

В. Ляшик, І. Шубін

МЕТОДИ АВТОМАТИЗАЦІЇ ТА ІНТЕРПРЕТАЦІЇ ЯКОСТІ ОСВІТНІХ ТЕСТІВ У СИСТЕМАХ РОЗПОДІЛЕНОГО ВІРТУАЛЬНОГО НАВЧАННЯ

Предметом дослідження є розроблення математичного та алгоритмічного забезпечення інтелектуального інструментарію аналізу наборів тестових завдань і моделювання процесу інтерпретації якості наборів тестових завдань, що дає змогу об'єктивно та всебічно проводити безперервний контроль знань студентів (суб'єктів навчання) за умови впровадження концепції віртуального розподіленого навчання. **Мета роботи** – створення методів автоматизації оцінювання якості освітніх тестів та інтерпретації результатів, подання математичного апарату, що допомагає якнайкраще описувати складні інтелектуальні процеси й методи оцінювання комплексів тестових завдань під час перевірки рівня знань суб'єктів навчання. У статті вирішуються такі **завдання**: формування моделі тестування в розподіленому віртуальному навчальному середовищі та моделі оцінювання валідності за змістом наборів тестових завдань. Використовуються такі **методи**: математична статистика й інтелектуальний аналіз інформації, алгебра скінченних предикатів і операцій. Здобуто такі **результати**: сформульовано принципи інтелектуального аналізу значення коефіцієнтів надійності, коефіцієнтів валідності, коефіцієнта дискримінативності; визначено індекс важкості завдань для оцінювання знань суб'єктів навчання. **Висновки**. Застосування методів формалізації оцінювання тестів підтверджують необхідність введення в практику освіти кількісних методів оцінювання знань студентів. Упровадження зазначених методів передбачає правильне визначення цілей контролю, предмета вимірювання та вибір засобів вимірювання. Застосування розробленого методу аналізу наборів тестових завдань сприяє ефективній реалізації всіх функцій контролю, відповідає основним його принципам для вирішення завдань оцінювання якості тестів і лежить в основі моделей розподіленого віртуального навчання та методів аналізу успішності суб'єктів навчання.

Ключові слова: програмна інженерія; бази знань; алгебра скінченних предикатів; надійність тестів; використання знань; логічні правила.

Вступ

У зв'язку з переходом на рейтингову систему контролю знань студентів, що є системою взаємопов'язаних правил і яка останнім часом все ширше використовується на різних рівнях освіти, зростають вимоги до якості педагогічного контролю. Актуальним стає розроблення педагогічного програмного інструментарію, що дає змогу об'єктивно та всебічно проводити безперервний контроль знань студентів. Використання рейтингової системи без тестування взагалі неможливе, оскільки саме на тестових оцінках ґрунтується ця система. Під адаптивним тестовим контролем розуміють комп'ютеризовану систему науково обґрунтованої перевірки й оцінювання результатів навчання, яка має високу ефективність завдяки оптимізації процедур генерації, подання й оцінювання результатів виконання адаптивних тестів і яка оснований на методах побудови та оптимізації логічних мереж. Алгоритми підбору й подання завдань будуються за принципом зворотного зв'язку, коли в разі правильної відповіді чергове завдання обирається важчим, а за умови неправильної

відповіді пропонується наступне завдання, легше, ніж попереднє, на яке учасник тестування не відповів. Також існує функція додаткових запитань із тем, які суб'єкт навчання знає не дуже добре. Це потрібно для більш точного з'ясування рівня знань у певній науковій галузі. Вибір алгоритмів тестування наразі фактично обмежений формами подання тестових завдань і оцінювання результатів. Досягнення більш високих результатів і підвищення мотивації навчання в остаточному підсумку є основною метою тестування.

Аналіз створення локальних інформаційно-освітніх середовищ дав змогу сформулювати основні принципи професійної дистанційної освіти: універсальність; інваріантність; масштабованість; відкритість; технологічність процесів створення, зберігання й використання ресурсів; спадкоємність тощо. Огляд гіпермедійних засобів адаптивного навчання показав, що це новий дослідницький напрям у сфері інженерії програмного забезпечення підтримки дистанційного навчання, завдання якого – додати до дистанційних навчальних систем можливості індивідуалізації. Аналітичний огляд проблем електронного навчання, зокрема

інтенсифікація, індивідуалізація, послідовність викладення матеріалів, адаптивність і гнучкість, дав змогу виокремити основні підходи до їх вирішення: формування індивідуальних траєкторій навчання, розроблення моделей і суб'єктів навчання, а також методів адаптивного навчання.

Контроль знань або тестування – процес, що проводиться для визначення рівня знань суб'єкта навчання [2]. Це найбільш стандартизований і об'єктивний метод контролю й оцінювання знань, умінь і навичок, який позбавлений традиційних недоліків інших методів контролю знань, наприклад, неоднорідності вимог, суб'єктивності екзаменаторів, невизначеності системи оцінювання тощо. Рівні знань зазвичай дискретизуються. За такого підходу тестування може розглядатися як деякий діагностичний процес, а стани, що визначають оцінки знань суб'єкта навчання, – як діагностичні стани. Тести є ефективним засобом перевірки якості знань, що набули студенти, і оперативного контролю етапів навчання [3].

У роботі запропоновано завдання підвищення ефективності інформаційно-аналітичного забезпечення підсистеми контролю знань суб'єктів освітнього процесу за умови побудови індивідуальної траєкторії навчання, що передбачає такі етапи: 1) розроблення концептуальної моделі електронного адаптивного навчання в розподіленому віртуальному середовищі; 2) створення моделі побудови індивідуальної освітньої траєкторії; 3) побудова математичної моделі навчального об'єкта; 4) розроблення моделі визначення тестових об'єктів перевірки засвоєння навчальних матеріалів відповідно до індивідуальних вимог користувача; 5) розроблення методу побудови адаптивної освітньої траєкторії; 6) створення прототипів програмних рішень для реалізації інформаційної технології побудови індивідуальних траєкторій навчання.

Для визначення базових алгоритмів необхідно навести алгебро-логічні моделі системи, в основі яких лежить модель приймання іспиту викладачем у студента (суб'єкта навчання) як модель адаптивного тестування. Такий вибір сценарію роботи системи зумовлений тим, що, по-перше, ця процедура історично добре формалізована, по-друге, у процесі проектування тестів розробнику необхідно спиратися на загальноприйняті, відомі та використовуваним методи з мінімальною модифікацією.

Аналіз проблеми й наявних методів

Виокремлюють такі функції педагогічного контролю: контролююча, навчальна (освітня), діагностична, прогностична, розвивальна, орієнтувальна, виховна. Вони, як правило, тісно взаємозв'язані [1]. У навчальному процесі ці функції виявляються в різному ступені та в неоднакових поєднаннях. Реалізація перелічених функцій робить контроль більш результативним, а також ефективнішим стає і сам освітній процес.

Навчальна функція контролю полягає в удосконаленні та систематизації знань і вмінь. У процесі перевірки студенти повторюють і закріплюють вивчений матеріал. Вони не тільки відтворюють раніше вивчене, але й застосовують знання та вміння в нових ситуаціях. Перевірка допомагає виокремити основні положення досліджуваного матеріалу, зробити знання та вміння більш чіткими й точними. Контроль сприяє також узагальненню та систематизації знань [2].

