, In Proc. of "EDMEDIA-2004", Lugano, Switzerland.
- [4] Varenne H., 2001. On pedagogical authority: One teacher's choice. [online]. Available from: http://varenne.tc.columbia.edu/hv/pol/pedagog_author.html
- [5] Linser R., 2004. Suppose you were someone else: The learning environment of a web-based role-play simulation. In SITE 2004 Conference Proceedings, Atlanta GA, USA.
- [6] SPERO, 2001-2004. Leonardo Da Vinci programme, Themes 5: E-Learning. [online]. Available from: <http://www.image.ntua.gr/spero>
- [7] Skordalakis E.. 1991. Introduction to Software Engineering. Athens.
- [8] IEEE LTSC, 2002. P1484.1 Architecture and Reference Model. WG1 LTSA. Available from: <http://ieeeltsc.org/>
- [9] IEEE LTSC, WG11 Computer Managed Instruction, 2007. Available from: <http://www.ieeeltsc.org/working-groups/wg11CMI/>
- [10] Mylonas Ph., Tzouveli P. and Kollias S., 2004. Towards a personalized e-learning scheme for teachers. 4th IEEE International Conference on Advanced Learning Technologies, Joensuu, Finland.
- [11] Ph. Mylonas, M. Wallace, S. Kollias, 2004. Using k-nearest neighbour and feature selection as an improvement to hierarchical clustering. Proceedings of 3rd Hellenic Conference on Artificial Intelligence, Samos, Greece.
- [12] Godoy D., Amandi A., 2005. User profiling in personal information agents: a survey. *The Knowledge Engineering Review*, Vol. 20 (4), pp. 329 – 361, Cambridge University Press.
- [13] 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
- [14] Theodoridis S. and K. Koutroubas, 1999. Pattern Recognition. Academic Press. San Diego
- [15] 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
- [16] Wallace M. and Stamou G., 2002. Towards a Context Aware Mining of User Interests for Consumption of Multimedia Documents. IEEE International Conference on Multimedia and Expo (ICME), Lausanne, Switzerland

- [17] Wallace M. and Kollias S., 2003. Soft Attribute Selection for Hierarchical Clustering in High Dimensions. International Fuzzy Systems Association World Congress (IFSA), Istanbul, Turkey.
- [18] Blackboard Academic Suite™: Course Management System. Available from: http://www.blackboard.com/products/Academic_Suite/index