An advanced system for automated certification

Ioannis Samoladas, Alexandros Nanopoulos, Athanasis Karoulis, Ioannis Stamelos and Yannis Manolopoulos

Aristotle University of Thessaloniki, Greece

Key words: *Tele-learning, Distance Certification Systems, Data Mining*

Abstract:

Internet has revolutionized many aspects of our everyday life. Education could not stay intact; Internet has introduced the great ability of distance learning, which allows the traditional educational process to overcome several limitations. Among the requirements that accrue naturally, is the assessment of the results for the new type of educational process. This requirement is denoted as distance certification. While numerous organizations have emerged to support the certification process, the presence of systems that will automate this process and will render it clear and objective, is limited. In this paper, we first describe the requirements that have to be addressed for the development of such a system. For a better understanding of the requirements we modelled the certification process and simulate it. The certification model was built according to the most recent standards and processes that major certification providers follow. Next, we propose ASPIS, which is a system that have been developed to meet these requirements. ASPIS takes into account various learning parameters and makes use of the feedback of the process. The proposed system is web based and it is built with a three tier architecture, it is highly modularized and it can be easily extended and modified. The whole application utilizes free/open source software, it is written in php, it functions upon the Apache web server and its database system is MySQL. Furthermore its web interface is compliant with the most recent web standards, separating content from presentation, making it accessible by all operating systems that utilize a web standard browser and with low effort by other handheld devices. Additionally, ASPIS exploits the advantages offered by the data-mining technology, to further assist the distance certification. The advantages of the proposed system are illustrated through several results of the data mining we applied on the system's preliminary data.

1 Introduction

During the last decade the wide spread of the Internet has changed many areas of everyday life, especially in communications and news spreading. Additionally the way business is conducted has also changed and electronic commerce has gained a lot of attention. Another important area of our lifes that has changed with the introduction of the Internet is the way we learn. Distance learning systems are now a reality and a lot of organizations are, now, offering courses that are conducted over the web and a large number of people are enrolled in them.

Additionally, big multinational companies use the Internet to keep their employees up-to-date with their job area, offering to them life long education. These courses cover a wide range of subjects, but the majority of them involve information technology areas.

All these organizations that provide distance learning courses need a mechanism to support this life-long learning process in order to track the performance of their students, a mechanism that provides distance certification. While there are a number of companies in other countries that offer a distance certification application and procedure, this is not the case in Greece. Furthermore, the support for the certification process is limited and there are certain problems with the inflexible process. The certification process does not take account of the learning process (teaching methods, training centers, educational staff, etc.) and there is not a formal process to accredit learning providers (or learning centers) and examination centers. Additionally, the results of the certification process are not used as a feedback in order to evaluate and make the process better.

In order to address the problems mentioned, we propose the ASPIS (the name ASPIS is the Greek acronym for "Automated Certification System") system. To our best knowledge, this is the first system in Greece that offers an online automation of both the certification process and the accreditation process for learning centers and the examination centers. The system keeps track of the whole learning and examination process.

The ASPIS system provides data (e.g., about examinees, education centers, tests, etc) that can be exploited in order to find useful information, which can be used to assess the whole educational process and carefully monitor specific parts of it. For this purpose, we describe the use of the analysis module that is included in ASPIS, which performs data mining on the data collected by the system. Up to the authors knowledge, relatively little interest has been given during the past years in applying data mining for educational-related purposes. We detail the data-mining methodology that is used and illustrate some sample results, with emphasis on the qualitative characterization of the results according to three parameters of the human cognition.

2 Certification - accreditation processes

Before implementing the ASPIS system it was necessary to define a certification process both for the examination process and the accreditation process of the examination centers and learning providers by a central authority (in our case the Aristotle University of Thessaloniki). These processes were used in order to specify the system's requirements [5]. To take an idea of such process we have conducted interviews with the industrial partners of the ASPIS project, from those interviews we have modeled a general framework that can be easily applied and altered in order to capture a specific process of a learning organization.

Organizations wishing to offer courses leading to certificates have to be accredited by the central authority mentioned above, according to already established standards [4], [7]. Likewise, organizations wishing to implement the process of examinations and offer it to candidates have to be accredited. Once accredited, learning centers (learning providers) and examination centers are monitored in a variety of ways to ensure that standards are fulfilled.

The general process for certifying the candidate is shown in Figure 1. The candidate is registered in the examination center by bringing complete probative elements of his identity. He acquires a unique username and password to log into the examination system and take exams in the required module(s) (e.g. word processing, spreadsheets). The examinations can be written offline or online (in our case, the ASPIS system accommodates only online examinations). Examination results are published shortly after.(e.g. either online or with regular post) and candidates have the right to appeal to any step of the process.