Сутність діагностичної функції полягає в отриманні інформації про помилки, недоліки й прогалини в знаннях і вміннях студентів, про причини, що породжують труднощі в опануванні навчального матеріалу, і про кількість, характер помилок. Результати діагностичних перевірок допомагають обрати найбільш інтенсивну методику, а також уточнити напрям подальшого вдосконалення змісту, методів і засобів навчання.

Прогностична функція слугує отриманню випереджальної інформації про навчально-виховний процес. За результатами перевірки здобувається прогноз про певний етап навчального процесу: чи достатньо сформовані конкретні знання, уміння й навички для засвоєння наступного навчального матеріалу (розділу, теми); розробляється модель подальшої поведінки студента.

Розвивальна функція контролю полягає в стимулюванні пізнавальної активності студентів, формуванні їхніх творчих здібностей. Контроль має виняткові можливості для розвитку студентів. У процесі контролю розвивається мова, пам'ять, увага, уява, воля та мислення молоді людини.

Сутність орієнтувальної функції контролю – отримати інформацію про ступінь досягнення мети навчання окремим студентом і групою загалом, тобто рівень вивчення навчального матеріалу. Розкриваючи прогалини, помилки й недоліки

в підготовленні студентів, визначаються напрями щодо вдосконалення знань і вмінь. Контроль допомагає суб'єкту навчання краще пізнати самого себе, оцінити свої знання й можливості [4].

Виховна функція контролю формує в студентів відповідальне ставлення до навчання, дисципліну, акуратність, чесність. Перевірка спонукає до більш серйозного й регулярного виконання завдань.

Система перевірки результатів навчання й виховання містить три основних види контролю: поточний, проміжний і підсумковий [5].

Для опису процесу вибору індивідуальної траєкторії навчання необхідні засоби інтелектуального оброблення інформації. Теорія інтелекту, засоби моделювання інтелектуальних функцій людини на основі методу компараторної ідентифікації та інструментарію алгебри скінченних предикатів, розроблені науковою школою професора Ю. Шабанова-Кушнарєнка, дають таку можливість та добре зарекомендували себе у вирішенні завдань інтелектуального оброблення інформації, зокрема текстового та візуального [6].

Ця мета в створенні методів і алгоритмів побудови віртуального розподіленого навчального середовища насамперед пов'язана з визначенням якості засвоєння студентами навчального матеріалу – рівня набуття знань, умінь і навичок, передбачених програмою. По-друге, конкретизація основної мети контролю пов'язана з навчанням прийомів взаємоконтролю та самоконтролю та формуванням потреби в них. По-третє, ця мета передбачає виховання в студентів таких особистісних якостей, як відповідальність за виконану роботу, вияв ініціативи.

Мета статті – створення методів автоматизації оцінювання якості освітніх тестів і впровадження цих методів у системи підтримки дистанційного навчання.

Вирішення завдання

Серед форм контролю знань у процесі впровадження віртуального навчального середовища все більшого розвитку набуває тестовий контроль, рекомендований Міністерством освіти і науки України та документами Болонської конвенції. Тестовий контроль у цих документах розглядається як один із засобів удосконалення процесу навчання. Актуальними є питання узагальнення, систематизації наявного досвіду й вироблення рекомендацій у сфері створення тестових завдань і проведення тестового контролю знань, умінь і навичок [6].

Для подання знань у інформаційних системах використовуються різні формальні мови, зокрема обчислення предикатів. Вона має однозначну формальну семантику й операційну підтримку у вигляді досконалого механізму виведення. Мовою алгебри предикатів будь-яка множина $U = \{a_1, a_2, \dots, a_n\}$ може бути записана у вигляді рівняння $x^{a_1} \vee x^{a_2} \vee \dots \vee x^{a_n} = 1$, де x – предметна змінна [7]. Сукупність усіх коренів цього рівняння збігається з множиною U . Рівняння є формальним записом твердження $x \in U$. Методи ідентифікації та подання знань із використанням алгебри предикатів можуть одночасно подати знання у вигляді рівнянь алгебри предикатів [9].

Алгебра предикатів описує лише знання про факти. Алгебра операцій над предикатами або алгебра предикатних операцій має формалізувати операції над знаннями, поданими як відношення на деякому предметному просторі. Алгебра предикатів визначає декларативний складник знань, а алгебра предикатних операцій – процедурний складник набуття знань, умінь і компетенцій суб'єктом навчання в умовах індивідуалізації освітнього процесу в розподіленому віртуальному середовищі [9].

Комп'ютерне тестування має свої переваги й недоліки. Саме об'єктивність отриманих оцінок сприяє розширенню сфери впровадження комп'ютерного тестування – воно успішно застосовується в школах, вишах, під час складання іспитів, співбесіди тощо.

До таких рішень належать: програма *NetTest* для комп'ютерного тестування знань у мережі [9]; *MyTestXPro* – система програм для створення та проведення комп'ютерного тестування, збору та аналізу їх результатів [10]; комп'ютерна система тестування знань *OpenTest* [11]; конструктор тестів *Keepsoft* – універсальної системи перевірки знань [12]; програма *x-TLS* [13]. Основні відмінності таких систем: централізований збір і аналіз результатів тестування за наявності мережі; виставлення оцінки із використанням довільної системи та шкали оцінювання; організація локального та мережного тестування; стеження за процесом тестування. Застосовується підтримка п'яти типів запитань; використання необмеженої кількості тем, запитань і відповідей; наявність запитань, що містять текст необмеженої довжини; застосування в запитаннях гіпермедійних і мультимедійних об'єктів; можливість декільком

користувачам проходити тестування на одному комп'ютері; збір і аналіз результатів за допомогою локальної мережі та синхронізації бази даних; копіювання/резервне копіювання та збереження й аналіз результатів. Після розгляду аналогів, присутніх на ринку, було прийнято рішення розробити новий продукт для аналізу якості тестів, у якому буде реалізовано можливість складання тестів, проведення тестування за допомогою мережі Інтернет, з подальшим упровадженням таких моделей у розподілені віртуальні навчальні системи.

Тестові технології контролю знань і вмінь широко використовують у ЗВО України на проміжних етапах навчального процесу, але на кінцевому етапі використовуються нечасто. Упровадження тестування на іспитах стримується через відсутність інформації про якість тестів [14].

Тільки якісний тест забезпечує відповідний контроль знань і дає змогу скласти індивідуальні програми навчання. Належна структура та помірна важкість тестових завдань суттєво розширюють інформативність здобутих результатів як в плані необхідності корегування методик викладання дисциплін (розділів), так і в гуманістичному напрямі, допомагаючи студентам визначитися щодо ефективності своєї реалізації у відповідній дисципліні (зокрема порівняти рівень своєї підготовки з іншими студентами групи).

Для мінімізації похибок тест має розроблятися за певними правилами та процедурами, що, наприклад, передбачають перевірку якості як самого тесту, так і його окремих завдань [13].

Тому процес оцінювання якості освітніх тестів має достатню кількість повторюваних кроків і потребує значних матеріальних та інтелектуальних витрат, що важливо звести до мінімуму, а це, безумовно, визначає необхідність автоматизації процесу [14].

