

## **Розробка методів проведення тестування й оцінки їхніх результатів з використанням сучасних інформаційних технологій**

*Проведено аналіз сучасних мережних систем тестування. Розглянуто три методи проведення тестування: оцінювальний, адаптивний й екстремальний і формалізовані моделі оцінки знань для кожного із цих методів. Розроблено програмне забезпечення системи тестування знань.*

*The analysis of modern network systems testing. Three methods of testing: intelligence, adaptive and extreme formal models and knowledge assessment for each of these methods. A software testing knowledge.*

**Ключові слова:** *тест, тестування, знання, інформаційні технології.*

**Вступ.** Процес одержання інформації нерозривно пов'язаний з необхідністю перевірки якості її засвоєння. Тестування знань стало найважливішим компонентом сучасних освітніх технологій. Це обумовлено тим, що грамотно побудований тест дає об'єктивну оцінку знань й є високопродуктивним інструментом контролю.

**Постановка завдання.** На сьогодні існує велика кількість різноманітного доробку для проведення перевірки рівня знань суб'єкта, що проходить тестування. Основними їхніми недоліками є: зв'язаність із конкретним матеріалом і неможливість застосування в інших сферах, що не дозволяє стандартизувати формати й форми зберігання даних; обмеженість схем тестування або навпаки; погано продуманий інтерфейс; складність ряду програм в освоєнні; специфічний формат зберігання інформації, що викликає необхідність повторного уведення тестових даних.

Крім того, більшість систем використовують найпростіші методи проведення тестування й оцінки знань, що робить актуальним питання про розробку більш складних й інформативних методів, що дозволяють підвищити надійність й об'єктивність результатів контролю.

Під тестом будемо розуміти інструмент, що складається із системи стандартизованих завдань, стандартизованої процедури проведення й задалегідь спроектованої технології обробки й аналізу результатів [1].

Одна із вагомих вимог до тесту, що відрізняє його від іспиту й від інших методів контролю, - це обов'язкова перевірка його якості. Існують відповідні науково обґрунтовані критерії, які дозволяють оцінити якість тесту: об'єктивність, надійність, валідність, складність, ефективність [2].

Методи проведення контролю знань можна розділити на три класи: неадаптивні методи; частково адаптивні методи; повністю адаптивні методи:

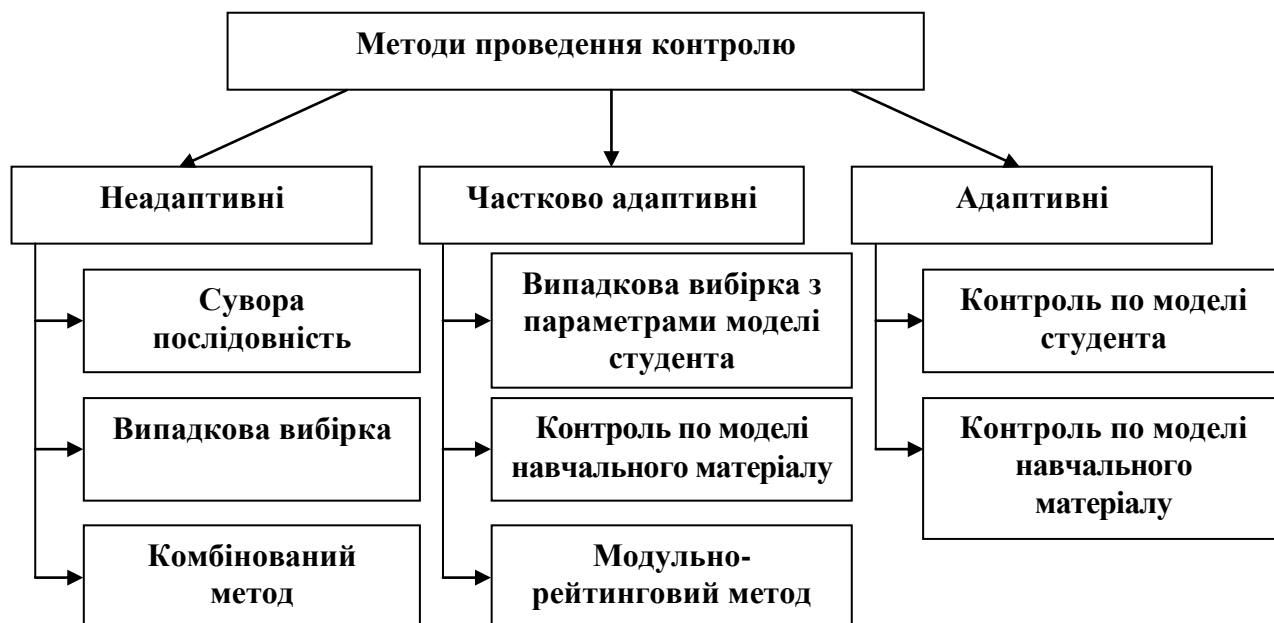


Рис. 1. Методи проведення контролю знань

Що стосується методів оцінювання, то виділяють наступні: найпростіша модель; модель, що враховує час виконання завдання й/або загальний час контрольної роботи; модель на основі рівнів засвоєння й так далі.

**Результати.** Розглянемо наступні методи проведення тестування: «оцінювальний», «адаптивний» й «екстремальний».

«Оцінювальне» тестування застосовується в тих випадках, коли необхідно довідатися рівень знань суб'єкту, що тестується, по деяким темам у певній області знань. При проходженні тестування питання вибираються випадковим чином з деякого набору. Це дозволяє для кожного суб'єкту, що тестується, використовувати свій окремий варіант тесту. Застосовується для оцінки знань, при самоперевірці, для проміжного контролю й т.д.

«Адаптивне» тестування відрізняється від «оцінювального» тем, що враховує рівень знань суб'єкту, що тестується, й залежно від цього вибирає питання, тобто адаптується під нього. У якості першого береться питання

середньої складності. Потім залежно від правильності отриманої відповіді, суб'єкту, що тестується, пропонується більш складне або більш легке питання. Даний метод проведення тестування має деякі переваги в порівнянні з «оцінювальним». «Адаптивний» метод дозволяє більш чітко розмежовувати суб'єкти, що тестується, по їхньому рівню знань теми, тому що суб'єкти з вищим рівнем підготовки будуть відповідати на більш складні питання й у підсумку одержувати більш високу оцінку в порівнянні з тими, хто буде відповідати на легкі питання. Крім того, даний метод дозволяє до деякої міри моделювати «живого» екзаменатора, дозволяючи вести тестування, динамічно змінюючи хід процесу. Наприклад, тестування може бути перервано, коли випробуваний досить довго відповідав на питання високого рівня складності, а також можуть бути задані більше легкі питання, якщо суб'єкт, що тестується, зазнає труднощі що до відповіді на запитання з даним рівнем складності. Таким чином, процес тестування з використанням даного методу для кожного суб'єкту, що тестується, буде індивідуальним, і залежати від рівня знань у даній області.

*«Екстремальне»* тестування використовується для визначення рівня реакції суб'єкту, що тестується (швидкості відповіді на поставлені питання). Він може застосовуватися для визначення швидкості дій при аваріях, катастрофах, інших надзвичайних ситуаціях або для визначення, як швидко може суб'єкт, що тестується, зреагувати на ситуацію й прийняти правильні рішення. При наявності надлишкового часу, у суб'єкта, що тестується, з'являється можливість скористатися додатковими джерелами, а це у свою чергу дає погіршеність в оцінці його дійсних знань. При перебільшенні часу відповіді суб'єкту, що тестується, на поставлене питання характеристика правильності відповіді знижується, тому що це можливо розглядати, як використання підказки.

Основний недолік «оцінювального» й «екстремального» методів – варіант тесту генерується випадково без урахування складності питань, що потрапили в нього. І може так трапитися, що набір питань для одного суб'єкту, що тестується, буде включати лише складні запитання, а для іншого - тільки легкі. Це часто приводить до неадекватності результатів контролю знань.

Для того щоб обійти цей недолік, при складанні тесту пропонується враховувати його загальну складність. У цьому випадку тест буде доступний

для проходження, якщо його складність не нижче певного числа, що задає як один з параметрів тесту.

Розглянемо математичну модель оцінки знань для даних методів. Нехай експерт склав тест. Тест складається з  $p$  тем. У цілому тест містить  $n$  питань. Суб'єкту, що тестується, необхідно відповісти на  $m$  питань,  $m \leq n$ .

Вагомість тем у тесті обчислюється за допомогою методу попарних порівнянь тем між собою і являє собою множину  $W = \{w_j\}$ ,  $j = \overline{1, p}$ .