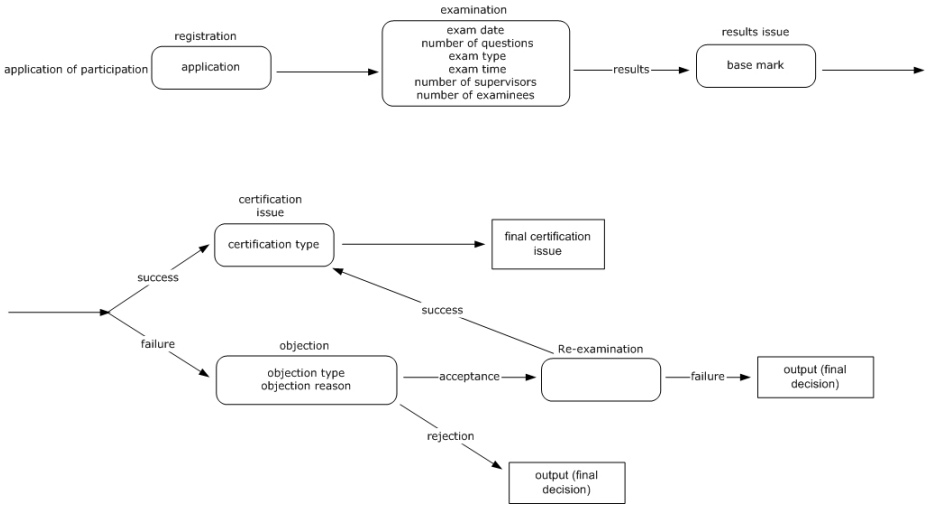

Fig. 1. Process diagram for the exam

The general process for accrediting the organizations wishing to act either as learning providers or examination centers is shown in Figure 2. The interested organization submits an application to the central authority. After a visit of evaluation, the authority decides whether to grant the accreditation. If the accreditation is not granted, the rejected organization has the right to appeal. If the accreditation is granted, the organization that was accredited must comply with the accreditation rules.

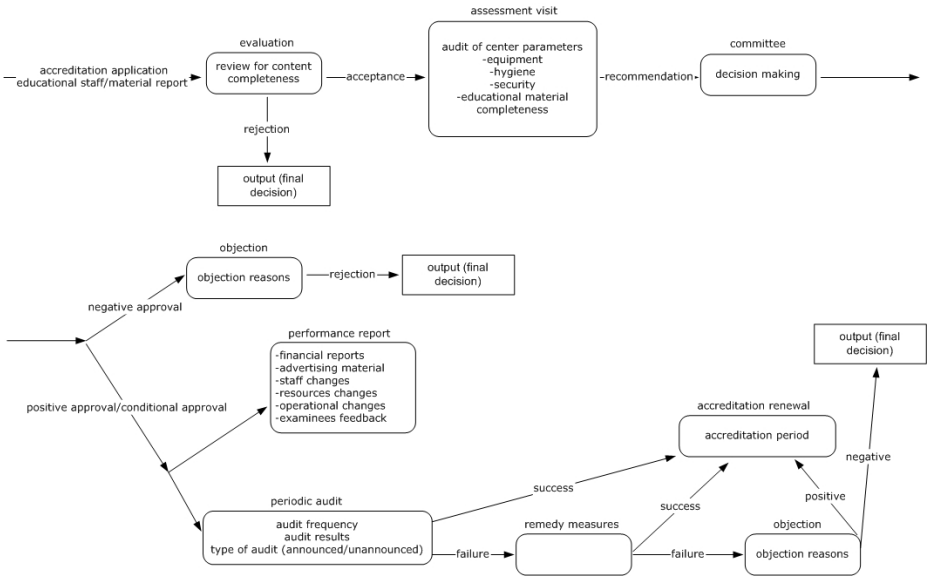

Fig. 2. Process diagram for center accreditation

3 The ASPIS System

3.1 Overview

The ASPIS system is a modern web based system, i.e. a web based database application. The system supports two functions, accreditation and examination. For the first, it offers automation regarding the tracking of the process for the accreditation and the maintenance of the existing accredited centers (e.g. reminders for the periodic audit). It does not support cases and processes that include strong human interference, for example decision making regarding the grant of the accreditation. For the second function the system support almost full automation of the certification process (the automation is not supported when the examination process demands a written exam, which is not the usual case for such system).