Методи автоматизації та інтерпретації якості освітніх тестів мають виконувати такі завдання: формувати тест (добирати тестові завдання з дисципліни); коректувати матриці результатів, а саме: вилучати невдалі завдання, відповіді студентів, знання яких не можуть бути оцінені з допомогою набору тестових завдань, що підлягає аналізу; оцінювати якість тестів, зокрема розраховувати коефіцієнти надійності тесту, коефіцієнт валідності тесту та якість окремого завдання з допомогою статистичних показників моди й медіани, дисперсії,

середньоквадратичного відхилення, коефіцієнтів асиметрії та ексцесу.

Також як результати застосування методів можна отримати коефіцієнти дискримінативності, індекси та коефіцієнти важкості завдань.

Методи інтерпретації наборів тестових завдань мають виконуватися в проведенні тестування знань у режимі реального часу.

Вхідною інформацією є тест і відповіді студентів на тест, оцінки експертів. Вихідна інформація – це значення коефіцієнта надійності, коефіцієнта валідності, коефіцієнта дискримінативності, індекс важкості завдання та оцінки студентів – учасників тестування.

Сучасна теорія тестування (IRT) [15] має такі основні припущення: існують приховані параметри особистості, не доступні для безпосереднього спостереження (у тестуванні це рівень підготовленості студента й рівень важкості завдання); залежно від моделі існують індикаторні змінні, пов'язані з прихованими параметрами й доступні для безпосереднього спостереження. За значеннями таких індикаторів можна судити про значення прихованих параметрів.

Якщо обрано тестування з пріоритетом "швидке навчання", тоді рішення додати навчальний об'єкт до набору тестових завдань проводиться за формулою:

$$lo_j = \begin{cases} 1, & lo_{j_{time}} \leq 1, lo_{j_{importance}} > 2, lo_{j_{level}} \geq K_{cu_{level}}, \\ 0, & lo_{j_{time}} > 1, lo_{j_{importance}} \leq 2, lo_{j_{level}} < K_{cu_{level}}. \end{cases} \quad (1)$$

Кількість часу на вивчення навчальних об'єктів має бути мінімальна: $\sum_{j \in J} lo_{j_{time}} \rightarrow \min$.

Якщо обрано тестування з пріоритетом на "глибоке навчання", тоді рішення додати навчальний об'єкт до тренінгу проводиться за формулою:

$$lo_j = \begin{cases} 1, & lo_{j_{level}} \geq K_{cu_{level}}, \\ 0, & lo_{j_{level}} < K_{cu_{level}}. \end{cases} \quad (2)$$

Кількість навчальних об'єктів має бути максимальна: $\sum_{j \in J} lo_{j_{time}} \rightarrow \max$.

Якщо обрано пріоритет "лише найважливіше", тоді рішення додати навчальний об'єкт до тренінгу проводиться за формулою:

$$lo_j = \begin{cases} 1, & lo_{j_{importance}} > 2, lo_{j_{level}} \geq K_{cu_{level}}, \\ 0, & lo_{j_{importance}} \leq 2, lo_{j_{level}} < K_{cu_{level}}. \end{cases} \quad (3)$$

Кількість часу на вивчення навчальних об'єктів має бути мінімальна, а обсяг інформації має бути максимальним:

$$\begin{cases} \sum_{j \in J} lo_{j_{time}} \rightarrow \min \\ \sum_{j \in J} lo_{j_{time}} \rightarrow \max. \end{cases} \quad (4)$$

Розроблений метод містить і методи розрахунку надійності тестів – жодний з наявних тестів не є ідеальним щодо надійності.

Показники надійності тесту можуть істотно змінюватись у разі незначних, на перший погляд, варіацій умов проведення тестування, зміни характеру тестових завдань, а також залежно від ступеня важкості завдань для суб'єкта навчання.

На ознаки надійності суттєво впливає характер вибірки, що використовується для перевірки тесту за показниками надійності. Особливе значення тут має діапазон розбіжності в оцінках. Так, якщо оцінки студента розміщуватимуться у відносно вузькому діапазоні значень і будуть близькі одна до одної, можна сподіватися, що й під час повторного дослідження ці оцінки також розміщуватимуться щільною гомогенною групою. Імовірно зміни

рангових місць окремих піддослідних будуть зовні незначними. Коефіцієнт надійності у такому разі буде завищений. Таке саме завищення коефіцієнтів може виникнути під час аналізу надійності, що проводиться на матеріалі вибірки, яка складається з контрастних груп, наприклад, з осіб, що мають найвищі та найнижчі оцінки з тесту. Тоді ці віддалені один від одного результати не перекриватимуться випадковими чинниками.

Оцінювання надійності тестів проводиться різними методами, які можна поділити на дві групи: оцінка, що основана на двократному тестуванні, й оцінка, яка передбачає однократне тестування. Дворазове тестування проводиться за допомогою одного й того самого тесту або двох паралельних тестів (рис. 1).

Оцінка надійності найчастіше будується на обчисленні кореляції між двома наборами результатів виконання одного й того самого тесту або двох паралельних форм. У цьому разі логіка міркувань ґрунтується на принципі: що вища кореляція, то більш надійним є тест [19].

