

## ОЦІНЮВАННЯ ЗАВДАНЬ ВІДКРИТОЇ ФОРМИ З РОЗГОРНУТОЮ ВІДПОВІДДЮ ПІД ЧАС КОМП'ЮТЕРНОГО ТЕСТУВАННЯ

© Шкіль О.С., Кучеренко Д.Ю., 2012

Розглянуто питання перевірки ІКТ-компетентності претендентів під час вступу до магістратури (аспірантури). Як тестові завдання під час комп'ютерного тестування запропоновано використовувати завдання з розгорнутою відповіддю. Оцінювання виконання таких завдань ґрунтується на експертному оцінюванні за обраними критеріями з використанням правил та процедур нечіткої логіки. Запропоновано методикку складання продукційних правил нечіткого виводу. Розроблено інструментальні засоби експертного оцінювання та порівняно результати їх роботи з наявними системами нечіткого виведення.

**Ключові слова:** компетентність, тестові завдання, експертне оцінювання, нечітка логіка, продукційні правила.

**The question of the ICT-competence test of the applicants for admission to the graduate or postgraduate study is considered. During computer testing the tasks with a detailed answer is proposed as test tasks. The assessment of such tasks performance is based on the expert review by selected criteria. The tools of the expert review are developed and the results of their work are compared with existing systems of the fuzzy inference.**

**Key words:** competence, test tasks, expert review, fuzzy logic, condition-action rules.

### Вступ

Комп'ютерне тестування як особливий інноваційний вид проведення контрольних заходів використовується не тільки для контролю, перевірки та самоперевірки знань студентів, але й для відбору, рейтингування та сертифікації (відбір для вступу до ВНЗ, розподіл за кваліфікаційними рівнями, видача професійних сертифікатів тощо). І саме об'єктивна оцінка результатів тестування претендентів залишається актуальною проблемою. Високий рівень суб'єктивізму під час оцінювання обумовив різке підвищення інтересу до тестових форм контролю. У вищій школі існує необхідність тестувати не тільки знання, вміння та навички (ЗВН), але й здатність застосовувати їх на практиці. Сукупність знань, вмінь, навичок та способів їх застосування, які дають змогу особистості адекватно сприймати та обробляти інформацію у предметній галузі, визначається як компетентність. Саме рівень компетентності претендентів має оцінюватися зокрема на іспитах з обраної спеціальності під час вступу до магістратури (аспірантури).

Для оцінювання компетентності претендентів необхідно підготувати і провести діагностичний експеримент із застосуванням знакової системи (тестів, кваліфікаційних завдань) і технічних засобів (комп'ютерів, тренажерів). Під час таких випробувань основною формою кваліфікаційних завдань (КЗ) є завдання відкритої форми з розгорнутою відповіддю, а спосіб оцінювання – експертне оцінювання. Незважаючи на те, що подібні завдання використовуються навіть у зовнішньому незалежному оцінюванні (ЗНО) в Україні, їх оцінювання є непрозорим та суб'єктивним. Відповідно до прийнятої у комп'ютерному тестуванні концепції оцінювання для завдань відкритої форми з розгорнутою відповіддю оцінювання має зводитися до визначення “відсотка правильності відповіді” [1]. Це необхідно враховувати вже на етапі формулювання

завдань шляхом виділення відповідних критеріїв оцінювання та визначення діапазонів оцінок визначених критеріїв. Введення системи критеріїв є основою для роботи експерта (експертної системи) під час оцінювання таких завдань. При цьому експерти, як правило, оперують поняттями природної мови, що також більш зрозуміло і самому випробуваному. З погляду випробуваного нечітка оцінка його компетентності у вигляді рівнів “високий”, “достатній”, “низький” і т.д. є зрозумілішою, ніж чітка кількість балів, які він набрав у результаті випробування. Саме це і зумовило використання нечіткої логіки як математичного апарату під час експертного оцінювання результатів виконання таких завдань. Отже, **метою роботи** є формалізація процедур експертного оцінювання завдань із розгорнутою відповіддю з використанням правил та процедур нечіткої логіки, та створення інструментальних засобів для експертного оцінювання.

### Експертне оцінювання

У складі комп’ютерної системи тестування знань OpenTEST2 створено модуль експертного оцінювання (рис. 1), який дає змогу залучати оцінки експертів для завдань з розгорнутою відповіддю під час випробувань у формі комп’ютерного тестування та формувати єдину оцінку незалежно від форми тестових завдань.