3.2 Architecture

The ASPIS system is a three-tier application, like most of the recent web based systems. The base of the system is the relational database management system (RDBMS) or the database tier, which is the MySQL database management system [reference]. Above that tier there is the application logic or the middle tier, which performs most of the systems functions and it is also responsible for the communication between the upper tier of the system and the database tier. The language we used to write this tier was PHP, a server side scripting language suitable for building web based applications. This tier also includes the web server that stands above the PHP engine. On the top there is the client tier which is the web browser that interacts with the user. All these three tiers are cooperating in order to carry out the functions that the ASPIS system performs. The general architecture is shown in Figure 3.

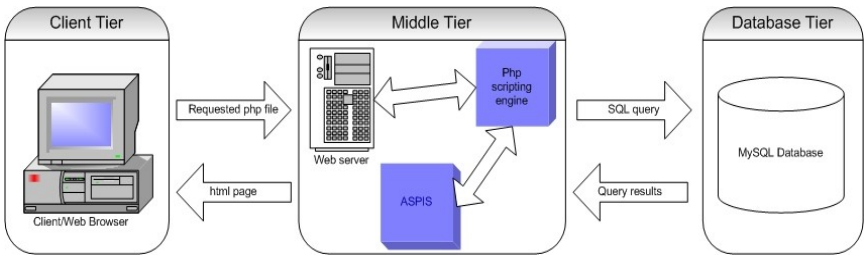

Fig. 3. The 3-tier architecture

Particularly, when a client or more precisely a web browser makes a request to the web server (middle tier) for a specific page, five steps occur:

1. The web server passes the request to the PHP engine's web server interface.
2. The web server interface calls the PHP engine and passes the parameters to the engine.
3. The appropriate PHP script from the ASPIS system is retrieved from the disk by the engine.
4. The script is executed by the engine and the output is returned to the web server interface.
5. The web server interface returns output to the web server which in turn returns the output as an http (common html file) response to the client's web browser.

Although PHP can be easily embedded in html web pages, for the ASPIS system this was not the case. Instead, PHP was used to build the web pages from scratch without using existing html code. These pages are built according to specific parameters as an input to PHP and the user's rights. This means, for example, that there are no separate pages for the main page of the student or the administrator, but a single script builds the page according to the role of the user logged into the system. This feature gives an advantage to the whole system in terms of maintainability. New features can be easily implemented by adding new function to the module that builds the pages instead of adding new raw script files to the application.

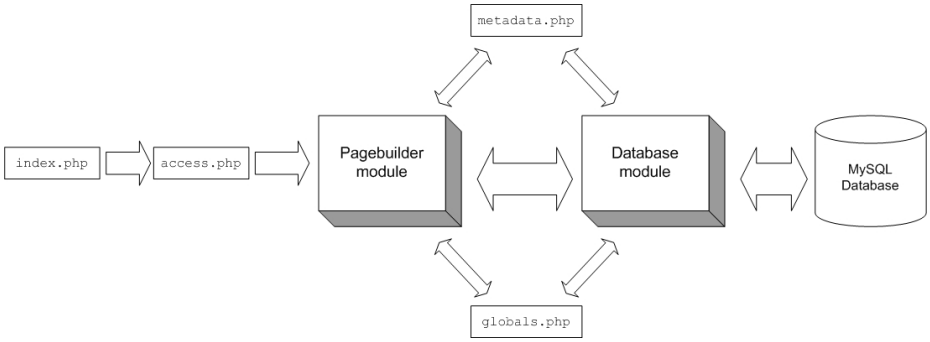

Fig. 4. ASPIS architecture

The ASPIS system consists of two main modules and one auxiliary module. The main modules are the *Pagebuilder module* and the *Database module*. The architectural view is shown in Figure 4. The auxiliary or Data Mining module is separate from the other two (not shown here). This module is not written in PHP and it directly interacts with the database, bypassing the PHP engine that the rest of the system uses. The functionality contained in the first two modules is further divided into other scripts or modules:

- There are three important scripts in ASPIS that contain information for the two main modules. The first script is `access.php`, which is responsible for user authentication and passes the appropriate role parameters to the Pagebuilder module. The second important script, `metadata.php`, contains all of the system's information - the database schema and all the information that the second script wants in order to build pages, e.g. information about forms. Additionally, it implements all the validation functionality, e.g. whether the user has entered an incorrect email address. The `globals.php` script is in fact the configuration file for the application and the database connections, containing information about how to log into the database and other system and database specific constants and parameters.
- The Pagebuilder module, as its name implies, is responsible for the construction of the majority of the web pages that the ASPIS consists of. It constructs the pages that data is entered or displayed in, i.e. the whole application. For each of one displayed item, it uses a separate function; we could say that it is the API for these functions. For example, the function `buildMenu` constructs the menu of the main page of the user according to his rights (for example, the student has a menu in order to take a test and the administrator has options like adding a new user to the system). It also adds the appropriate html tags, cascading style sheets and javascript code to the page constructed and uses the scripts mentioned in the first bullet.
- The Database module contains all the appropriate functions that the system uses in order to communicate with the MySQL database management system and either

