UDC 004.02:007

Ye.V. NIKITENKO*, O.V. TRUNOVA*

QUALITY TESTING AT ASSESSMENT OF PROFESSIONAL COMPETENCIES IN WEB LEARNING MANAGEMENT SYSTEM

*Chernihiv National University of Technology, Chernihiv, Ukraine

Анотація. *У статті пропонуються шляхи підвищення якості тестування при оцінюванні рівня професійних компетенцій у веб-системі управління навчанням, зокрема, розглянуто архітектуру системи та алгоритм її роботи. Запропонований модуль перевірки критеріальної валідності тестів ґрунтується на аналізі показників асиметрії, ексцесу та індексу дискримінації; сукупність наведених характеристик вказує на способи валідизації, що дозволяє підвищити якість тестування. Розроблений модуль надання та оцінки тестів на основі нечітких множин надає можливість достатньо точно виставити оцінку при проходженні тестування і при цьому проводити інтегральний облік як кількісних, так і якісних факторів.*
Ключові слова: *веб-система управління навчанням, модуль перевірки валідності, модуль надання та оцінки тестів.*

Аннотация. *В статье предлагаются пути повышения качества тестирования при оценке уровня профессиональных компетенций в веб-системе управления обучением, в частности, рассмотрены архитектура системы и алгоритм ее работы. Предложенный модуль проверки критериальной валидности тестов основывается на анализе показателей асимметрии, эксцесса и индекса дискриминации; совокупность приведенных характеристик указывает на способы валидизации, что позволяет повысить качество тестирования. Разработанный модуль предоставления и оценки тестов на основе нечетких множеств дает возможность достаточно точно выставить оценку при прохождении тестирования и при этом проводить интегральный учет как количественных, так и качественных факторов.*
Ключевые слова: *веб-система управления обучением, модуль проверки валидности, модуль предоставления и оценки тестов.*

Abstract. *The article offers ways to improve the quality of testing in the assessment of professional competence in Learning Management System, including the reviewed system architecture and algorithm of its work. The proposed module of checking the criterion validity tests based on the analysis of indicators of asymmetry, kurtosis and discrimination index, set of these characteristics indicates on the validation methods that can increase the quality of testing. Designed module of providing and assessment of tests based on fuzzy sets allows you assessing accurately at the testing and thus pursue integrated account both of quantitative and qualitative factors.*
Keywords: *web learning management system, the module of checking the validity, module of providing and assessment of tests.*

1. Problem statement

At the present stage of development of Ukraine, it is essential to improve the system of traditional university education. Creation and implementation of state standards of higher education require further improvement and the search for effective forms, tools and training methods, new methods of evaluating student achievement, creation an appropriate learning opportunities of students learning model in higher education.

One way to improve the quality of the educational process in institutions of higher education is the introduction of automated training system (ATS). ATS – a system that helps to master new material performs control of knowledge and helps teachers in preparing teaching material.

However, the results of such systems are perceived ambiguously, that hinders their widespread use [1].

One of the aspects that should be considered in these systems is evaluation of level of professional competencies that is the result of specific existing knowledge and skills, that is, what a student should have at the end of the course. The elimination of subjectivity in the examination of professional competence, promotes new advanced forms of electronic (computer) control (testing), which is part of the ATS.

The current education system is characterized with intensive development of the theory and practice of computer testing, which becomes an integral and important part of the educational information environment of any educational institution.

Computer testing is a method of control, which is a standardized procedure of using tests on a computer under a special software control that provides the required presentation of tests and test results processing to solve complex task goals.

The most significant claim to the currently existing distance learning systems such as Moodle, ILIAS, Web Tutor, CoureLab, Blackboard Learning System is a primitive control and evaluation of test results.

From the mentioned above we can make a conclusion about feasibility of allocation of individual independent components in the architecture of the developing system. This will make changes directly to individual modules without affecting the system as a whole.

The aim of article is to improve the quality of testing in the assessment of level of professional competence in Web-Learning Management System.

2. System architecture

The server of authorization processes the requests of unauthorized users to the system. Following the link to the site in the browser, the user can enter a name and password for authentication. Thanks to simple logic of work, the request to search a user in the system can be performed directly in the database. Depending on the type of user, a redirection takes place to one of the modules – test or viewing the results and quality of tests. A special key will be written to the browser's session, without which further work with the system will be impossible.

Server of web-services contains a set of Web services that are required for the work with database, services of providing and processing test results, also services of providing data about their results and quality. Modules of passing tests and module of viewing their results and quality will access data services.