Рис. 1. Структура системи OpenTEST2 з модулем експертного оцінювання

Експертне оцінювання ґрунтується на системі нечіткого виведення. Найвні алгоритми нечіткого виведення розрізняються за видом правил використання логічних операцій і способом дефазифікації, а вибір моделі визначається, як правило, характером вирішуваних завдань. Необхідність отримання підсумкової оцінки на основі експертних оцінок окремих критеріїв обумовила вибір алгоритму нечіткого виведення Мамдані [2].

Для експертного оцінювання відповідей випробуваного введено вхідну лінгвістичну змінну (ЛЗ) <ОЦІНКА КРИТЕРІЮ>, яка містить три терми {“низький” (Н), “середній” (С), “високий” (В)}. Як функція приналежності використовується стандартна симетрична гауссова функція

$$\mu(x) = e^{-\frac{(x-m)^2}{2s^2}}, \text{ де } m - \text{ середнє значення діапазону, } s - \text{ стандартне відхилення.}$$

Для оцінювання рівня компетентності претендента під час експертного оцінювання введено вихідну лінгвістичну змінну <ОЦІНКА ЗАВДАННЯ>, і для неї визначено п'ятирівневу градацію ступенів правильності виконання завдання (п'ять термів) {“дуже низький” (ДН), “низький” (Н), “середній” (С), “достатній” (Д), “високий” (В)}. Діапазон значень ЛЗ <ОЦІНКА ЗАВДАННЯ>  $E = \{0 - 100\}$  балів. Якщо прийняти, що для ЛЗ <ОЦІНКА ЗАВДАННЯ> діапазон значень  $E = \{0 - 100\}$  є 100-бальною шкалою оцінювання, а терми – діапазонами шкали оцінювання кожного критерію, то межі діапазонів (термів) підпорядковуються “центральної тенденції”, тобто “ширина” діапазонів термів різна і звужується до центру шкали оцінювання. Параметри функцій приналежності вказаних лінгвістичних змінних, а також межі їх діапазонів наведено на рис. 2.

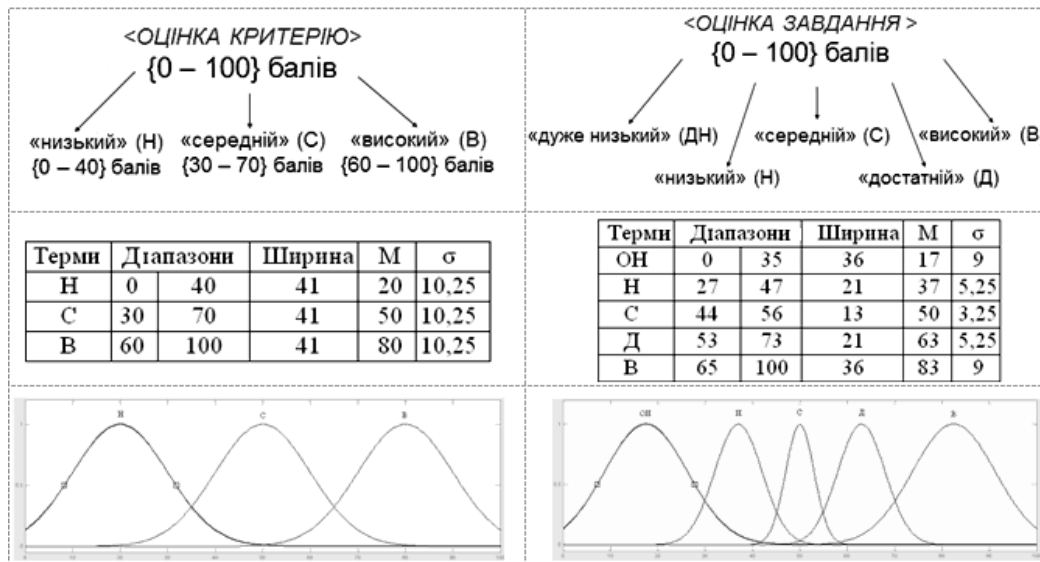


Рис. 2. Вхідні та вихідні лінгвістичні змінні, їх терми та функції приналежності

Згідно з алгоритмом Мамдані, продукційні правила (ПП) у формі диз'юнкцій значень термів ЛЗ мають вигляд:

$$\bigcup_{p=1}^{k_j} \left( \bigcap_{i=1}^n \{x_i = a_{i,jp}\} \right) \rightarrow y = d_j, j = \overline{1, m}, \quad (1)$$