inserts new data or processes old ones. This module uses the information from the `metadata.php` and the `globals.php` mentioned before. This module has all the functions that interact with the database in the lower level through the PHP engine, i.e. the sql queries. Whenever access to the database is needed, this script is used through the function it provides. If a new query has to be inserted to the system, here is where we implement it. This allows, apart from the modularity of the system, future migration to another database system.

Of course there are web pages that do not belong to the above, like the login page, which passes parameters to the Pagebuilder module.

It worths mentioning that the whole application is written according to the most up-to-date web standards as these are specified by the World Wide Web Consortium (W3C). This means that the content is separate from the presentation. The content is coded in pure xhtml (a reformulation of html as an xml application) and the presentation or the style is applied by various CSS (Cascading Style Sheets) without affecting the content. This means, first that by applying different style sheets the user interface can be replaced easily, leaving content untouched, and second that having the content free from the way it is presented, it can be easily displayed in other media than the computer screen, i.e. handheld devices e.t.c.

3.3 ASPIS Functionality

As mentioned earlier, at the beginning of this section, the ASPIS system supports automation both for the accrediation and the certification process. The accreditation process, as it can be seen from the previous sections, is mostrly bureaucratic, it evolves filing applications and granting accreditations. Thus the accreditation subsystem is actually a system that tracks the progress of an accreditation application and its status. Here we are going to present the certification process subsystem.

3.3.1 ASPIS Users for the Certification Subsystem

Regarding the certification process, the ASPIS system has three type of users: the students who give tests and take exams, the educator who specify tests, and the administrators who perform tasks such as adding new students to the system. The specific funtions that each user can perform are:

- *Student*. After a successful login to the system a student can (a) see the tests that are available for him and choose the test he/she wants to take, (b) To see all the past tests he/she has given. We have to note here that a student cannot give a test that has already given, but when he chooses a past test he/she sees his/her performace for that specific test.
- *Instructor*. After a successful login an instrustor can (a) input a new test specification to the system, (b) see the performance of the students assigned to him (c) correct scenarios that students have given and assign a mark to them.
- *Administrator*. After successful login and administrator can perform tasks such as (a) insert a new question into the database and edit the existing ones, (b) add new users to the system and edit the existing ones and (c) organize the testing material into subjects and sections and edit them

Here we have to mention that the system gives the administrator the ability to assign to specific instructors additional rights in order to insert, view and edit the questions that the system has in his database. Fig. 5 presents the form that adds a new user to the system.