The module of passing tests is a web-server that allows the passage of the tests through a graphical Web interface. The results of passing the tests can be stored in the database by calling web-service server with web services. Module of viewing the results and quality tests is a web-server. It uses a graphical web interface and allows you to view the results of passed tests, that stores in a database, as well as their quality.

When using the module of creation the graphs, it can visually learn the results. Control Module of quality tests – is servers with applications that enable to call Web- services of processing data and receiving results.

Module of competency assessment allows evaluating professional competence of students.

The system architecture is shown in fig. 1.

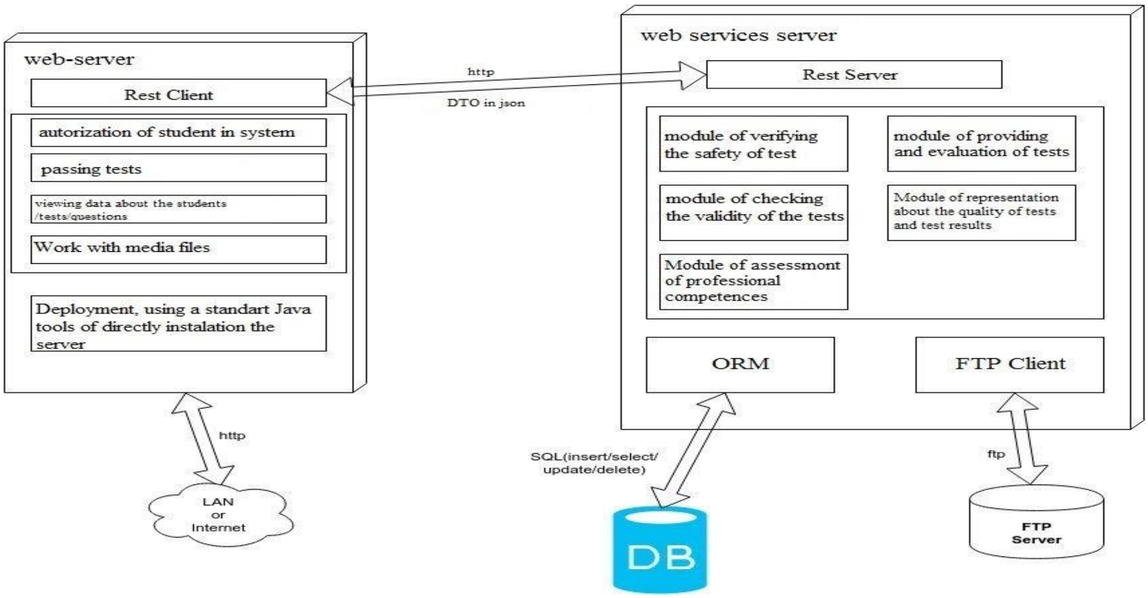

Fig. 1. System Architecture

3. The algorithm of the system

The procedure of detection of professional competencies in automated mode is implemented in two phases: the first phase includes the formation of test for students (from a set of test's tasks). It consists of two parts:

1. Experts offer a set of questions of custom test.
2. The test is verified on quality (validity and reliability).

2.1. Construction and analysis of matrix of results.

2.2. Determination of complexity and item determination of tests.

2.3. Determination of variation of marks (difference of maximum and minimum dialed scores), asymmetry, kurtosis and discrimination index).

Based on the questions that have been passed quality control, the base of the questions of user test is made, thus questions are divided into categories for the evaluation of a set of professional competencies.

The second phase involves obtaining the final test result in dependence on the importance of specific competence for individual student after passing user's test.

Block diagram of previously given process is shown in fig. 2.

4. The module of the validity checking

The validity in theory of testing means compliance of form and content of the test to that it has to evaluate or measure on a plan of its creators.

Validity – one of the main criteria for a quality of test, the degree of coverage by the test the set of objects subject area compared with the standard. Validation – is not about gathering evidence of validity of the test, but about the process of actions, that raise its validation. As a result, the evidence base of the validity of the test will grow.

In the ATS the following analysis scheme of criteria l validity (provides the availability of external criterion – standard, which defines the validity of test) test's tasks are offered, that are based on the results of statistical data of testing with building a curve of distribution of correct answer's number on specific task of test. Not valid tasks considered those tasks, which had been received during the testing the following statements the right answers more than 84% of tested persons; correct answers less than 16% of tested persons.