де  $a_{1,jp}$  – лінгвістичний терм, яким оцінюється змінна  $x_i$  у рядку з номером  $jp$  ( $p = \overline{1, k_j}, j = \overline{1, m}$ );  $k_j$  – кількість рядків–кон'юнкцій, у яких вихід у оцінюється лінгвістичним термом  $d_j$ ;  $m$  – кількість термів, що використовуються для оцінювання лінгвістичної вихідної змінної. Під час складання ПП найважливіше завдання – зробити ці правила повними і несуперечливими одне одному. Загалом для заданих умов експертного оцінювання завдання (кількості критеріїв оцінювання та термів вхідної ЛЗ) існують  $243$  ( $3^5 = 243$ ) продукційні правила.

### Методика створення правил нечіткого виведення

Модель продукційних правил можна подати у класичній формі, що робить запис вказаних правил наочним і придатним для використання в стандартному пакеті нечіткого виведення:

$$\begin{aligned} & \{x_1 = a_{1,j1}\} \text{ I } \{x_2 = a_{2,j1}\} \text{ I } \dots \text{ I } \{x_n = a_{n,j1}\} \quad \text{АБО} \\ \text{ЯКЩО} & \{x_1 = a_{1,j2}\} \text{ I } \{x_2 = a_{2,j2}\} \text{ I } \dots \text{ I } \{x_n = a_{n,j2}\} \quad \text{АБО} \\ & \text{LLLLLLLLLLLLLLLLLLLLLLLLLLLL} \quad \text{АБО} \\ & \{x_1 = a_{1,jk_j}\} \text{ I } \{x_2 = a_{2,jk_j}\} \text{ I } \dots \text{ I } \{x_n = a_{n,jk_j}\} \\ \text{ТО} & \quad y = d_j, \quad j = \overline{1, m}. \end{aligned} \quad (2)$$

Для формалізації складання ПП скористаємося принципом призначення ваги кожному терму вхідної ЛЗ <ОК> [2]. Зауважимо, що вага кожного з критеріїв у підсумковій оцінці при цьому має бути однаковою. Для кожного з п'яти термів вихідної ЛЗ <ОЗ> визначимо ваговий діапазон у 2 бали ( $(15 - 5) / 5 = 2$ ), а з урахуванням перекриття меж нечітких діапазонів ваговий діапазон для кожного терма ваговий діапазон буде три бали (табл. 1).

Діапазони та набір кон'юнкцій для термів вихідної ЛЗ &lt;OЗ&gt;

Ваговий діапазон	Терм	Набір кон'юнкцій	Кількість кон'юнкцій
(5 - 7)	ДН	ННННХ	5
(7 - 9)	Н	ННССХ	30
		НННВХ	20
(9-11)	С	ССССХ	5
		ССВНХ	60
		ВВННХ	30
(11 - 13)	Д	ВВВНХ	30
		ВВССХ	20
(13 - 15)	В	ВВВВХ	5

У цій моделі кожен вектор (варіант моделі) містить множину кон'юнкцій, обумовлену поєднанням експертних оцінок за різними критеріями. Кількість кон'юнкцій обчислюється як  $\frac{n!}{R_B!R_C!R_H!}$ , де  $n$  – число змінних,  $R_H, R_C, R_B$  – число значень (букв) Н, С та В, які повторюються в кожному термі. Наприклад, для вектора ННННХ число змінних  $n = 5$ , число букв, які повторюються,  $R_H = 4, R_C = 0, R_B = 0$ , тому число різних кон'юнкцій буде  $\frac{5!}{4!0!0!} = 5$ . Отже, для ЛЗ <OЗ> терм ДН  $\rightarrow$  {ННННХ} можна подати у формі (2) так:

$$\begin{aligned} & \{OK_1 = H\} I \{OK_2 = H\} I \{OK_3 = H\} I \{OK_4 = H\} \text{ АБО} \\ & \{OK_1 = H\} I \{OK_2 = H\} I \{OK_3 = H\} I \{OK_5 = H\} \text{ АБО} \\ \text{ЯКЩО} & \{OK_1 = H\} I \{OK_2 = H\} I \{OK_4 = H\} I \{OK_5 = H\} \text{ АБО} \\ & \{OK_1 = H\} I \{OK_3 = H\} I \{OK_4 = H\} I \{OK_5 = H\} \text{ АБО} \\ & \{OK_2 = H\} I \{OK_3 = H\} I \{OK_4 = H\} I \{OK_5 = H\} \\ \text{ТО} & \quad OЗ = ДН, \end{aligned}$$

де  $OK_1, \dots, OK_5$  – оцінки експертів за п'ятьма критеріями відповідно;  $OЗ$  – оцінка завдання.

