

УДК 519.854:004.8

О.Є. Коноваленко¹, В.О. Брусенцев²¹ Національний технічний університет «Харківський політехнічний інститут», Харків² Харківська державна академія культури, Харків

ПРОБЛЕМА АВТОМАТИЗОВАНОГО ТЕСТУВАННЯ ТА ОЦІНЮВАННЯ ЗНАТЬ

Розглядаються та аналізуються проблеми впровадження сучасних автоматизованих систем тестування. Пропонується підхід до організації тестування, пов'язаний із застосуванням методів інтелектуальних обчислень та інженерії знань. Формулюються принципи інтелектуального тестування, покладені в основу нового класу програмних систем контролю та оцінювання знань.

Ключові слова: тестування, освіта, дистанційне навчання, обробка інформації.

Вступ

Сучасний стан вітчизняної системи навчання характеризується досить високою насиченістю вищих та інших навчальних закладів засобами обчислювальної техніки, що змушує задуматися над ефективністю її застосування в навчальному процесі. Один з найпоширеніших напрямків – створення та експлуатація автоматизованих систем контролю знань. У цей час відома безліч практичних реалізацій систем автоматизованого тестування як за окремими дисциплінами (предметні тести), так і універсальних систем оцінювання знань (“конструктори тестів”), повністю або частково інваріантних до конкретних дисциплін, що допускають їхнє інформаційне наповнення викладачами – організаторами тестування.

Основний матеріал

Аналіз ефективності автоматизованого тестування у вищому або іншому навчальному закладах показує, що багато викладачів насторожено і навіть негативно ставляться до подібних систем. Серед найбільш істотних недоліків сучасних підходів до автоматизованого тестування, названих як причини такого негативного відношення, можна відзначити:

– необхідність формулювання варіантів відповідей на тестові завдання за принципом “один абсолютно правильний” – “N абсолютно неправильних”. Це не дає можливості організувати повноцінне тестування за слабо формалізованими дисциплінами, для яких характерна діалектичність знань (дисципліни суспільно-політичного, гуманітарного, соціально-економічного і т.ін. циклів);

– примітивність і негнучкість процедур розрахунку підсумкової оцінки, що зводять або до визначення відносини кількості правильних відповідей до кількості заданих питань, або до підсумовування балів, призначуваних за кожну правильну відповідь;

– неможливість автоматизації різноманітних методик контролю знань, широко застосовуваних у педагогічній практиці (оцінка широти або глибини

знань, облік відносної важливості окремих тем або розділів досліджуваної дисципліни, вибір складності тесту з урахуванням рівня підготовленості і самооцінки тестуємого, стимуляція правильних відповідей і т.п.);

– значна трудомісткість ручного формування такої безлічі тестових завдань і варіантів відповідей на кожне з них, що дозволить виключити або мінімізувати ймовірність пред'явлення того самого завдання різним тестуємим при паралельній перевірці їхніх знань.

Особливо яскраво зазначені недоліки автоматизованого тестування проявляються при контролі знань за дисциплінами гуманітарного, соціально-економічного і суспільно-політичного циклів. Ступінь формалізації знань за цими дисциплінами у силу діалектичності занадто низька, щоб їхня наявність могла визначатися по тому, наскільки добре пам'ятає екзаменуємий окремі факти, точні визначення або конкретні формули й правила їхнього застосування.

Можливість постійного оперативного моніторингу якості знань на всьому протязі освітнього процесу – у середній школі або ВНЗ – істотно підвищила б ефективність навчання.

Система оцінювання знань – система оцінювання якості освоєння освітніх програм учнем, найважливіший елемент освітнього процесу. У цей час у світі використовується безліч шкал оцінювання знань. У деяких шкалах прийнято використовувати цифрові позначення розрядів, причому допускаються дробові оцінки, інші шкали (наприклад, американська) за традицією мають справу з літерними позначеннями. Американська шкала також має чисельну інтерпретацію, при якій вищим оцінкам A і A+ відповідає бал 4.