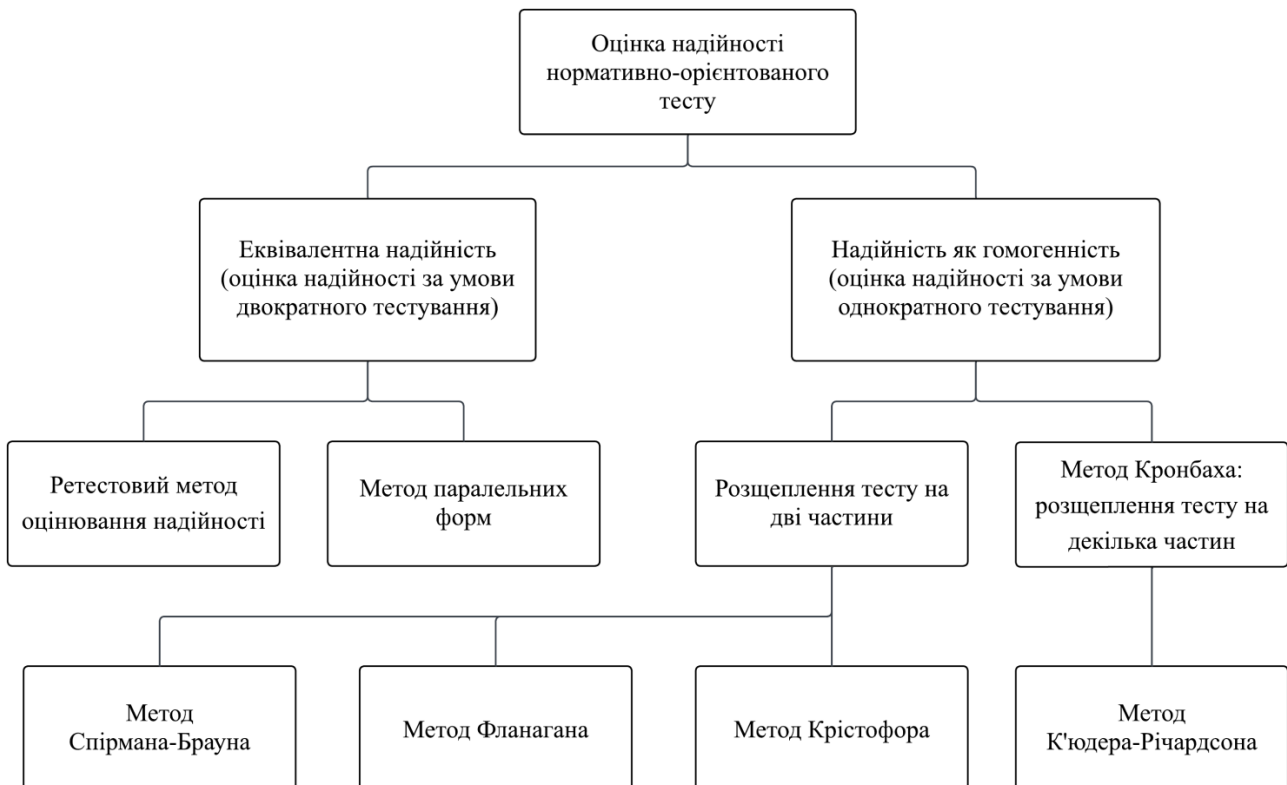


Рис. 1. Методи оцінювання надійності тесту

Розмаїття характеристик і показників надійності тесту так само велике, як і розмаїття умов, що можуть впливати на результат тесту, але найширшого практичного застосування набули:

- ретестова надійність:

$$r_n^{rem} = \frac{n \sum_{i=1}^n X_i Y_i - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right)}{\sqrt{n \left(\sum_{i=1}^n X_i \right)^2 - \left(\sum_{i=1}^n X_i \right)^2} \times \sqrt{n \left(\sum_{i=1}^n Y_i \right)^2 - \left(\sum_{i=1}^n Y_i \right)^2}}; \quad (6)$$

- надійність паралельних форм тесту:

$$r_n = \frac{2(r_n^{расц})}{1 + r_n^{расц}}; \quad (7)$$

– надійність частин тесту – розщеплення завдань тесту на рівноцінні половини є лише частковим випадком аналізу надійності частин тесту. Дійсно, можливе розщеплення на три, чотири й більше частин. У граничному випадку кількість таких частин дорівнює кількості пунктів тесту. Тоді аналіз надійності, власне, стає аналізом внутрішньої узгодженості, де $r_n^{расц}$ – кореляція між частинами тесту.

Для мінімізації похибок тест має розроблятися за певними правилами й процедурами, що, зокрема, передбачають перевірку якості як самого тесту, так і його окремих завдань.

Тому оцінювання якості освітніх тестів передбачає значну кількість повторюваних кроків і потребує чималих матеріальних та інтелектуальних витрат, які важливо звести до мінімуму, а це, безумовно, визначає необхідність автоматизації процесу.

Основним завданням запропонованого методу, згідно з *IRT*, є перехід від індикаторних змінних до прихованих змінних. Перелічимо переваги:

- відповідно до стандарту *IRT* перетворено вимірювання, що виконані в дихотомічних і порядкових шкалах, у лінійні вимірювання, потім якісні показники аналізуються за допомогою кількісних методів;

- міра вимірювання параметрів моделей є лінійною, що дає змогу використовувати широкий спектр статистичних процедур для аналізу результатів вимірювань;

- інваріантність завдань: оцінка важкості тестових завдань не залежить від вибірки досліджуваних, на яких вона була отримана;

- інваріантність здібностей: оцінка рівня підготовки студентів не залежить від використовуваного набору тестових завдань;

- неповнота інформації (пропуск деяких комбінацій "студент – тестове завдання") не є критичною.

Показники надійності тесту можуть істотно змінюватись у разі незначних, на перший погляд, варіацій умов проведення тестування, зміни характеру тестових завдань, а також залежно від ступеня важкості завдань для суб'єкта навчання.

На ознаки надійності суттєво впливає характер вибірки, що використовується для перевірки тесту за показниками надійності. Особливе значення тут має діапазон розбіжності в оцінках. Так, якщо оцінки студента розміщуватимуться у відносно вузькому діапазоні значень і будуть близькі одна до одної, можна сподіватися, що й під час повторного дослідження ці оцінки також розміщуватимуться щільною гомогенною групою. Імовірні зміни рангових місць окремих піддослідних будуть зовні незначними. Коефіцієнт надійності у такому разі буде завищений. Таке саме завищення коефіцієнтів може виникнути під час аналізу надійності, що проводиться на матеріалі вибірки, яка складається з контрастних груп, наприклад, з осіб, що мають найвищі та найнижчі оцінки з тесту. Тоді ці віддалені один від одного результати не перекриватимуться випадковими чинниками.

Призначення системи. На першому етапі необхідно визначити цілі навчання, що подаються в термінах компетенцій. Далі важливо провести початкове тестування суб'єкта навчання, щоб визначити його рівень. Тестування ґрунтується на використанні відповідного фреймворку, що зазвичай розробляється експертами професійних спільнот. Тематика навчальних і тестових матеріалів, яка відповідає компетенціям, подається у вигляді когнітивних карт і дає змогу визначити необхідні навчальні матеріали на основі застосування розроблених моделей. Тестування та адаптивний вибір навчальних об'єктів є перманентним процесом для досягнення цілей навчання.