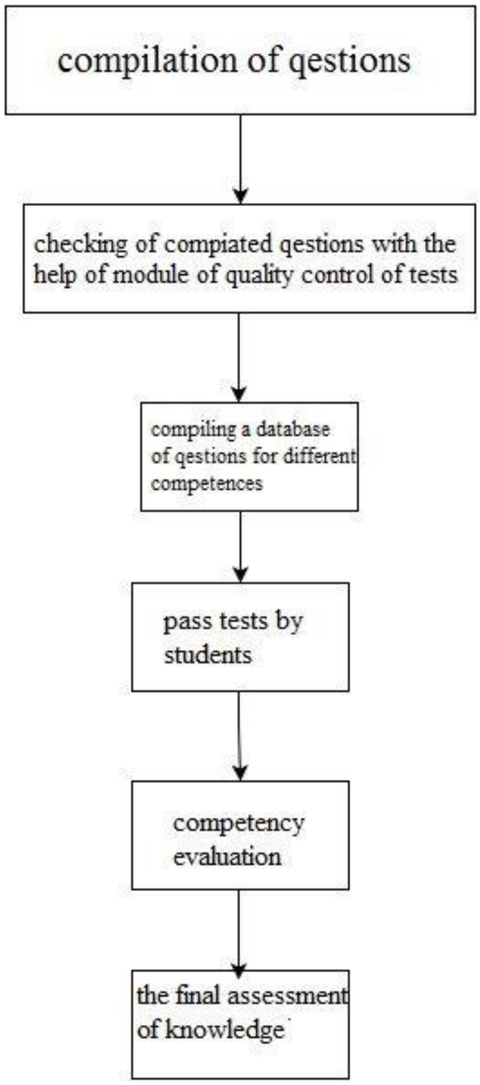

Fig. 2. Block diagram of the algorithm of the system

Distributional (discriminatory) ability of the test task defines by the index (discrimination index), which indicates to what extent the results obtained in the performance of individual test tasks correlate with the results of all test and how well it's test task distinguishes tested persons with high mark and tested persons with low mark.

Distributional ability can be calculated as the correlation coefficient between the mark for whole test and a mark for a particular test task. The correlation coefficient is calculated as follows:

$$r_{x,y} = \frac{1}{s_x}(X_B - X_H) \cdot \sqrt{p_j \cdot q_j},$$

where X_T – arithmetic mean score of tested persons, that successfully completed the j-th task of test, X_N – the arithmetic mean score of tested persons, that not coped with the j-th task, $\sqrt{p_j \cdot q_j}$ – standard deviation of the j-th task, S_x – standard deviation throughout the test. The value of the correlation coefficient is interpreted as follows: 0,7–1 – the relationship is very strong; 0,5–0,7 – average; 0,3–0,5 – weak.

The index can range from -1 to +1. When the index is «0», it means that all tested persons responded the same way (good or bad). If the correlation is positive (higher than 0), it is a sufficient discriminatory ability of this test, it means that it has the sufficient ability to determine strong and weak of the tested persons. It is believed that difficult tasks have to be solved successfully by more prepared tested persons and easy tasks – less prepared, but good prepared also. If test results show a different picture, this test does not meet the characteristics and needs high-quality processing.

General characteristic of test for validity provides comparisons of indicators of asymmetry, kurtosis and discrimination index see table 1, which is directly implemented in the system.

Table 1. Analysis of tests for validity

As	*Ex*	$r_{x,y}$	Characteristic	Methods of validation
= 0	= 0		The valid test	
< 0	> 0	0,3-0,5	easy test with low resolution	to complicate the task to increase the resolution of problems improve administration of procedure testing
< 0	< 0	0,7-1	easy test with high resolution	to complicate the task
< 0	= 0		easy test	to complicate the task
> 0	< 0	0,7-1	complex test with high-resolution	remove or simplify tasks that make testing difficult
> 0	> 0	0,3-0,5	complex test with low-resolution	validation of tasks
> 0	= 0		complex test	validation of tasks
= 0	> 0	0,3-0,5	heterogeneous test (presence of heavy and light tasks) with low resolution	validation of the test is the division into two separate tests.
= 0	< 0	0,7-1	heterogeneous test (presence of heavy and light tasks) with high resolution	validation of the test is the division into separate tests.

5. Module of provision and assessment of tests

Getting the final test result has the goal to objectively assess the level of training of students, in particular to reduce the subjectivity that appears between teachers and students, and eliminate other (probably hidden) factors that interfere. Evaluation of the test task is that the student accumulates points for the test, and then total result is translated into all the usual estimate.

But, first, various questions from the proposed list obviously are measured in different units, and therefore the direct storage system is problematic. Second, as a rule, the translation method of total result into the assessment remains outside of using fuzzy logic [2], or applies the standard approach of working with random variables [4].