Тест – це інструмент для оцінки знань, тому, як будь-який інструмент виміру, він повинен мати вимірювальну шкалу й погрішність виміру. Крім того, він повинен бути надійним інструментом (не давати збоїв), тобто результати виміру знань людей з однаковим рівнем підготовки повинні бути однакові.

Слово «тест» англійського походження й мовою оригіналу означає «випробування», «перевірка». Тест навченості – це сукупність завдань, зорієнтованих на визначення (вимір) рівня (ступеня) засвоєння певних аспектів (частин) змісту навчання.

Правильно складені тести навченості повинні задовольняти ряду вимог. Вони повинні бути:

- відносно короткостроковими, тобто не витрачати більших витрат часу;
- однозначними, тобто не допускати довільного тлумачення тестового завдання;
- правильними, тобто виключати можливість формулювання багатозначних відповідей;
- відносно короткими, потребуючими стислих відповідей;
- інформаційними, тобто такими, які забезпечують можливість співвіднесення кількісної оцінки за виконання тесту з порядковою або навіть інтервальною шкалою вимірів;
- зручними, тобто придатними для швидкої математичної обробки результатів;
- стандартними, тобто придатними для широкого практичного використання – виміру рівня навченості можливо більш широких контингентів тих, кого навчають, що опановують однаковий обсяг знань на тому самому рівні навчання.

Визначення параметрів тесту – це процес прив'язки результатів до бажаної шкали користувача і визначення погрішності виміру та надійності тесту.

Визначивши параметри тесту, потрібно зрівняти їх з виробленими практикою критеріями. Якщо параметри не задовольняють потрібним критеріям, склад тесту змінюють і визначають параметри ще раз. Така процедура повторюється доти, поки необхідні критерії не будуть задоволені.

Тести – найвідоміші та розповсюджені з оцінних процедур. Якщо говорити про професійно розроблені тести, то надійність і точність інформації, отриманої на їхній основі, становить 0,35 – 0,45. Це не багато, але вище, ніж точність неструктурованих інтерв'ю (0,1) і рекомендацій (0,23).

Також серйозним обмеженням застосування тестів є те, що вони добре оцінюють відносно прості психологічні фактори – формальний рівень інтелекту, емоційну стійкість, комунікабельність. І дуже погано вимірюють складні фактори – лідерство, іноваційність, відкритість до навчання, орієнтацію на результат.

Тестування часто проходить у комп'ютерному варіанті. Можливе його дистанційне проведення із централізацією на єдиному сайті. Результати повинні генеруватися автоматично. Кількість одночасно тестуємих фахівців обмежене тільки технічними можливостями.

При розробці тестів важливо, наскільки вони відповідають запроєктованим цілям навчання, утворення, розвитку тих, кого навчають. Найважливішими критеріями діагностичних тестів навченості є дієвість (валі-

дність, показовість), надійність (імовірність, правильність), диференційованість (розрізнення).

Дієвість тесту за своїм змістом близька до вимоги повноти, всебічності перевірки, пропорційного подання всіх елементів досліджуваних знань, умінь. Термін «дієвість» має, як мінімум, два синоніми – валідність (від англ. *valid* – що має значення, коштовний) і показовість, показність, трактуємі так само, як репрезентативність. Завжди мається на увазі, що укладач тесту зобов'язаний ретельно вивчити всі розділи навчальної програми, навчальні книги, добре знати мету і конкретні завдання навчання. Лише тоді він зможе скласти тести, які будуть діючими для певної категорії тих, кого навчають.

Чітка і ясна постановка питання в межах освоєних знань – невід'ємна умова дієвості тесту, що сильно впливає на результати тестування. Якщо тест виходить за межі освоєного змісту або ж не досягає цих меж, перевищує запроєктований рівень навчання, то він не буде діючим для тих, кого навчають, котрим він адресований. Дієвість тесту визначається статистичними методами. Величина 0,7 – 0,9 свідчить про високу дієвість тестів навченості. Якщо ж коефіцієнт кореляції досягає 0,45 – 0,55, то дієвість тесту вважається задовільною, при більш низьких значеннях вона вважається незадовільною.