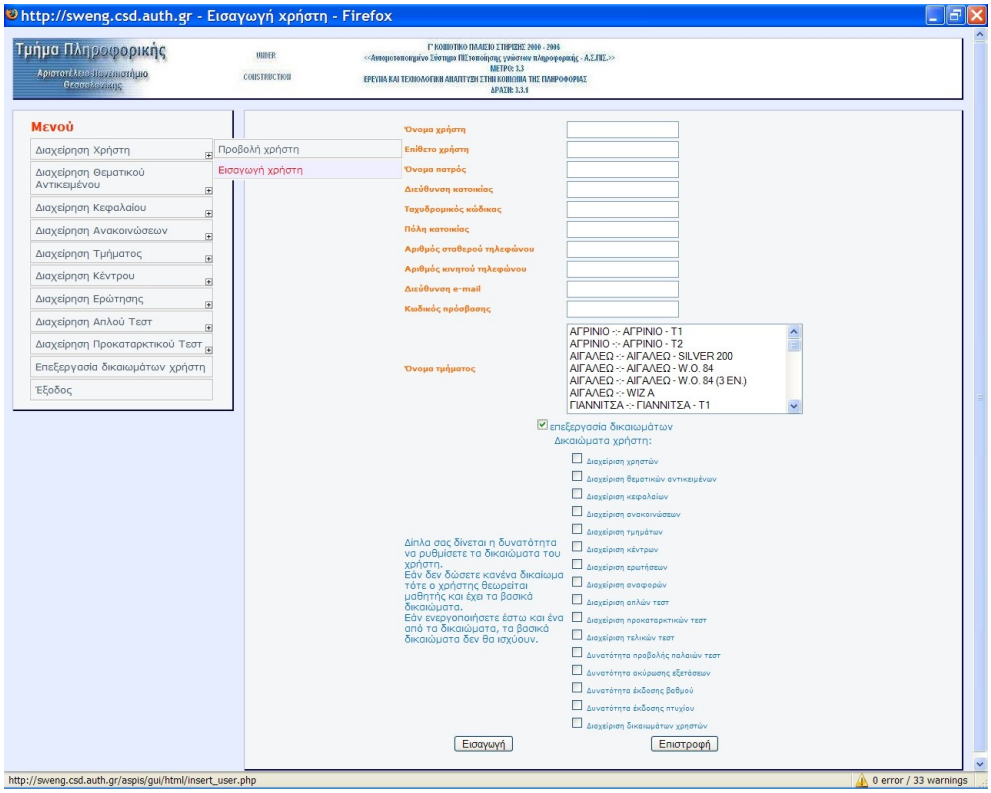

Fig. 5. Add new user to the system

Additionaly for organizational purposes students belong to classes, with every class to be unique. The same is true for the instructors. This creates a relationship between students and instructors by creating these virtual classes.

3.3.2 Question Characteristics

A question in the ASPIS system can be one of the following three types: multiple choice (including true/false type questions), fill in the blank and matching questions. Apart from its type, every question has the following characteristics:

- Question text
- Any images that accompany the test
- Possible answer (in case of multiple choice and matching)
- Correct answer
- To which subject and section of the teaching material the question refers to
- Difficulty level of the question

Apart from the question the systems gives tha ability to have scenarios. Scenarios are specific tasks that students have to perform in order to achive a desirable result, that will enable them to pass an exam. For example a scenario could contain task such as to create a Word document containing material from several resources (e.g. Other files and pictures) and format it according with certain instructions, mentioned in the scenario. Fig. 6 shows the form that adds a new question to the system.

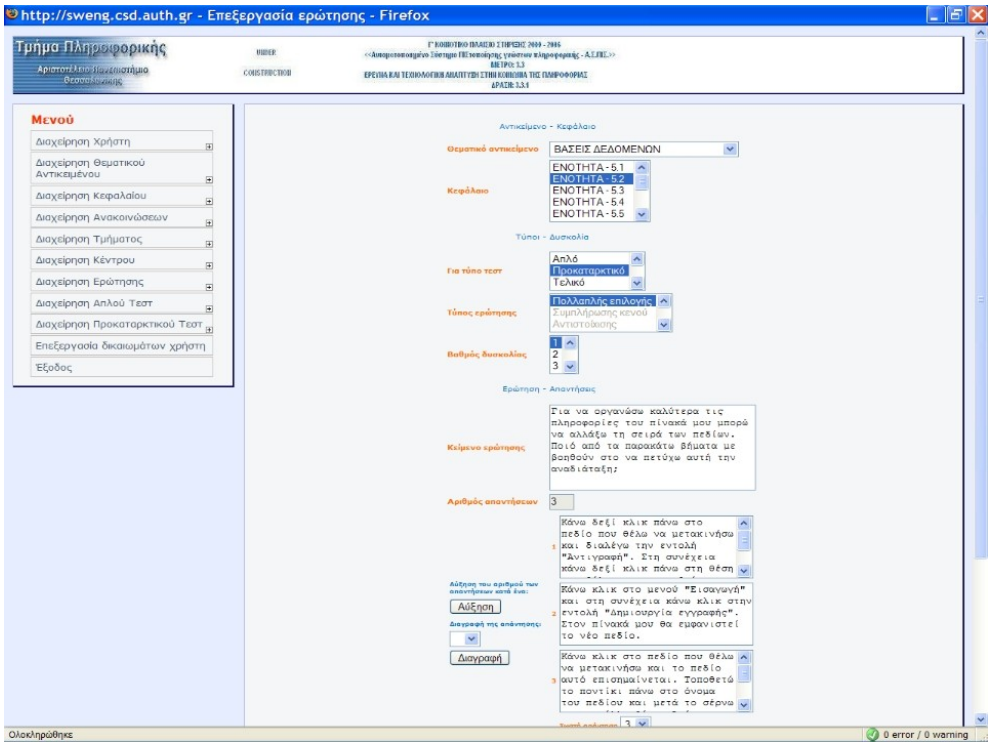

Fig. 6. Add new question to the system

3.3.3 Test Characteristics

In ASPIS when an instructor inserts to the system a new test, what he/she actualy does is to insert a new specification of a test and not the actual test with specific questions and scenarios. In particular when a test is inserted to the system the instructor enters the following data:

- A descrite name for the test, in order to refer uniquely to it
- A small textual discription of the test
- The subject and the sections that this test examines
- The distribution of the question into the various difficulty categories (e.g. ten questions with difficulty one, five with two and three with one)
- The expiration date of the test. A test can be active for a certain period and it is automatically deactivated after that date
- Whether or not the test includes a scenario

When a student chosses to give a test, the actual test is created in real time, dynamically , according with its specification. Thus the questions are chosen randomly and every student has a different test in front of him, all with the same specifications. Of course some questions are the same, but the order is surely different. Fig. 7 shows a form that it used to submit a test specification.