When building a model of formation of linguistic evaluation of student's competences, in our view, as input variables should be used as a quantitative factors (Q – number of questions, b_k – the number of correct answers, Σ – total score), and qualitative factors (x_1 – mastering level on 2 points. (unsatisfactory); x_2 – level – on 3 (satisfactory); x_3 – level – on 4 (good); x_4 – level – on 5 (excellent)) [2].

Model of linguistic assessment of competence of the student (F) in general can be represented as follows (fig. 3).

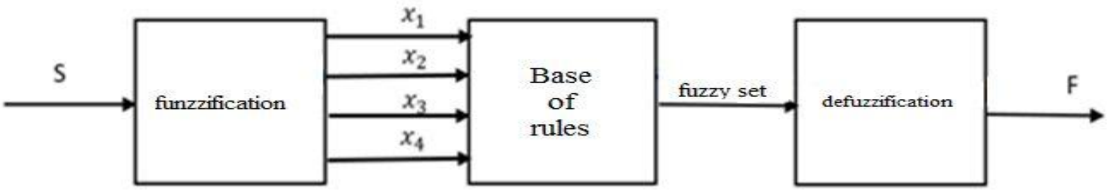

Fig. 3. The formation model of linguistic assessment of competence level, where *S* – the sum of points obtained by the student, $x_1 - x_4$ – the level of assessment

The algorithm of formation of linguistic competence assessment of students is as follows:

1. To carry out the rationing of the accumulated amount of points accumulated on the interval [0;100] by levels.

Determine auxiliary variables x_i, $i=\overline{1,4}$ normalizing the accumulated amount of points on the interval [0;100] regardless of the number of questions in the test:

$$x_1{}' = S-(2k_1(\text{int}(\frac{Q}{15})+\text{int}(\frac{Q-5}{15}))+k_1\,\text{int}(\frac{Q-10}{15})),$$

$$x_2{}' = \begin{cases} S-(3k_1(\text{int}(\frac{Q}{10})+2k_1(\frac{Q}{10}-1{,}5)), & if \quad Q-unpaired \quad (Q\,\text{mod}\,2 <> 0); \\ S-(3k_1+2k_1)\cdot(\frac{Q}{10}-1), & if \quad Q-pair \quad (Q\,\text{mod}\,2=0); \end{cases}$$

$$x_3{}' = \begin{cases} S-(2k_2(\text{int}(\frac{Q}{10})+3k_2(\frac{Q}{10}-1{,}5)), & if \quad Q-unpaired \quad (Q\,\text{mod}\,2 <> 0); \\ S-(2k_2+3k_2)\cdot(\frac{Q}{10}-1), & if \quad Q-pair \quad (Q\,\text{mod}\,2=0); \end{cases}$$

$$x_4{}' = \frac{S\cdot(k_3\cdot(Q-2)+1)}{((k_3\cdot(Q-2)+1)+4\cdot\frac{Q-10}{5}))},$$

$$S = k_1b_1+k_2b_2+k_3b_3,$$

where S – cumulative total score;

Q – the number of questions in the test, at $Q \geq 10$, $Q\,mod$ 2=0;

b_1 – the number of correct answers to the difficulty level 1;

b_2 – the number of correct answers to the difficulty level 2;

b_3 – the number of correct answers to the difficulty level 3;

k_1 – coefficient for the correctness of answer to questions of 1 level of complexity;

k_2 – coefficient for the correctness of answer to questions of 2 level of complexity;

k_3 – coefficient for the correctness of answer to questions of 3 level of complexity, at $0 < k_1 < k_2 < k_3$;

int – division function that returns an integer;

mod – division function that returns the remainder.

The variable x_i, $i=\overline{1,4}$ (x_1 – unsatisfactory level; x_2 – satisfactory level; x_3 – good level; x_4 – excellent level) standardized depending on meanings of the auxiliary variable x_i on interval [0;100]:

$$
x_i = \begin{cases} 20x_i', & if \quad x_i' \le 1{,}2, \\ 24 + 62{,}5(x_i' - 1{,}2), & if \quad 1{,}2 < x_i' \le 1{,}6, \\ 49 + 20(x_i' - 1{,}6), & if \quad 1{,}6 < x_i' \le 2, \\ 57 + 12{,}8(x_i' - 2), & if \quad 2 < x_i' \le 2{,}7, \\ 66 + 11{,}4(x_i' - 2{,}7), & if \quad 2{,}7 < x_i' \le 3{,}4, \\ 74 + 4{,}5(x_i' - 3{,}4), & if \quad 3{,}4 < x_i' \le 5{,}4, \\ 100 - 3(7{,}4 - x_i'), & if \quad 5{,}4 < x_i' < 7{,}4, \\ 100, & if \quad x_i' \ge 7{,}4. \end{cases}
$$

Formulas compiled in accordance with the 100-point scale that used to convert points into traditional assessment and numeric value x_i, $i = \overline{1,4}$ obtained in the experiment (see table 2).