Ступінь надійності тестування характеризується стабільністю, стійкістю показників при повторних вимірах за допомогою того ж тесту або його рівноцінного замітника. Кількісно цей показник характеризується ймовірністю досягнення запроєктованих результатів (правильністю значень). Грамотно складені та апробовані тести навченості дозволяють досягти коефіцієнта надійності 0,9. Встановлено, що надійність тесту підвищується при збільшенні кількості тестових завдань.

Встановлено також, що чим вище тематична, змістовна розмаїтість тестових завдань, тим нижче надійність результату тестування. Це варто розуміти так: тест, націлений на перевірку засвоєння конкретної теми, завжди буде більш надійним, ніж тест, спрямований на перевірку всього розділу (курсу), що охоплює значну кількість матеріалу – закономірностей, концепцій, фактів. Відбувається це саме тому, що змістовна розмаїтість останнього вище.

Надійність тестів навченості значно залежить від труднощів їхнього виконання. Труднощі визначаються за співвідношенням правильних і неправильних відповідей на тестові питання. Включення до складу тестів таких завдань, на які всі ті, кого навчають, відповідають правильно або ж, навпаки, неправильно, різко знижує надійність результату тестування в цілому. Найбільшу практичну цінність мають завдання, на які правильно відповідають 45 – 80% учнів.

Характеристика диференційованості (розрізнення) пов'язана з використанням таких тестів, де потрібно вибирати правильну відповідь із декількох можливих

альтернатив. Якщо, скажемо, всі учні безпомилково знаходять правильну відповідь на одне питання і також дружно не можуть відповісти на інше, то це сигнал для вдосконалювання тесту в цілому. Його необхідно диференціювати, зробити помітним. Інакше подібні завдання не допоможуть відокремити тих, хто засвоїв матеріал на необхідному рівні, від тих, хто заданих рівнів не досяг. Практично диференціюють тести за результатами тестування, порівнюючи результати виконання тесту в цілому з результатами виконання окремих завдань. Якщо коефіцієнт кореляції між відповідями на конкретні завдання і на тест в цілому більше 0,5, то це свідчить про достатню диференційованість тесту.

Тести навченості застосовуються на всіх етапах дидактичного процесу. З їхньою допомогою забезпечується попередній, поточний, тематичний і підсумковий контроль знань, умінь, облік успішності, академічних досягнень.

1. Попередній контроль. Успіх вивчення будь-якої теми (розділу або курсу) залежить від ступеня засвоєння тих понять, термінів, положень і т.д., які вивчалися на попередніх етапах навчання. Якщо інформації про це у викладача нема, то він не має змоги проектування і керування в навчальному процесі, вибору оптимального його варіанту. Необхідну інформацію викладач одержує, застосовуючи попередній контроль (облік) знань. Останній необхідний ще й для того, щоб зафіксувати (зробити зріз) вихідний рівень навченості. Порівняння вихідного початкового рівня навченості з кінцевим (досягнутим) дозволяє вимірювати «приріст» знань, ступінь сформованості вмінь і навичок. Якщо відомі вхідні та вихідні характеристики системи, проблеми її оптимізації вважаються багато в чому вирішеними. Зібрати максимальний обсяг інформації про вхідні характеристики навченості і оцінити їх у кількісних показниках вдається за допомогою тестування, що здійснюється за допомогою спеціально розроблених для цієї мети завдань.

2. Поточний контроль. Він необхідний для діагностування ходу дидактичного процесу, виявлення динаміки останнього, зіставлення реально досягнутих на окремих етапах результатів тестування із запроєктованими. Крім властиво прогностичної функції поточний контроль і облік знань, умінь стимулює навчальну працю учнів, сприяє своєчасному визначенню пробілів у засвоєнні матеріалу, підвищенню загальної продуктивності навчальної праці.

3. Тематичний контроль. Складання тематичного тестового завдання вимагає кропіткої і ретельної праці. Адже мова йде не просто про перевірку засвоєння окремих елементів, а про розуміння системи, що поєднує ці елементи. Значну роль при цьому грають синтетичні, комплексні завдання, що поєднують питання про окремі поняття теми, спрямовані на виявлення інформаційних зв'язків між ними.