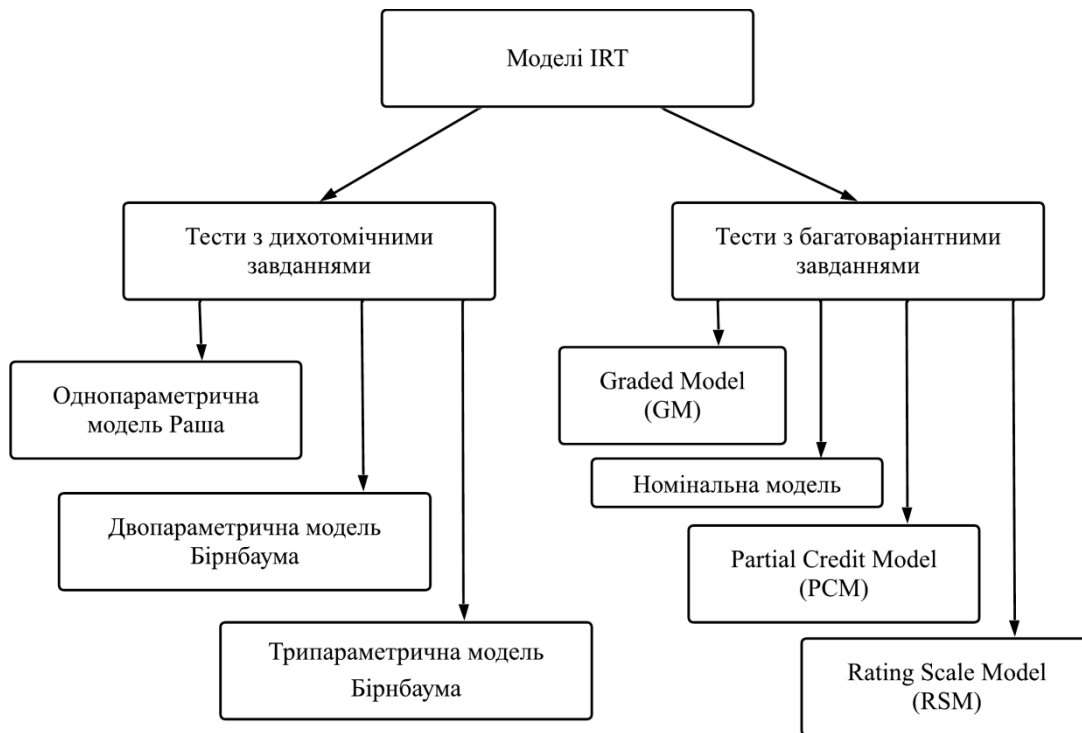


Рис. 2. Моделі IRT

Завдання реалізації запропонованих методів оцінювання якості тестів із певної дисципліни має бути успішно розв'язане способом автоматизації необхідних для її вирішення процесів. Розроблене програмне забезпечення може бути використане в розробленні викладачами нормативно-орієнтованих тестів. Відповіді на завдання тесту можуть надходити до програмної системи після проходження студентом тестування в режимі реального часу або способом їх додавання до бази даних. Оброблення результатів передбачає розрахунок оцінки якості та інтерпретацію наборів тестових завдань на основі відповідей студентів. ПЗ може бути автономним модулем або програмним компонентом, що інтегрується з підсистемою віртуального розподіленого навчального програмного ресурсу.

Приклад моделювання завдання оцінювання якості комплексів тестових завдань

Алгоритм вирішення завдання передбачає такі етапи:

1. Розроблення завдань.
2. Створення тесту.
3. Проведення пробного тестування.
4. Вибракування й корегування тестових завдань.

5. Розрахунок інтеркореляцій завдань.

6. Видалення або перероблення завдань із високою однорідністю.

7. Проведення тестування за новим тестом, у якому всі завдання відповідають умовам якості.

8. Розрахунок валідності тесту.

9. Розрахунок надійності тесту.

10. Проведення тестування за якісним тестом.

Алгоритм вибракування тестових завдань передбачає такі етапи:

1. Перевірка на нормальність розподілу балів завдань.

2. Видалення тих завдань, розподіл яких не є нормальним.

3. Розрахунок індивідуального бала суб'єктів навчання.

4. Розрахунок середніх результатів сумарних балів суб'єктів навчання.

5. Розрахунок середніх результатів суб'єктів навчання з кожного завдання.

6. Розрахунок коефіцієнта важкості завдання.

7. Розрахунок індексу важкості завдання з огляду на відгадування.

8. Розрахунок дисперсії та стандартного відхилення сумарних балів суб'єктів навчання.

9. Розрахунок стандартного відхилення сумарних балів суб'єктів навчання.

10. Вилучення завдань, для яких коефіцієнт важкості близький до 1 або до 0.

11. Розрахунок дисперсії результатів випробуваних за j -м завданням.

12. Розрахунок стандартного відхилення результатів випробуваних за j -м завданням.

13. Вилучення завдань із високою дисперсією.

14. Розрахунок індексу важкості завдань.

15. Розрахунок коефіцієнта дискримінативності завдання та вилучення завдань, для яких оцінка близька до 0,9 або до 0.

Розроблення системи опису основних інформаційних об'єктів та їх властивостей для конкретної предметної галузі – бізнес-правил – це наступний етап проектування програмного забезпечення аналізу наборів тестових завдань. Сформульовано такі бізнес-правила: кожен суб'єкт тестування має унікальний ідентифікатор; кожний тест може мати багато запитань; у кожному запитанні може бути багато відповідей; необхідно, щоб кожне запитання мало одну правильну відповідь; кожний тест може використовуватися в багатьох тестуваннях; кожна відповідь має текстовий опис; кожна помилкова

відповідь може коригуватись на відгадування; для кожного суб'єкта навчання за тестом можна розрахувати індивідуальний бал; для кожного завдання тесту можна обчислити коефіцієнт важкості завдання; для кожного завдання тесту можна обчислити індекс важкості завдання; для кожного завдання тесту можна обчислити дискримінативність; для кожного тесту можна обчислити надійність за обраним методом та за оцінкою експертів, у формулі для розрахунку валідності тесту береться середнє i -го суб'єкта навчання. Отже, для кожного тесту можна обчислити валідність.

Користувачами системи є викладач, студент та адміністратор. Викладач розробляє тести, вводить результати проведеного тестування, розраховує оцінки надійності тестів та оцінки знань студентів. Суб'єкт навчання (студент) – виконавець тесту. Адміністратор відповідає за реєстрацію користувачів і супровід БД.

Основні функціональні вимоги до програмної системи з погляду викладача, студента, адміністратора подані у вигляді діаграм варіантів застосування (рис. 3).

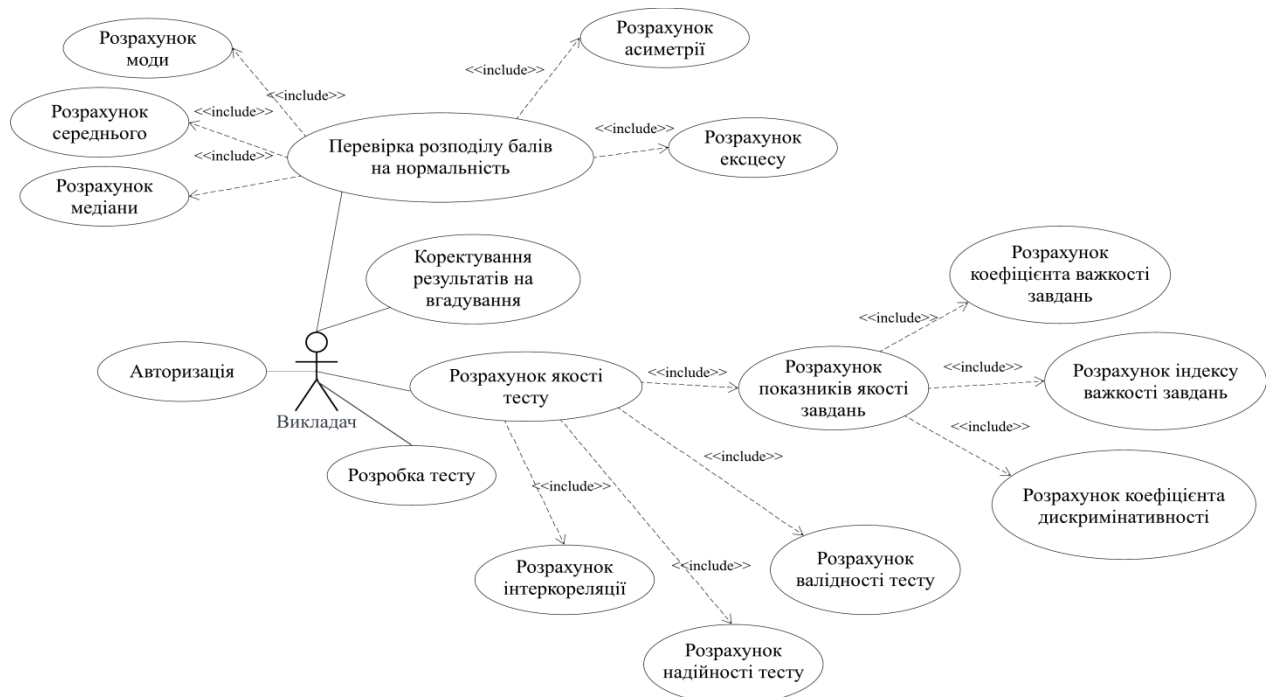


Рис. 3. Діаграма використання системи викладачем

Діаграма послідовності для прецеденту "розрахунок надійності" зображена на рис. 4. Розроблена діаграма класів – статичне подання структури моделі – відображає статичні елементи, такі як: класи, типи даних, їх зміст і відношення.