Table 2. The numerical values necessary for the formation of the argument x of membership functions of quality of output variable

Transition coefficient			1,2		16		2,0		2,7		3,4		5,4		7,4
The numerical value of membership function		20		62,5		20		12,8		11,4		4,5			
Total points for all types of the educational activities			24		49		57		66		74		83		100
Evaluation by national scale	x_1 – unsatisfactorily							x_2 – satisfactorily			x_3 – good				x_4 – excellent
Assessment ECTS	F				FX			E	D		C		B		A

For example, the value 7,4 corresponds to the maximum value x_4 at the passage of the adaptive test which begins from 1 level of complexity; values 3,4; 2 and 1,2 – the average value x_3, x_2, x_1, the values 1,6; 2,7 and 5,4 obtained by adding the value of the left value and the difference between neighboring values divided in half (1,6=1,2+(2–1,2)/2; 2,7=2+(3,4–2)/2, and so on. d.).

On each of the received segments there were compliance coefficients of numerical values to marks. For example, between [3,4; 5,4] are points from 74 to 83, that's why a numerical value on this interval corresponds to 4.5 points ((83–74)/(5,4–3,4)).

2. Write the classification scale of linguistic variables x_i, $i=\overline{1,4}$ Linguistic variables: x_1 – level of 2 points (unsatisfactory); x_2 – level – of 3 (satisfactory); x_3 – level – of 4 (good); x_4 – level – of 5 (excellent), interpreting as a term-set with three-dimensional scale T2={NL, CS, CC}, where value NL – «does not correspond to the level of (2, 3, 4 or 5)», CS – «corresponds to slightly» and CC – «completely corresponds».

3. Ask a membership function of quality of variables x_i, $i=\overline{1,4}$.

Each of linguistic variables «level on 2 (3, 4, 5)» has one triangular curve of membership (1) and two T-curves of membership (2) ($\mu_{x_i}^{NL}, \mu_{x_i}^{CS}, \mu_{x_i}^{CC}$, $i=\overline{1,4}$), that in general may be given by expressions:

$$\mu_{\Delta}^{j}(x_i,a,b,c)=\begin{cases} 0, & x\le a, \\ \dfrac{x-a}{b-a}, & a<x\le b, \\ \dfrac{c-x}{c-b}, & b<x\le c, \\ 0, & c<x, \end{cases} \qquad (1)$$

where $i=\overline{1,4}$; $j\in\{T2\}$; a, b, c – some numerical parameters that characterize base of the triangle a, c and its top (b), moreover, the condition $a\le b\le c$ must be met.

$$\mu_{T}^{j}(x_i,a,b,c,d)=\begin{cases} 0, & x\le a, \\ \dfrac{x-a}{b-a}, & a<x\le b, \\ 1, & b<x\le c, \\ \dfrac{d-x}{d-c}, & c<x\le d, \\ 0, & d<x, \end{cases} \qquad (2)$$

where $i=\overline{1,4}$; $j\in\{T2\}$; a, b, c, d – some numerical parameters that characterize the lower base of the trapezoid a, d and the upper base of the trapezoid b, c, that acquire arbitrary real values and organized by ratio: $a\le b\le c\le d$.

With considering (1) and (2) the membership function of fuzzy-term set of linguistic variable «level 2 (3, 4, 5)» x_1, x_2, x_3, x_4 will be as follows:

1) $\mu_{T x_1}^{NL}(x_1, 0, 0, 15, 30)$; $\mu_{\Delta x_1}^{CS}(x_1, 20, 35, 50)$; $\mu_{\Delta x_1}^{CC}(x_1, 40,55,100, 100)$;

2) $\mu_{T x_2}^{NL}(x_2, 0, 0, 20, 40)$; $\mu_{\Delta x_2}^{CS}(x_2, 30, 50, 70)$; $\mu_{\Delta x_2}^{CC}(x_2, 60, 70, 100, 100)$;

3) $\mu_{T x_3}^{NL}(x_3, 0, 0, 50, 70)$; $\mu_{\Delta x_3}^{CS}(x_3, 55, 70, 85)$; $\mu_{\Delta x_3}^{CC}(x_3, 70, 90, 100, 100)$;

4) $\mu_{T x_4}^{NL}(x_4, 0, 0, 70, 80)$; $\mu_{\Delta x_4}^{CS}(x_4, 70, 80, 90)$; $\mu_{\Delta x_4}^{CC}(x_4, 80, 90, 100, 100)$.