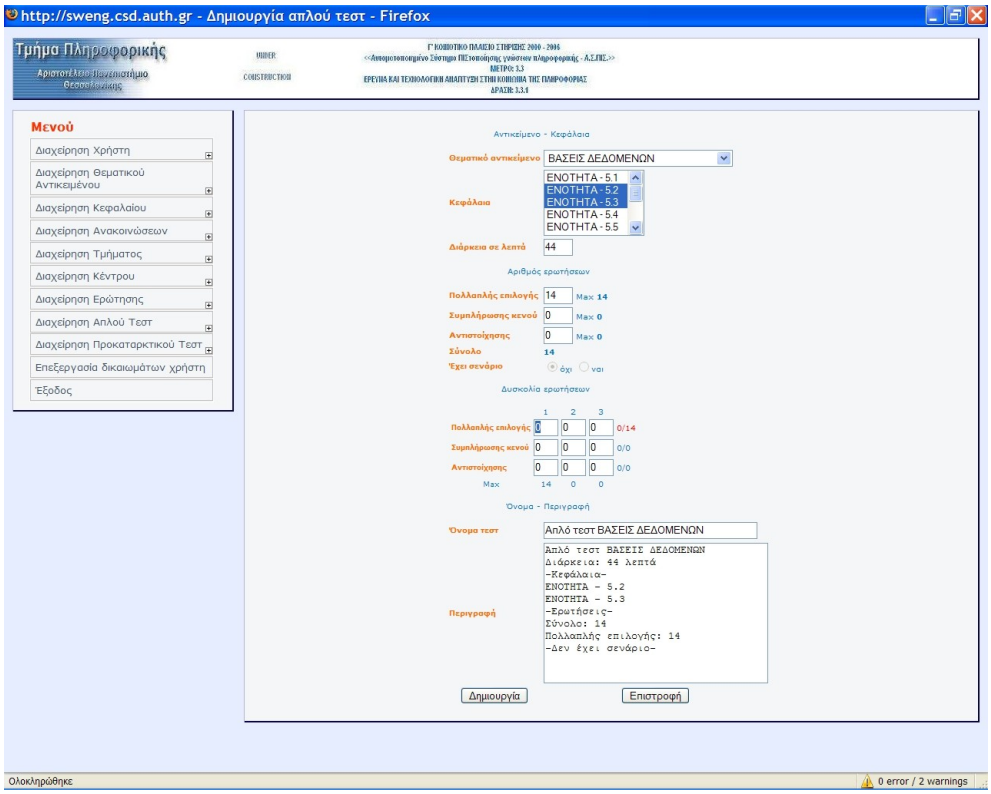

Fig. 7. Submit a new test specification

4 Case Study of ASPIS

Table 1. Usage statistics of the ASPIS system

Total number of accounts	**2403**
Total number of students	**2309**
Total number of educators	**77**
Total number of administrators	**17**
Total number of classes	**448**
Total number of study areas	**16**
Total number of chapters (of study areas)	**80**
Total number of questions	**1329**
Total number of test specifications submitted	**27077**
Total number of tests submitted	**81561**

In order to evaluate the ASPIS system, we first implemented a prototype and performed a pilot operational use in a real educational environment. The system was installed by a large Greek educational organization (and an industrial partner of the ASPIS project) that offers computer training in a number of cities across Greece. During that period a lot of data was collected and users provided a lot of feedback about the system itself. This feedback helped in improving the system, especially its user interface which changed, without altering its core design and functionality. The final system is also online and used by the same educational

organization mentioned before. Some numbers that indicate the massive usage of the system until March 2005 are shown in Table 1. The size of the database has reached 291.4 megabytes.

5 Data Mining

5.1 Objectinves of applying data mining in ASPIS

During its operation, the ASPIS system produces a relatively large volume of data. This includes data about examinees (personal data like address, education center); data about education centers (name, address); data about the tests, i.e., the contents of the database of questions asked to and answers provided by the examinees during their tests; etc. Evidently, such data can be transformed to valuable information, which can be used to (i) improve the ability to assess the whole process and (ii) to monitor specific parts of it in order to adjust accordingly.