Метою реалізації алгоритмів аналізу даних (*AnalysisServlet*) є вирішення підзадачі, що пов'язана з аналізом показників тестування для отримання результатів.

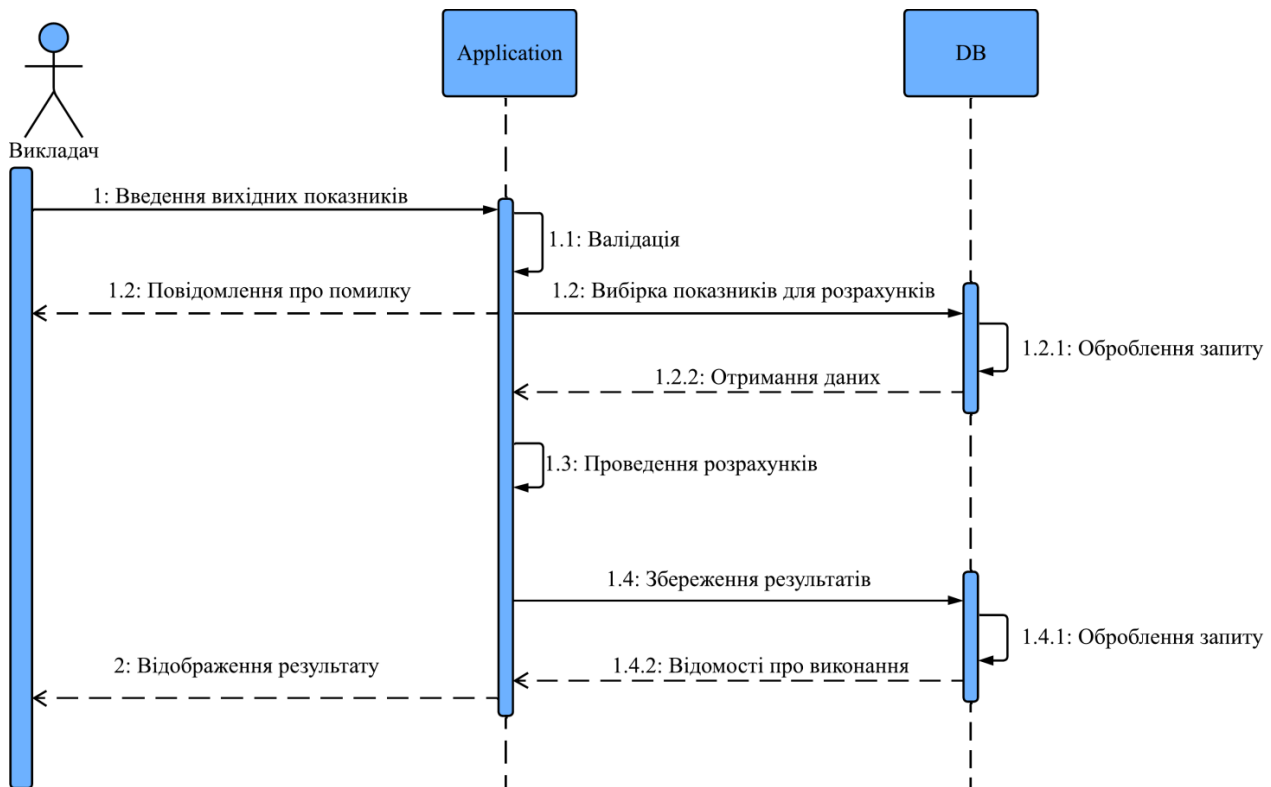


Рис. 4. Діаграма послідовності

До функцій компоненти "управління анкетами" належать: перегляд списку розроблених тестів; розроблення нових тестів; редагування змісту тесту.

Функції компоненти "збір даних онлайн" передбачають: відображення тесту; автоматичний збір даних за допомогою моделей штучних агентів; збереження відповідей на тест у базі даних.

До функцій компоненти "аналіз даних" належить: проведення аналізу завдань; аналіз розподілу балів; розрахунок коефіцієнта дискримінативності, індексу важкості завдань, коефіцієнта важкості завдань.

Компонента "аналіз тесту" містить такі підзадачі: розрахунок коефіцієнта надійності, коефіцієнта валідності та інтеркореляції, побудова гістограм.

Результати розрахунків мають показати значення коефіцієнтів надійності. Високі значення таких коефіцієнтів, отриманих унаслідок різних методів розрахунків надійності, свідчать про високу надійність розробленого методу.

Потім викладач має визначити розрахунок валідності тесту, де він обирає зі списку дисципліну й тест, потрібне тестування та обчислює валідність тесту. Отримано достатньо високі значення коефіцієнта валідності.

Наступний крок реалізації передбачає розрахунок якості завдань. Відкривається сторінка, на якій

обирається дисципліна й тест, потрібний вид тестування (проміжне, підсумкове тощо), набір запитань, обчислюється дискримінативність і коефіцієнти важкості завдань тесту, аналізується розподіл балів на основі розрахованих статистик для корегування на відгадування результатів тестування або для первинних результатів тестування.

Розробляється тест, у якому на першому етапі 14 запитань. Результати тестування, отримані групою з 50 студентів, наведені далі. Було проведено тестування функціональності програмних рішень. Так, за індексом важкості завдань, скоректованих на відгадування, не одне із завдань не вилучатиметься з тесту.

Аналіз показників дискримінативності завдань – критичне значення коефіцієнта дискримінації знаходиться зі статистичних таблиць. Воно дорівнює 0,28. Отже, всі запитання цього тесту валідні. Унаслідок чого надійність тесту становить 0,78545. Отримані високі значення надійності, що пов'язано з високою однорідністю запитань.

Аналіз результатів розроблених моделей доводить, що для 5 % рівня значущості знайдено критичний рівень кореляції, який дорівнює 0,40. Це свідчить про високий рівень надійності тесту, що відповідає показникам якості.

Висновки

Серед факторів, які формують ситуацію у сфері педагогічного контролю, можна визначити основний – необхідність упровадження в практику освіти кількісних методів оцінювання знань студентів. Методи ідентифікації та подання властивостей наборів тестових завдань із використанням алгебри предикатів можуть одночасно подати знання у вигляді рівнянь алгебри предикатів.

Упровадження кількісних методів оцінювання наборів тестових завдань передбачає правильне визначення цілей контролю, предмета вимірювання та вибір засобів вимірювання. Використання педагогічних тестів сприяє ефективній реалізації всіх функцій контролю знань і відповідає основним його принципам та принципам побудови віртуального розподіленого навчального простору.