Implementation of vague conclusions is based on logic algorithm of Mamdani [Mamdani], because the problem is solved of gaining knowledge from data (in the form of linguistic rules). The algorithm works like a «black box». The input received numerical value, is the output [3].

The parameters a, c, b, d can be adjusted according to experimental data.

When using classifiers (triple and quaternary scales) on the carriers of fuzzy set the importance of linguistic variables is defined on the real axis interval [0; 1].

4. Determine the classification scale and membership functions of quality of investigated parameter (output variable) «Evaluation of student competencies» as a term-set of values T1={unsatisfactory (*U*), satisfactory (*S*), good (*G*) excellent (*E*)}. Linguistic variable «Evaluation of student competencies» has two triangular curves of membership and two T-shaped curves of membership $\mu_F^U, \mu_F^S, \mu_F^G, \mu_F^E$.

With considering (1) and (2) the membership functions of fuzzy-term set of linguistic variable «Evaluation of student competencies» (*F*) will have the following form: μ_F^U (x, 0, 0, 25, 50); μ_F^S (x, 40, 55, 70); μ_F^G (x, 60, 75, 90); μ_F^E (x, 80, 90, 100, 100).

5. Identify the knowledge base that needed to set fuzzy production rules of assessment of student competencies is formed by specialists of subject area. It is as follows (table 3).

Table 3. Fuzzy production rules

Rules	View term	Variable				F
		x_1	x_2	x_3	x_4	
PR1	NL	+				F^E – excellent
	CS		+	+	+	
	CC				+	
PR2	NL	+				F^G – good
	CS		+		+	
	CC			+		
PR3	NL				+	F^G – good
	CS		+			
	CC		+	+		
PR4	NL					F^G - good
	CS	+				
	CC					
PR5	NL	+			+	F^S – satisfactory
	CS	+		+		
	CC		+			
PR6	NL	+		+	+	F^S – satisfactory
	CS	+	+			
	CC					
PR7	NL		+	+	+	F^U – unsatisfactory
	CS	+				
	CC	+				
PR8	NL				+	F^U – unsatisfactory
	CS		+	+		
	CC	+				

In our view, at the development of competency level assessment module the most appropriate is integration already existing in environment MATLAB package of Fuzzy Logic Toolbox in created software addition. Fig. 4–5 shows windows of the editor of membership functions and the editor of fuzzy inference rules of membership of environment MATLAB Fuzzy Logic Toolbox, where alternatively defined membership functions for each of the terms of the input and output variables.