Складність питань задається явно експертом й являє собою множину  $V = \{v_{ij}\}$ ,  $i = \overline{1, n}$ ,  $j \in \overline{1, p}$  і показує який з  $p$  тем належить питання, область визначення  $v_{ij}$  – інтервал  $(0, 1]$ .

Множину відповідей на  $i$ -е питання позначимо через  $R = \{r_{iq}\}$ ,  $i = \overline{1, n}$ ,  $q = \overline{1, z_i}$ , де  $z_i \geq 2$  – кількість запропонованих відповідей в  $i$ -ому питанні. Для кожної відповіді експертом явно задається ступінь правильності відповіді, що належить відрізьку  $[-1, 1]$ .

Розглянемо схему «оцінювального» тестування. Нехай тест складається з  $m$  питань, обраних випадковим чином з  $n$  можливих. У цей тест потрапили питання з деякої підмножини тем й їхньої вага згідно предметної області задається множиною  $W' = \{w'_j : w'_j \in W, j \in J'\}$ , тут  $J'$  – множина індексів тем у одержаному тесті. Складність питань, що потрапили в тест, позначимо через множину  $V' = \{v'_{ij} : v'_{ij} \in V, i = \overline{1, m}, j \in J'\}$ . Множину отриманих відповідей позначимо через  $\tilde{R} = \{\tilde{r}_i : \tilde{r}_i \in R, i = \overline{1, m}\}$ .

Для початку необхідно обчислити складність одержаного тесту. Вона обчислюється по формулі:

$$T = \sum_{\substack{i=1 \\ j \in J'}}^m v'_{ij} w'_j, \quad (1)$$

де  $v'_{ij}$  – складність питання,  $w'_j$  – важливість теми, який належить дане питання.

Для проходження може надаватися тест, складність якого не менше деякого  $T_0$ , котре задається як один з параметрів тесту експертом.  $T_0$  задається у вигляді нечіткого значення: дуже слабкий, слабкий, середній, складний, дуже складний. Для використання цього параметра в розрахунках, його необхідно перевести в числовий вигляд. Інтервал, якому може належати  $T_0$ , обчислюється

в такий спосіб: необхідно знайти мінімально й максимально можливе значення складності тесту (при випадковій вибірці  $m$  питань із  $n$ ). Для цього обчислюються всі можливі  $s_i = v_{ij}w_j$ ,  $i = \overline{1, n}$ ,  $j \in \overline{1, p}$  і впорядковуються по зростанню, тоді  $s_{\min}$  рівняється сумі перших  $m$   $s_i$ , а  $s_{\max}$  – сумі останніх  $m$ :

$$s_{\min} = \sum_{i=1}^m s_i, \quad (2)$$

$$s_{\max} = \sum_{i=n-m+1}^n s_i. \quad (3)$$

Обчислюємо кватиль – одну четверту довжини відрізка  $[s_{\min}, s_{\max}]$ :

$$kv = \frac{1}{4}(s_{\max} - s_{\min}). \quad (4)$$

Остаточний відрізок, якому може належати  $T_0$ :

$$T_0 \in [s_{\min}, s_{\max} - kv]. \quad (5)$$

Далі обчислюється  $d$  – числовий аналог параметра  $T_0$ , якому рівняється  $T$ . Наприклад:  $d = \{(0 - \text{дуже слабкий}), (0.25 - \text{слабкий}), (0.5 - \text{середній}), (0.75 - \text{складний}), (0.99 - \text{дуже складний})\}$ .

Тоді остаточне  $T_1$  обчислюється по формулі:

$$T_1 = s_{\min} + (s_{\max} - kv - s_{\min}) * d. \quad (6)$$

Тест допускається для проходження, якщо  $T \geq T_1$ .

Оцінка відповіді суб'єкту, що тестується:

$$O = \frac{\sum_{\substack{i=1 \\ j \in J'}}^m \tilde{r}_i v'_{ij} w'_j}{T}, \quad (7)$$

де  $T$  обчислюється по формулі (1) на основі сформованого тесту, на який відповідав суб'єкт, що тестується,  $\tilde{r}_i$  – ступінь правильності відповіді. Область визначення  $O$  – відрізок  $[-1, 1]$ . Далі це число можна перевести в п'ятибальну систему, у відсотки й так далі (залежно від прийнятої шкали оцінювання).