Педагогічні тести відкривають перспективні напрями підвищення якості дистанційної освіти

завдяки вдосконаленню системи контролю та посилення мотивації до навчання. З упровадженням системи електронного тестування цю проблему вдається вирішити, крім того, унеможливаються негативні ситуації, пов'язані з необ'єктивністю викладачів. Статистичні методи визначення рівня знань на основі тестів не мають проблем суб'єктивізму експертного оцінювання, але пов'язані з певною похибкою вимірювань. Тому питання щодо розроблення нових методів інтелектуального оцінювання якості таких вимірювань, що містять оцінки надійності та валідності наборів тестових завдань, є актуальним.

У роботі розглянуто питання формалізації оцінювання тестів, аналізу вимог до програмного забезпечення тестової підсистеми віртуального розподіленого навчального середовища, розроблення програмного забезпечення для розв'язання задачі оцінювання якості тестів.

Список літератури

1. Соколовський Я. І., Сторожук О. Л., Крошний І. М. Застосування сучасних інформаційно-комунікаційних технологій для організації дистанційного навчання. Науковий вісник НЛТУ України. 2015. Вип. 25.6. С. 243–251. URL: http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&2_S21P03=FILE=&2_S21STR=nnvntu_2015_25
2. Learning management system. URL: https://en.wikipedia.org/wiki/Learning_management_system (дата звернення: 26.08.2023).
3. Backer R. Educational data mining and learning analytics. The Cambridge handbook of the learning sciences. 2019. 274 p. DOI: 10.1017/CBO9781139519526.016
4. A Guide to the Scrum Body of Knowledge (SBOK™ GUIDE) URL: <https://www.scrumstudy.com/SBOK/SCRUMstudy-SBOK-Guide-3rd.pdf> (дата звернення: 26.08.2023).
5. Gruzdo I., Kyrychenko I., Tereshchenko G., Shanidze O. Analysis of Models Usability Methods Used on Design Stage to Increase Site Optimization *Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems. (COLINS)*, Volume III: Intelligent Systems Workshop, 2023. In CEUR Workshop Proceedings, Vol-3403, P. 387–4093. URL: <https://ceur-ws.org/Vol-3403/paper31.pdf>
6. Shubin I. Development of conjunctive decomposition tools. CEUR Workshop Proceedings, 2021. P. 890–900. URL: <https://ceur-ws.org/Vol-2870/>
7. Karataiev O., Sitnikov D., Sharonova N. A Method for Investigating Links between Discrete Data Features in Knowledge Bases in the Form of Predicate Equations, CEUR Workshop Proceedings, 2023. P. 224–235. URL: <https://ceur-ws.org/Vol-3387/paper17.pdf>
8. Тест як інструмент педагогічного моніторингу. URL: <http://opentest.com.ua/test-kak-instrument-izmereniya-urovniya-znanij> (дата звернення: 20.08.2023).
9. MyTestXPro – система програм для створення та проведення комп'ютерного тестування, збору та аналізу їх результатів. URL: <http://mytest.net> (дата звернення: 28.08.2023).
10. Комп'ютерна програма тестування OpenTEST2. URL: <http://opentest.com.ua/kompyuternaya-programma-testirovaniya-znanij-opentest-2> (дата звернення: 21.08.2023).
11. Конструктор тестів Keepsoft. URL: <http://www.keepsoft.ru/simulator.htm>. (дата звернення: 20.08.2023.)
12. Безкоштовна програма для тестування знань та онлайн підготовки. URL: <http://xtls.org.ua/test.html> (дата звернення: 20.08.2023).
13. Sharonova N. et al. Issues of Fact-based Information Analysis. International Conference on Computational Linguistics and Intelligent Systems. 2018. 178 p. URL: http://web.kpi.kharkov.ua/iks/wp-content/uploads/sites/113/2021/10/preface_colins_volume2_2018.pdf
14. Williams P. E-learning: what the literature tells us about distance education. *An overview. Aslib Proceedings*. Vol. 57. 2005. P 109–122. DOI: <https://doi.org/10.1108/00012530510589083>

15. Omran P. G., Wang K., Wang Z. An Embedding-based Approach to Rule Learning in Knowledge Graphs, *IEEE Transactions on Knowledge and Data Engineering*. Vol. 33(4). 2021. P. 1348–1359. DOI: 10.1109/TKDE.2019.2941685
16. Pellissier-Tanon T., Weikum G., Suchanek F. YAGO 4: A Reasonable Knowledge Base, *17th International Conference, ESWC 2020*, Heraklion, Crete, Greece, May 31–June 4, 2020. P. 583–596. DOI:10.1007/978-3-030-49461-2_34
17. Kyrychenko I., Malikin D. Research of Methods for Practical Educational Tasks Generation Based on Various Difficulty Levels 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2022), May 12–13, 2022, Gliwice, Poland. CEUR Workshop Proceedings 3171, Volume I: Main, P. 1030–1042. URL: <https://ceur-ws.org/Vol-3171/paper74.pdf>
18. Omran P. G., Wang Z., Wang K. Scalable rule learning via learning representation, *Twenty-Seventh International Joint Conference on Artificial Intelligence. IJCAI-18*. 2018. P. 2149–2155. DOI:10.24963/ijcai.2018/297
19. Svato M., Schockaert S., Davis J. STRiKE: Rule-Driven Relational Learning Using Stratified k-Entailment, in: ECAI, 2020. URL: <https://ida.fel.cvut.cz/~kuzelka/pubs/ecai2020.pdf>
20. Sharonova N., Gruzdo I., Tereshchenko G. Generalized Semantic Analysis Algorithm of Natural Language Texts for Various Functional Style Types. 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2022), May 12–13, 2022, Gliwice, Poland. CEUR Workshop Proceedings 3171, Volume I: Main, P. 16–26. URL: <https://ceur-ws.org/Vol-3171/paper4.pdf>
21. Barkovska O. Research into Speech-to-text Transformation Module in the Proposed Model of a Speaker's Automatic Speech Annotation. *Innovative Technologies and Scientific Solutions for Industries*. 2022. № 4 (22). P. 5–13. DOI: <https://doi.org/10.30837/ITSSI.2022.22.005>