Кількість продукційних правил для п'яти критеріїв буде 205 за загальної кількості правил 243. Слід зазначити, що розроблену модель побудови продукційних правил можна використати для будь-якого типу критеріїв та завдань, що говорить про її універсальність та зводить до мінімуму суб'єктивність експерта при складанні бази продукційних правил.

Всі нечіткі підмножини, отримані для вихідної змінної ( $y$  всіх правилах), об'єднуються разом, що формує одну нечітку підмножину для кожної вихідної змінної (етап агрегації або композиції). За такого об'єднання можна використовувати два види операцій: максимум (max) та гранична сума (sum), правила виконання яких наведено в табл. 2. Проаналізувавши два види об'єднання нечітких множин, зроблено висновок, що як метод композиції слід обрати метод граничної суми, який, своєю чергою, дає змогу враховувати вплив кожного продукційного правила на вихідну нечітку множину.

Таблиця 2

## Правила виконання операцій агрегації

Операція max (максимум)	Операція sum (гранична сума)
$m_{\Sigma}(y) = \bigcup_{i=1}^M m_{B_i}(y) = \max_{i=1}^M (m_{B_i}(y)),$	$m_{\Sigma}(y) = \bigoplus_{i=1}^M m_{B_i}(y) = \min((\sum_{i=1}^M m_{B_i}(y)), 1),$
де « $\cup$ » – операція логічного максимуму (max).	де « $\oplus$ » – операція граничної суми (sum).

На рис. 3 наведено наочну структуру системи нечіткого виведення, яка використовується для оцінювання кваліфікаційних завдань. На рисунку зображено 5 вхідних лінгвістичних змінних

(оцінка критерію  $\langle OK \rangle$ ), кожна з яких відповідає критеріям оцінювання  $OK_1, \dots, OK_5$  і має 3 терми (Н, С, В), вихідна лінгвістична змінна (оцінка завдання  $\langle O3 \rangle$ ), яка має п'ять термів (ДН, Н, С, Д, В). Як було розраховано вище, загальна кількість продукційних правил – 205.

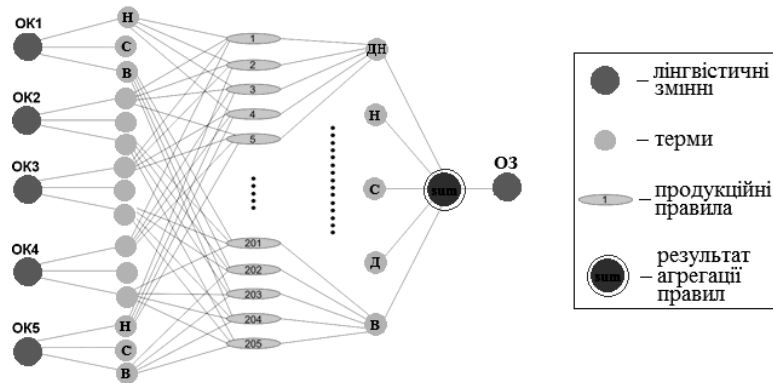


Рис. 3. Система нечіткого виводу для оцінювання кваліфікаційного завдання

### Реалізація методики експертного оцінювання завдань

Розроблену методику оцінювання завдань з розгорнутою відповіддю реалізовано у програмному модулі FuzzyLogic.exe.

Структуру та інтерфейс інструментальних засобів нечіткого виведення наведено на рис. 4.