As it has been recognized during the previous years, the technology to attain objectives analogous to the aforementioned ones is data mining. Data mining is the extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from data in large databases [2]. Data mining has been primarily used in financial/marketing applications like target marketing, determination of customer purchasing patterns, and cross-market analysis. Other application areas include telecommunications, insurance, astronomy, etc. However, relatively little interest has been given during the past years in applying data mining for education-related purposes.

Given the two (i and ii) objectives described previously, we consider the following data mining tasks:

1. Mine relationships between data about the examinees (e.g., region of lodging, level of class, etc), data about the examination centers (their names), and data about their examination results (total score). The discovered relationships of this kind will allow for better assessment of the educational process (objective i), since it is helpful to understand how the aforesaid factors interrelate and which of them affect the success or failure of examinees.
2. Mine relationships and investigate patterns between the answers of the participants. The discovered relationships allow for better monitoring of the educational part of the system, that is, the questions used in the tests and, accordingly, their educational value. If, for example, one can find questions, or categories of questions, that tend to be answered in the same way (i.e., correctly or incorrectly), then one may have a better understanding on how the taught subjects interrelate and obtain a picture of the educational model the particular student develops through the interaction with the system. So, one could adjust the initial educational approach in order to meet students' expectations and developed cognitive models on the particular domain.
3. To advocate the understanding of the resulting patterns, we consider their characterization according to three parameters of the human cognition process, which can interpret phenomena of cognitive associations.

The data mining method that we used to examine the previous issues is association-rules mining, which is detailed in the following.

5.2 Mining association rules in ASPIS

Association-rules mining concerns the finding of frequent patterns called associations (other names: correlations, causal structures) among sets of items or objects in transaction databases,

relational databases, and other information repositories. An almost legendary example of a discovered association rule is:

- buys(x, “diapers”) → buys(x, “beers”) [0.5%, 60%]
- which states that if one purchases diapers, then he is also likely to purchase beer. As shown, an association rule is characterized by two measures:
- The support s% (in previous example: 0.5%), which reflects the statistical significance of the rule, that is, how many transactions include the items of the rule.
- The confidence c% (in previous example: 60%), which reflects the assurance of the rule, that is, how sure we are that the items in the head and the body of the rule really relate.

The problem of association-rules mining is, therefore, to find all rules with support higher than a given threshold s% and a confidence higher than a given threshold c%. This problem has attracted a lot of interest and many algorithms have been developed [1]. We focus on the family of algorithms that are based on *pattern-growth* and more particularly on *FP-growth* [3], since they have been reported to have very good properties.
In ASPIS, we created a separate module in the architecture, which is responsible for probing the database of questions and answers, and for finding association rules of type 1 and 2, that have been described previously. In particular, the analyst can query the database and formulate a view containing information about the total scores (case 1) or the correctly/incorrectly answered questions per examinee (case 2). The selection can be further focused by including the time period of interest. Each record in the aforementioned views corresponds to transactions, and we use the FP-growth algorithm to mine associations. Examples of discovered patterns, which indicate the usefulness of this module, are given in the following.

5.3 Examples of mined results

By using a sample of the tests’ database, consisting of 81561 tests, we mined association rules for cases 1 and 2 that were described previously.
For the former case (1), we have discretized the ratings of examinees (originally in 0-100 scale) in 8 levels: A, B, … G, by equi-partitioning the original scale. We focused on relationships between the city of residence of examinees and their rating. The rules we detected (support threshold 1%, confidence threshold 55%) were the following:

- Residence(x, “Kozani”) → Rating(x, “G”) [2%, 67%]
- Residence(x, “Thessaloniki”) → Rating(x, “A”) [6%, 66%]
- Residence(x, “Florina”) → Rating(x, “A”) [1%, 58%]

The utility of this kind of rules is evident, because they are significant (the original sample was large) and one can detect clear relationships. Thus, such rules can help in monitoring the educational process, since regions that tend to have low ratings can be looked more carefully to examine the reasons. Moreover, it is also useful to know the regions where the education process has good results.
For the case 2, we have mined association rules between the correctly and incorrectly answers by each examinee. A sample of the detected rules for the former case is the following:

- “What appears after then end of booting?”, “Give a way to terminate a program that got stuck” → “Which icon do we use to undelete a file?” [76%]
- “What appears after then end of booting?”, “Give a way to terminate a program that got stuck” → “Which of the listed actions cannot be performed with the mouse?” [76%]

A sample of detected rule for the latter case is the following:

- "Is it possible to minimize a dialog box?", "In which way can we minimize all opened windows?" → "From the listed items, which is not a component of MS Windows?" [73%]

From such rules one can see topics that tend to be understood (correctly answered) or not understood (incorrectly answered), for instance the minimization procedure, and react accordingly by emphasizing such topics (and their relation) during classes.