Розглянемо схему адаптивного тестування. Нехай  $n$  – загальна кількість питань, наявних у тесті,  $m$  – кількість питань, на які необхідно відповісти. Позначимо  $s_i = v_{ij}w_j$ ,  $i = \overline{1, n}$ ,  $j \in \overline{1, p}$ , де  $v_{ij}$  – складність питання,  $w_j$  – важливість теми, якій належить питання. Нехай  $l$  – кількість рівнів складності, один з параметрів тесту, задається явно експертом й  $l \geq 3$ .

Упорядковуємо  $s_i$  по зростанню й знаходимо  $s_{\min}$  й  $s_{\max}$  по формулах (2) і (3).

Обчислюємо довжину кожного з рівнів складності:

$$in = \frac{s_{\max} - s_{\min}}{l}. \quad (8)$$

Починаємо вибирати питання із середнього інтервалу, нехай його індекс рівняється  $h \in [1, l]$ .

На кожному рівні задається  $k$  питань. У даній моделі розглядається  $k = 3$ .

У результаті відповідей суб'єкту, що тестується, на кожному етапі одержуємо набір  $\tilde{R}^u = \{\tilde{r}_f^u : \tilde{r}_f^u \in R, f = \overline{1, k}\}$ , – ступень «правильності» відповідей, причому  $\tilde{r}_f^u \in [-1, 1]$ ,  $f = \overline{1, k}$ . Індекс  $u$  визначає порядковий номер етапу. Таким чином,  $\sum_{f=1}^k \tilde{r}_f^u \in [-k, k]$ .

Позначимо  $b^u = \sum_{f=1}^k \tilde{r}_f^u$ . Якщо  $0.75k \leq b^u \leq k$  – переходимо на наступний рівень складності (збільшуємо індекс  $h$  на 1, якщо це можливо). Якщо  $0.5k \leq b^u < 0.75k$  – залишаємося на тому ж рівні складності (індекс  $h$  не міняється). Якщо  $-k \leq b^u < 0.5k$  – переходимо на рівень нижче, знижуючи складність питань (зменшуємо індекс  $h$  на 1, якщо це можливо). Якщо суб'єкт, що тестується, перебуває на самому складному (легкому) рівні, і судячи з його відповідей потрібно підвищити (понижити) рівень, то необхідно продовжувати задавати питання самого складного (легкого) рівня.

Таким чином, суб'єкту, що тестується, закінчує тест, якщо:

- він відповів на  $m$  питань тесту,
- він відповів на всі питання самого складного (легкого) рівня й питань більше не залишилося, а необхідно підвищувати (знижувати) рівень або залишатися на тому ж рівні.

Розглянемо схему екстремального тестування. Тест формується подібно оцінної схеми тестування. Питання вибираються випадково, але таким чином, щоб складність усього сформованого тесту була не менше складності, що задає експертом як один з параметрів тесту.

На загальну оцінку суб'єкту, що тестується, впливає правильність і швидкість відповіді, його реакція й увага. Час відповіді на питання визначається як загальна сума часу всіх відповідей даного питання.

У результаті проходження тесту, одержуємо множину  $\tilde{R} = \{\tilde{r}_i : \tilde{r}_i \in R, i = \overline{1, m}\}$  – відповіді суб'єкту, що тестується, час відповіді на питання –  $Time = \{t_i\}$ ,  $i = \overline{1, m}$ . Експертом явно задається  $Time^{\max} = \{t_i^{\max}\}$ ,  $i = \overline{1, m}$  – час, відведений для відповіді на  $i$ -е питання.

Загальний час, що відводиться на проходження тесту  $t = \sum_{i=1}^m t_i^{\max}$ . Тобто якщо суб'єкт, що тестується, затримався на одному питанні, то втратив час для відповіді на інше питання.

Розрахуємо коефіцієнти для оцінки знань суб'єкту, що тестується, залежно від часу відповіді на питання:

$$E_i = \begin{cases} 1, \tilde{r}_i \leq 0; \\ 1, \tilde{r}_i > 0, t_i \leq t_i^{\max}; \\ 1 - \frac{t_i - t_i^{\max}}{t_i^{\max}} = \frac{2t_i^{\max} - t_i}{t_i^{\max}}, \tilde{r}_i > 0, t_i^{\max} < t_i < 2t_i^{\max}; \\ 0, \tilde{r}_i > 0, t_i \geq 2t_i^{\max}. \end{cases} \quad (9)$$

Область визначення  $E_i \in [0, 1]$ .