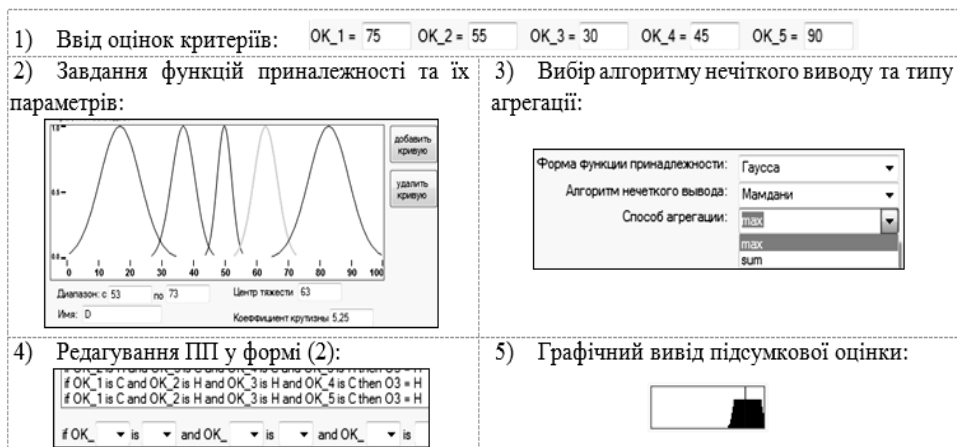


Рис. 4. Структура інструментальних засобів нечіткого виведення модуля FuzzyLogic.exe

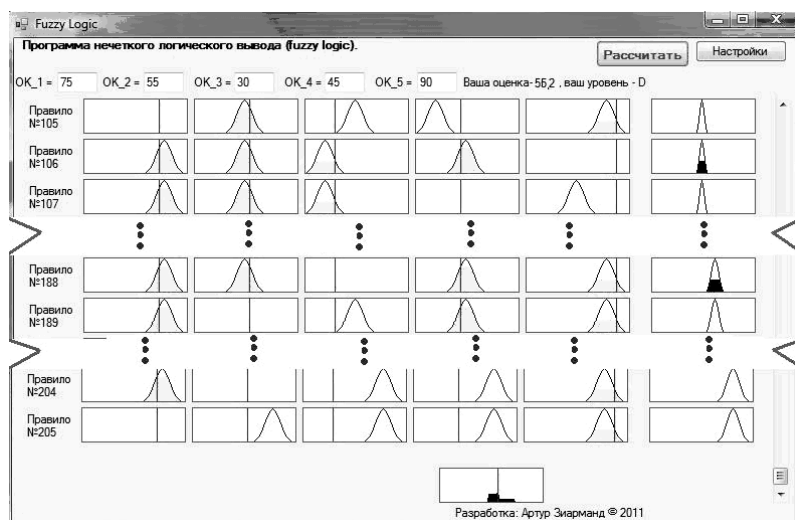


Рис. 5. Візуальне представлення результатів нечіткого логічного виводу

В основу роботи програми було покладено алгоритм нечіткого виведення Мамдані. Параметри та вид функцій приналежності обрано відповідно до правил, наведених на рис. 2. Візуально результати роботи програми для п'яти критеріїв оцінювання та способу агрегації продукційних правил за методом *sum* наведено на рис. 5.

### Приклад експертного оцінювання

На іспиті під час вступу до магістратури за спеціальністю “Комп’ютерні системи та мережі” претенденту було запропоновано таке завдання: “Запропонувати проект локальної офісної мережі та обґрунтувати вибір апаратного та програмного забезпечення для офісу на 20 робочих місць, розташованих в одному приміщенні, для роботи з програмами MS Office та бухгалтерською програмою 1С”. Зазначене завдання спрямоване на перевірку наявності у претендента компетентності у галузі інформаційно-комуникаційних технологій (ІКТ-компетентності).

Для аналізу ІКТ-компетентності застосовано тривимірну модель, яка складається з предметного виміру (що потрібно знати), діяльнісного виміру (що потрібно уміти) і практичного виміру (як використовувати набуті знання і уміння під час проектування та експлуатації комп’ютерних систем) [3]. Запропоновано таку змістовну структуру категорій предметного виміру ІКТ-компетентності: інформаційна компетентність, технічна компетентність, технологічна компетентність, алгоритмічна компетентність, мережева та телекомунікаційна компетентність, методична та дослідницька компетентність і компетентність у питаннях інформаційної безпеки. Діяльнісний вимір ІКТ-компетентності складається з алгоритмічних, технологічних, технічних та дослідницьких умінь. Практичний вимір ІКТ-компетентності, своєю чергою, складається з комунікативних застосувань, застосувань у комплектуванні комп’ютерних систем (КС), експлуатаційних застосувань та діагностичних застосувань.