5.4 Analyzing the patterns according to cognitive parameters

In this section we elaborate further on the third task that was described in Section 5.1. According to contemporary literature, human cognition is usually associated to following cognitive parameters: *perception*, *attention*, *memory*, *processing*, *reasoning*, *problem solving*, and *time perception*. For the purposes of data mining in the results of ASPIS, we focused on the following three parameters in order to elicit useful results from the derived patterns:

- Memory: indicates in how far the user can remember the semantic and/or the use of the questioned entity. A typical question of this kind is "which of the following completes the action…"
- Perception: refers to the ability of the user to recognize the presence and/or the utility of the entity in the interface. A typical question of this kind is "which of the following shoud one activate in order to…"
- Reasoning: refers to the ability of the user to interrelate the particular entity to others, which interfere in a way with the questioned one. In other words, to investigate relations between entities (and accordingly depict them correctly in the questions). A typical question of this kind is "which of the following statements is correct…"

Usually, while the user performs the test, the questioned interface is not available for experimentation, so a great deal of memory is employed by the user in all questions. On the other hand, there is not pure memory or perception or reasoning cognition which would satisfactorily answer the particular question, but merely a combination of these attributes. However, at any given question, one of these parameters clearly dominates over the others, characterizing, thus, the question as memory or perception or reasoning oriented. Bearing that in mind, one can attempt to classify the questions of the tests in one of these categories. Emerging patterns of interrelations between various questions of the tests indicate cognitive interactions between these particular subsets of questions, which can be further analyzed and discussed in detail later.

To exemplify the aforementioned procedure, we present the following characterization of interactions between the three categories in the results of the case-study that was described in the previous section. The association rules between correctly answered questions mainly involve questions in the memory and perception categories. This may indicate a problem with questions belonging to the third category (reasoning), which can help in improving the teaching methodologies towards increasing the aspect of interrelating entities. Regarding the associations between incorrectly answered questions, among the highly ranked rules (w.r.t. confidence and support thresholds) we found correlations between the three categories that are shown in Table 2:

Table 2. Correlations between cognitive parameters

	Memory	*Perception*	*Reasoning*
Memory	1	4	1
Perception	4	0	0
Reasoning	1	0	0

We notice a higher correlation between incorrectly answered questions that belong in categories memory and perception. This may indicate that the teaching process can be improved towards increasing the combined understanding of these two categories.

References:

[1]. Agrawal, R., Imielinski, T., Swami, A. N. Mining Association Rules between Sets of Items in Large Databases. ACM SIGMOD Conference (1993) 207-216

[2]. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press (1996)

[3]. Han, J., Pei, J., Yin, Y. Mining Frequent Patterns without Candidate Generation. ACM SIGMOD Conference (2000) 1-12

[4]. Kirkpatrick, D. Evaluating training programs: The four Levels, San Francisco, CA:Berrett-Koehler (1996)

[5]. Mooij, T., & Smeets, E. Modeling and supporting ICT implementation in secondary schools. Computers & Education, 36 (2001), 265-281

[6]. MySQL AB. http://www.mysql.com

[7]. Wisher, R. A., & Champagne, M. V. Distance learning and training: An evaluation perspective. In J. D. Fletcher and S. Tobias (Eds.), Training and Retraining: A Handbook for Business, Industry, Government and the Military. MacMillan Reference USA (2000)

Author(s):

Ioannis Samoladas, PhD Candidate
Aristotle University of Thessaloniki, Department of Informatics
Thessaloniki 541 24, Greece
ioansam@csd.auth.gr

Alexandros Nanopoulos, PhD
Aristotle University of Thessaloniki, Department of Informatics
Thessaloniki 541 24, Greece
alex@delab.csd.auth.gr

Athanasis Karoulis, PhD
Aristotle University of Thessaloniki, Department of Informatics
Thessaloniki 541 24, Greece
karoulis@csd.auth.gr

Ioannis Stamelos, Assistant Professor
Aristotle University of Thessaloniki, Department of Informatics
Thessaloniki 541 24, Greece
stamelos@csd.auth.gr

Yannis Manolopoulos, Professor
Aristotle University of Thessaloniki, Department of Informatics
Thessaloniki 541 24, Greece
manolopo@delab.csd.auth.gr