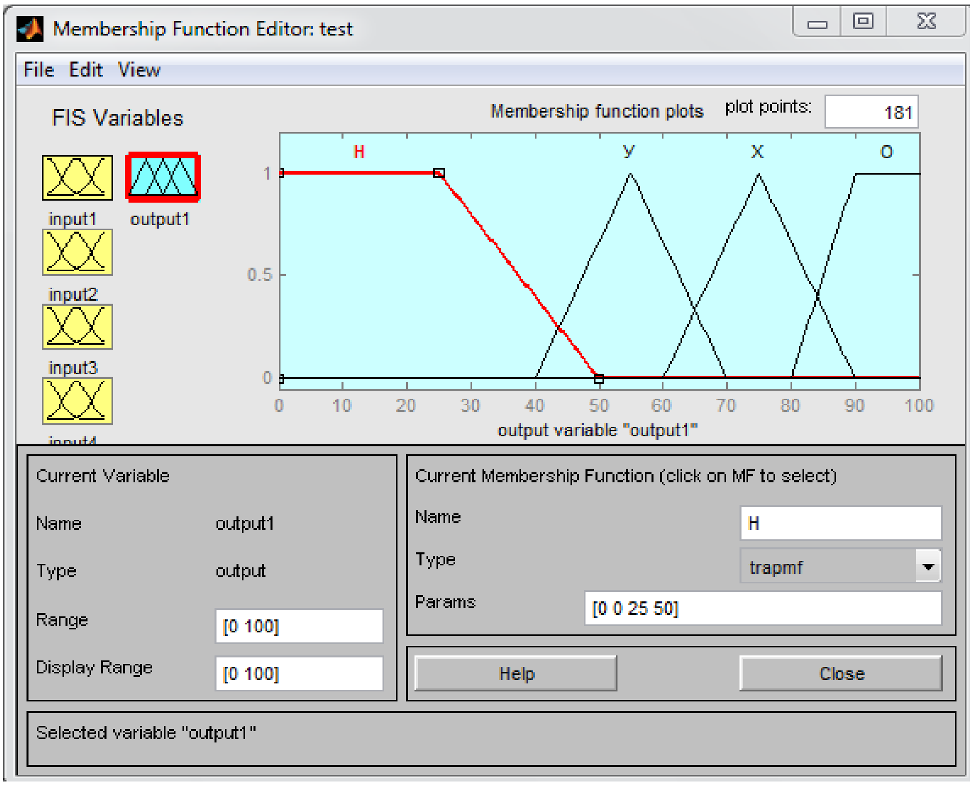

Fig. 4. The editor of membership functions

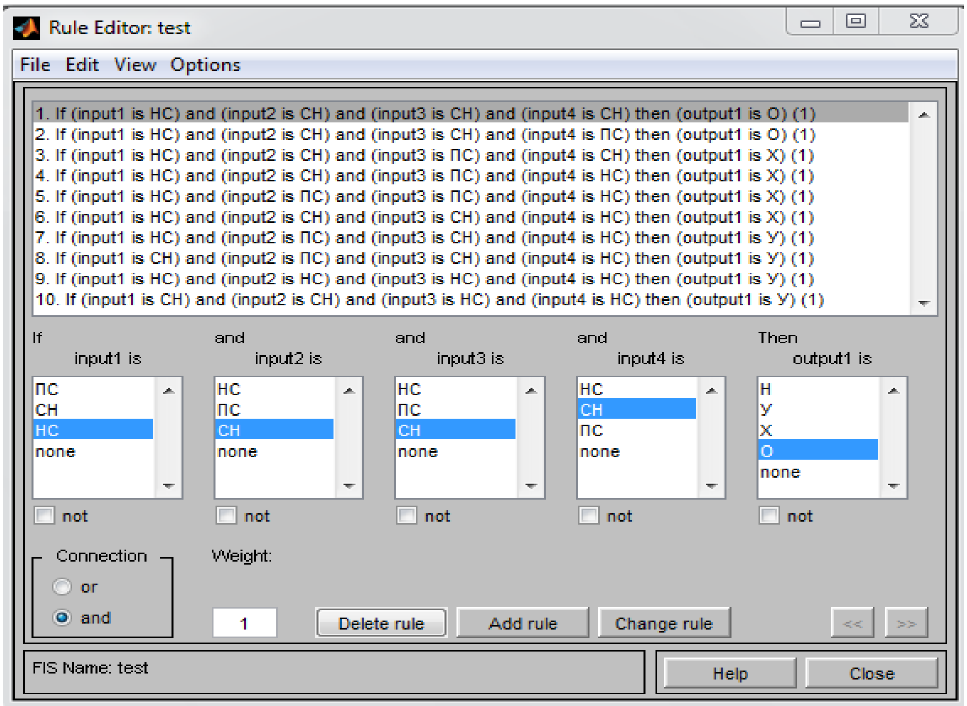

Fig. 5. Rule Editor of fuzzy inference system

6. Accumulation of conclusion by all rules carried out using max-disjunction operation. The main methods of defuzzification (transformation of fuzzy set of findings in a clear number) are a lot of methods of reduction to the definition: the method of selecting a maximum of membership function; method of center of gravity; median method; the method of choosing center of highs etc.

Experimental studies prove that the most accurate – is a method of center of gravity (see fig. 6) for a discrete set of values of membership functions

$$f^{g} = \frac{\sum_{r=1}^{f_{\max}} f_r \mu_Б(f_r)}{\sum_{r=1}^{f_{\max}} \mu_Б(f_r)},$$

where $f_{\max}$ – number of elements in f_r sampled for calculation of the «center of gravity» area F [5].