References

1. Sokolovskiy, Ya, Storozhuk, O., Kroshnyi, I. "The use of modern information and communication technologies for the organization of distance learning". Scientific Bulletin of the National Technical University of Ukraine. 2015. Issue 25.6. P. 243–251. available at: http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&S21P03=FILE=&S21STR=nvntu_2015_25
2. "Learning management system". available at: https://en.wikipedia.org/wiki/Learning_management_system (last accessed 26.08.2023).
3. Backer, R. (2019), *Educational data mining and learning analytics*. The Cambridge handbook of the learning sciences. 274 p. DOI: 10.1017/CBO9781139519526.016
4. "A Guide to the Scrum Body of Knowledge (SBOK™ GUIDE)", available at: <https://www.scrumstudy.com/SBOK/SCRUMstudy-SBOK-Guide-3rd.pdf> (last accessed 26.08.2023).
5. Gruzdo, I., Kyrychenko, I., Tereshchenko, G., Shandzhe, O. "Analysis of Models Usability Methods Used on Design Stage to Increase Site Optimization" Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems. (COLINS), Volume III: Intelligent Systems Workshop, 2023. In CEUR Workshop Proceedings, Vol-3403, P. 387–4093. available at: <https://ceur-ws.org/Vol-3403/paper31.pdf>
6. Shubin, I. "Development of conjunctive decomposition tools". CEUR Workshop Proceedings, 2021. P. 890–900. available at: <https://ceur-ws.org/Vol-2870/>
7. Karataiev, O., Sitnikov, D., Sharonova, N. "A Method for Investigating Links between Discrete Data Features in Knowledge Bases in the Form of Predicate Equations", CEUR Workshop Proceedings, 2023, P. 224–235. available at: <https://ceur-ws.org/Vol-3387/paper17.pdf>
8. "The test as a tool of pedagogical monitoring" ["Test yak instrument pedahohichnoho monitorynuh"] available at: <http://opentest.com.ua/test-kak-instrument-izmereniya-urovnya-znanij> (last accessed 20.08.2023.)
9. "MyTestXPro – system of programs for creating and conducting computer testing, collecting and analyzing their results". ["Systema prohram dlia stvorennia ta provedennia kompiuternoho testuvannia, zboru ta analizu yikh rezultativ"] available at: <http://mytest.net> (last accessed 28.08.2023).
10. "Computer testing program OpenTEST2". ["Kompiuterna prohrama testuvannia OpenTEST2"]. available at: <http://opentest.com.ua/kompyuternaya-programma-testirovaniya-znanij-opentest-2>. (last accessed 21.08.2023).
11. "Test designer Keepsoft". ["Konstruktor testiv Keepsoft"]. available at: <http://www.keepsoft.ru/simulator.htm>. (last accessed 20.08.2023).
12. "Free program for knowledge testing and online preparation". ["Bezkoshtovna prohrama dlia testuvannia znan ta onlain pidhotovky"]. available at: <http://xtls.org.ua/test.html> (last accessed 20.08.2023).
13. Sharonova, N. et al. "Issues of Fact-based Information Analysis". International Conference on Computational Linguistics and Intelligent Systems. 2018. 178 p. available at: http://web.kpi.kharkov.ua/iks/wp-content/uploads/sites/113/2021/10/preface_colins_volume2_2018.pdf
14. Williams, P. (2005), "E-learning: what the literature tells us about distance education". *An overview. Aslib Proceedings*. Vol. 57. P 109–122. DOI: <https://doi.org/10.1108/00012530510589083>
15. Omran, P. G., Wang, K., Wang, Z. (2021), "An Embedding-based Approach to Rule Learning in Knowledge Graphs", *IEEE Transactions on Knowledge and Data Engineering*. Vol. 33(4). P. 1348–1359. DOI: 10.1109/TKDE.2019.2941685
16. Pellissier-Tanon, T., Weikum, G., Suchanek, F. (2020), "F. YAGO 4: A Reasonable Knowledge Base", *17th International Conference, ESWC 2020*, Heraklion, Crete, Greece, May 31–June 4. P. 583–596. DOI:10.1007/978-3-030-49461-2_34

17. Kyrychenko, I., Malikin, D. "Research of Methods for Practical Educational Tasks Generation Based on Various Difficulty Levels" 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2022), May 12–13, 2022, Gliwice, Poland. CEUR Workshop Proceedings 3171, Volume I: Main, P. 1030–1042. available at: <https://ceur-ws.org/Vol-3171/paper74.pdf>

18. Omran, P. G., Wang, Z., Wang, K. (2018), "Scalable rule learning via learning representation", *Twenty-Seventh International Joint Conference on Artificial Intelligence. IJCAI-18*. P. 2149-2155. DOI:10.24963/ijcai.2018/297

19. Svato, M., Schockaert, S., Davis, J. "STRiKE: Rule-Driven Relational Learning Using Stratified k-Entailment", in: ECAI, 2020. available at: <https://ida.fel.cvut.cz/~kuzelka/pubs/ecai2020.pdf>

20. Sharonova, N., Gruzdo, I., Tereshchenko, G. "Generalized Semantic Analysis Algorithm of Natural Language Texts for Various Functional Style Types". 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2022), May 12–13, 2022, Gliwice, Poland. CEUR Workshop Proceedings 3171, Volume I: Main, P. 16–26. available at: <https://ceur-ws.org/Vol-3171/paper4.pdf>

21. Barkovska, O. (2022), Research into Speech-to-text Transformation Module in the Proposed Model of a Speaker's Automatic Speech Annotation. *Innovative Technologies and Scientific Solutions for Industries*. № 4 (22). P. 5–13. DOI: <https://doi.org/10.30837/ITSSI.2022.22.005>

Надійшла 28.08.2023

Відомості про авторів / About the Authors

Ляшик Володимир Андрійович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії, Харків, Україна; e-mail: volodymyr.liashyk@nure.ua; ORCID ID: <https://orcid.org/0000-0001-7326-0813>

Шубін Ігор Юрійович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри програмної інженерії, Харків, Україна; e-mail: igor.shubin@nure.ua; ORCID ID: <https://orcid.org/0000-0002-1073-023X>

Liashyk Volodymyr – Kharkiv National University of Radio Electronics, Postgraduate, Kharkiv, Ukraine.

Shubin Ihor – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Professor at the Software Department, Kharkiv, Ukraine.

METHODS OF AUTOMATION AND INTERPRETATION OF THE QUALITY EDUCATIONAL TESTS IN DISTRIBUTED VIRTUAL LEARNING SYSTEMS

The **subject matter** of the article is the development of mathematical and algorithmic support for intellectual tools, which allows you to conduct continuous control of the knowledge of students (subjects of study) objectively and comprehensively. The **goal** of the work is to create methods for assessing the quality of educational tests and automating such processes. The following **tasks** were solved in the article: formation of a testing model in a distributed virtual learning environment and a validity assessment model based on the content of sets of test tasks. The following **methods** used are – algebra of finite predicates and operations, methods of mathematical statistics and methods of intellectual data analysis. The following **results** were obtained – the principles of intellectual analysis of the value of reliability coefficients, validity coefficients, discriminability coefficient, the difficulty index of the task of assessing the knowledge of subjects of training are formulated. **Conclusions:** the application of methods of formalization of test evaluation, analysis of software requirements, software development confirm the need to introduce quantitative methods of assessing students' knowledge into educational practice. The introduction of quantitative methods involves the correct setting of control goals, the selection of the measurement object and the selection of measurement tools. The use of pedagogical tests contributes to the effective implementation of all control functions and corresponds to its main principles for solving the problem of assessing the quality of tests, based on distributed virtual learning models and analysis methods. A test example of calculations was developed, with the help of which the efficiency of the proposed methods was researched.

Keywords: software engineering; knowledge bases; algebra of finite predicates; use of knowledge; rules.

Бібліографічні описи / Bibliographic descriptions

Ляшик В. А., Шубін І. Ю. Методи автоматизації та інтерпретації якості освітніх тестів у системах розподіленого віртуального навчання. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 40–51. DOI: <https://doi.org/10.30837/ITSSI.2023.25.040>

Liashyk, V., Shubin, I. (2023), "Methods of automation and interpretation of the quality educational tests in distributed virtual learning systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 40–51. DOI: <https://doi.org/10.30837/ITSSI.2023.25.040>