Згідно з запропонованою моделлю ІКТ-компетентності з погляду предметного виміру це завдання перевіряє мережеву та телекомунікаційну компетентність, з погляду діяльнісного виміру – наявність технічних та технологічних умінь, а з погляду практичного виміру – компетентність у галузі комунікативних застосувань, а також застосувань у комплектуванні та експлуатації комп’ютерних мереж. Отже, критерії оцінювання виконання завдання мають відповідати переліченим розділам моделі ІКТ-компетентності. Тобто, експерти разом із автором завдання запропонували такі критерії оцінювання:

- 1) працездатність моделі розробленої мережі: *OK1*;
- 2) організація одночасної роботи з базою даних бухгалтерської системи 1С: *OK2*;
- 3) обґрунтування до вибору топології мережі, апаратного та програмного забезпечення: *OK3*;
- 4) простота реалізації (мінімізація витрачених ресурсів без шкоди для вирішення поставленої задачі): *OK4*;
- 5) можливість масштабування моделі мережі та зручність обслуговування: *OK5*.

Запропонована система критеріїв є основою для роботи експерта (експертної системи) під час оцінювання виконання зазначеного завдання.

За результатами виконання завдання претендент отримав від комісії такі оцінки:

*OK1* – висока оцінка (75 балів); *OK2* – середня оцінка (55 балів); *OK3* – низька оцінка (30 балів); *OK4* – середня оцінка (45 балів); *OK5* – висока оцінка (90 балів).

Результати роботи розробленої локальної версії програми FuzzyLogic.exe було порівняно з результатами, отриманими з використанням спеціального пакета нечіткого виведення Fuzzy Logic Toolbox математичної системи Matlab 7.5. На рис. 6 наведено результати роботи цих програм, а саме, вихідну нечітку множину для однієї і тієї самої моделі ПП за однакових вхідних значень оцінок критеріїв (75, 55, 30, 45, 90), але за різними методами агрегації продукційних правил.

Проаналізувавши результати оцінювання, можна зробити такі висновки:

- що більший внесок у активацію окремих ПП, то більша різниця між результатами дефазифікації за різних методів агрегації, що підтверджує переваги використання під час агрегації ПП методу граничної суми;

- різниця у 1,8 бала (за методом MAX) и 2,1 бала (за методом SUM) для різних програм обмовлена різницею у методиках реалізації алгоритмів агрегації, а також точністю округлення числових значень у розрахунках під час агрегації та дефазифікації.





<i>Fuzzy Logic Toolbox MATLAB</i>	<i>Fuzzy Logic OpenTEST</i>
Max:  56.2	Max:  58
Sum:  57.1	Sum:  55
Результати дефазифікації: за методом max – 56,2 бала, за методом sum – 57,1 бала.	Результати дефазифікації: за методом max – 58 бала, за методом sum – 55 бала.
Випробовуваний має <b>достатній</b> (Д) рівень компетентності.	

Рис. 6. Порівняння результатів дефазифікації за різних методів агрегації правил у розробленій програмі з пакетом FLT системи MATLAB

Загалом різниця між результатами роботи розробленої програми FuzzyLogic.exe та пакета Fuzzy Logic Toolbox незначна, вихідний діапазон для двох оцінок збігається (достатній рівень компетентності), тобто різниця (похибка) загалом не впливає на результат оцінювання.

### Висновки

У роботі запропоновано методику перевірки ІКТ-компетентності під час вступу до магістратури (аспірантури). В основу системи оцінювання покладено використання кваліфікаційних завдань відкритої форми з розгорнутою відповіддю. Як метод формалізації оцінювання таких завдань запропоновано використовувати правила та процедури нечіткої логіки, які реалізовано в алгоритмі нечіткого виведення Мамдані. Формалізація процедури складання правил нечіткого виведення дає змогу запобігти помилкам під час створення бази знань продукційних правил. Запропоновані процедури реалізовано у вигляді окремого програмного модуля FuzzyLogic.exe з можливістю його локального використання або введення до комп'ютерної системи тестування OpenTEST2.

1. Шкіль О.С. Єдиний тестовий сервер вищого навчального закладу як основа якісного проведення тестувань / О.С. Шкіль // Інформатизація вищого навчального закладу. Вісник Нац. ун-ту “Львівська політехніка” – 2011. – №703. – С. 54–59. 2. Кривуля Г.Ф. Нечеткая логика в экспертной оценке ИКТ-компетентностей / Г.Ф. Кривуля, А.С. Шкіль, Д.Е. Кучеренко, Е.В. Гаркуша // Вестник ХНТУ. – 2011. – №2 (41). – С. 13–22. 3. Кривуля Г.Ф. Процедуры диагностирования компетентностей пользователей компьютерных систем / Г.Ф. Кривуля, А.С. Шкіль, Д.Е. Кучеренко, Е.В. Гаркуша // Інформаційно-керуючі системи на залізничному транспорті. – 2011. – №1. – С. 68–75.