Далі підсумкова оцінка обчислюється, як і для «оцінювального» методу, тільки враховується ще й коефіцієнти  $E_i$ :

$$O = \frac{\sum_{\substack{i=1 \\ j \in J'}}^m E_i \tilde{r}_i v'_{ij} w'_j}{T} \quad (10)$$

Розглянуті схеми проведення тестування й моделі оцінювання знань дозволяють досить гнучко здійснювати оцінювання й ранжування суб'єктів, що тестуються, залежно від їхнього рівня знання за наданим матеріалом.

У результаті досліджень було розроблено програмне забезпечення системи тестування знань. Клієнтська частина написана з використанням PHP, а для зберігання інформації про тести, користувачів й іншу супровідну інформацію обрана СУБД MySQL.

Програмне забезпечення підрозділяється на наступні підсистеми:

- 1) ведення довідника студентів, викладачів, тем, питань і відповідей на них;
- 2) підсистема формування й настроювання тестів;
- 3) підсистема проходження тестів;
- 4) підсистема видачі результатів і статистики.

На рис. 2 представлена головна сторінка сайту:



Рис. 2. Головна сторінка сайту

Інтерфейс редагування й настроювання одного з тестів наведений на рис.

3.



Рис. 3. Інформація з окремого тесту

Загальний вигляд вікна проходження тестування наведений на Рис. 4.



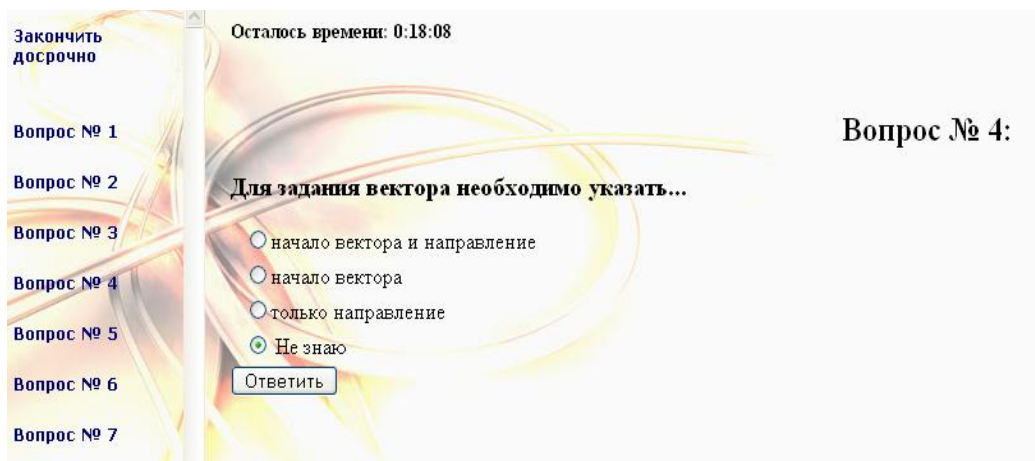


Рис 3. Проходження тесту

**Висновки.** Таким чином, розглянуті схеми проведення тестування й моделі оцінювання знань дозволяють досить гнучко здійснювати оцінювання й ранжування суб'єктів, що тестуються, залежно від їхнього рівня знання матеріалу. У підсумку був розроблений пакет програмного забезпечення, що надає великі можливості, але в той же час потребує мінімум дій з боку користувачів, яким необхідно пройти тестування, а також має досить потужні можливості для подальшої оптимізації, додавання нових функцій і розширення закладених можливостей.

### Література

1. Аванесов В.С. Композиция тестовых заданий. - М.: Адепт, 1998. 217с.
2. Анастаси А. Психологическое тестирование. СПб.: Питер, 2002. 688 с.
3. Зайцева Л.В. Некоторые аспекты контроля знаний в дистанционном обучении. – Образование и виртуальность, 2000. – 108 с.
4. Казаринов А.С., Култышева А.Ю., Мирошниченко А.А. Технология адаптивной валидности тестовых заданий: Учебное пособие. – Глазов, ГГПИ, 1999 г. – 62 с.
5. Нардюжев В.И., Нардюжев И.В. Модели и алгоритмы информационно-вычислительной системы компьютерного тестирования. – М.: Прометей, 2000. – 148 с.
6. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. – М.: Прометей, 2000. – 169 с.