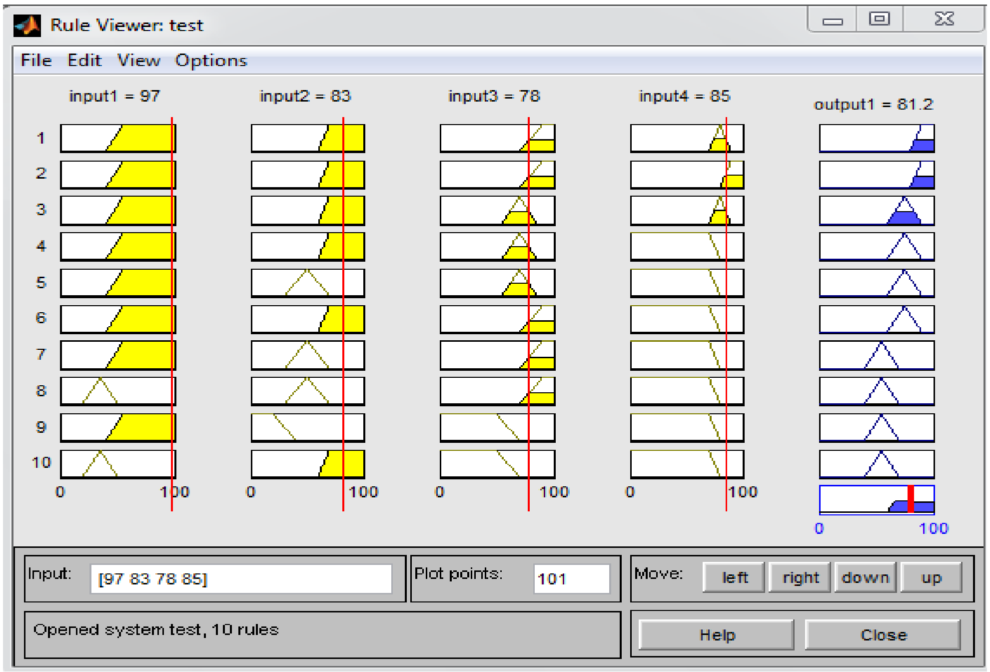

Fig. 6. Rules viewer program

7. Evaluation of student mastering level of competencies derives from the maximum amount of points that equal to 100 [4]. If each of level of test's tasks $F_1, F_2, ..., F_n$ is known a linguistic evaluation $\mu(x) = (\mu_1(x), \mu_2(x), ..., \mu_n(x))$ and the weight's coefficients were defined $p = (p_1, p_2, ..., p_n)$, then the operator of information's aggregation is the weighted sum and is characterized by its linguistic assessment, that defined by membership function on 01-classifier

$$\mu_F(x) = \sum_{i=1}^{n} \mu_i(x) \cdot p_i .$$

It should be noted that the results of the testing (evaluation of mastering level of student's competence) can be used for future monitoring of professional's level of competence and checking validity as a reference.

6. Conclusions

The results of approbation of created Web system, whose main function is to automate the process of determining the level of formation of professional competence of graduates in each of the selected areas indicate the reliability and efficiency of adaptive testing environment. Simplicity, adaptability and versatility of the software complex allow to use it for either examination of professional competence or to assess knowledge of the discipline.

It was experimentally established that using the algorithm of formation the linguistic evaluation of student's competences based on vague sets allows you to accurately enough set the assessment at the testing and thus pursue integrated account both quantitative (number of questions, the number of correct answers, total score) and qualitative factors (x_1 – the level of assimilation on 2 (unsatisfactory), x_2 – level on – 3 (satisfactory), x_3 – level on – 4 (good), x_4 – level on – 5 (excellent), considering the uncertainty of the last. By installing criterion significance of reliability's level of membership functions of input (output) variable's quality, you can change the final results depending on the level of preparedness of students.

The developed application has the following advantages: learning regardless of time and spatial location; checking tests for the validity allows teachers to choose more quality educational material and does not go beyond student's knowledge level; the ability to view material not just as text, but also through media increases the effectiveness of training program; the ability to view test results and demonstration of statistics as graphs are more clearly.

REFERENCES

1. Національна доктрина розвитку освіти України у XXI столітті: проект // Освіта. – 2001. – № 60 – 62. – 24–31 жовтня.

2. Домрачев В.Г. Нечеткие модели рейтинговых систем оценки знаний [Электронный ресурс] / В.Г. Домрачев, О.М. Полещук, И.В. Ретинская. – Режим доступу: http://www.ict.edu.ru/vconf/files/tm01_627.doc.

3. Дьяконов В. Алгоритмы нечёткого вывода: алгоритм Мамдани и алгоритм Сугэно / В. Дьяконов, В. Круглов // Математические пакеты расширения MATLAB. Специальный справочник. – Санкт-Петербург: Питер, 2001. – С. 307 – 309.

4. Ротштейн А.П. Влияние методов дефаззификации на скорость настройки нечеткой модели / А.П. Ротштейн, С.Д. Штовба // Кибернетика и системный анализ. – 2002. – № 5. – С. 169 – 174.

5. Трунова О.В. Алгоритм оцінки успішності засвоєння навчальної дисципліни студентом на основі теорії нечітких множин / О.В. Трунова, А.Г. Гребінник, І.С. Скітер // Математичне та імітаційне моделювання систем. МОДС 2014: Дев'ята міжнар. наук.-практ. конф. Тези доповідей (Чернігів-Жукін, 23–27 червня 2014 р.). – Чернігів: Чернігівський державний технологічний університет, 2014. – С. 296 – 299.

Стаття надійшла до редакції 07.06.2016