4. Підсумковий контроль. Здійснюється під час

заключного повторення, а також у процесі іспитів (заліків). Саме на цьому етапі дидактичного процесу систематизується та узагальнюється навчальний матеріал. З високою успішністю можуть бути застосовані відповідним чином складені тести навченості.

Природно, не всі необхідні характеристики засвоєння можна одержати засобами тестування. Такі, наприклад, показники, як уміння конкретизувати свою відповідь прикладами, знання фактів, уміння складно, логічно і доказово виражати свої думки, деякі інші характеристики знань, умінь, навичок діагностувати тестуванням неможливо. Це значить, що тестування повинне обов'язково сполучатися з іншими традиційними формами і методами перевірки, у такий спосіб результати тестування будуть максимально точними.

Для визначення ступеня правильності відповіді людська участь не потрібна, а об'єктивність оцінки – абсолютна і визначається тільки закладеними в роботу системи алгоритмами. Ці алгоритми були створені в рамках роботи над системою і характеризуються наступними властивостями:

1. Оцінка за відповідь на питання залежить від параметрів питання, таких як його складність, ступінь правильності кожного варіанта відповіді, тимчасових рамок і т.д.

2. Оцінка є не дихотомічною, а розподіленою на інтервалі від мінімальної до максимальної. Причому цей розподіл не рівномірний, а злегка зміщений убік максимальної оцінки, оскільки, виходячи з даних експериментів, викладачі звичайно не ставлять оцінки, рівні мінімальній за зовсім неправильну відповідь.

3. Підсумкова оцінка обчислюється, опираючись на оцінки кожного питання.

Відповіді на питання з вільно-конструюємих відповідей неможливо перевірити без участі людини. Але в цьому випадку перевіряючий не повинен бачити ім'я учня, відповідь якого він перевіряє, щоб уникнути необ'єктивної оцінки. Тому, у рамках даної методики, питання з вільно-конструюємою відповіддю пропонуються викладачеві анонімним набором. Необхідно помітити, що ця особливість відрізняє автоматизовану систему контролю знань від «ручної», де суб'єктивний фактор свідомо великий.

Одна з переваг автоматизованих систем контролю знань в тому, що вони можуть використати складні методики подання завдань учням, називані стратегіями тестування.

Щоб використання систем комп'ютерного тестування було виправдано, необхідне виконання декількох умов:

1. Комп'ютерне тестування повинне надавати результати не гірше, ніж його неавтоматизований аналог.

2. Оскільки будь-який вид тестування є не що інше, як спроба структурувати процес проведення іспиту або опитування, то результат комп'ютерного

тестування повинен якісно прагнути до результату іспиту. Тоді основною перевагою автоматизованого тестування буде помітне скорочення часу, відведеного на проведення іспиту, збільшення пропускної здатності самого процесу іспиту, а результат тестування можна буде вважати більш об'єктивним.

3. Для того, щоб одержати додаткові переваги перед іспитом, тестування повинне максимально абстрагуватися від людського фактору при перевірці робіт і при обчисленні оцінки. На жаль, цей момент практично не досяжний при використанні неавтоматизованого тестування, оскільки перевіряючий майже завжди знає, роботу якого учня він перевіряє, а, отже, існує ймовірність необ'єктивної оцінки. Для виходу зі сформованої ситуації необхідно залишати в тесті тільки найпростіші типи контрольних питань, а це, у свою чергу, сильно спрощує тест і знижує його якість.

Програма комплексного комп'ютерного тестування (ПККТ) для навчальних закладів повністю реалізує процес тестування та оцінювання, починаючи від складання завдань і до одержання звітів про результати. Вся інформація – тестові завдання, тести, результати, автори, експерти, учні і багато чого іншого – повинне зберігатися в єдиній базі даних і доступна в будь-який момент. Російський, український та англійський інтерфейс роблять ПККТ зручним для використання в будь-яких регіонах і для різних завдань.

Критерії валідності тесту і тестових завдань із погляду математики ПККТ повинна робити автоматично.

Основне призначення ПККТ – забезпечити повний технологічний цикл процесів оцінки знань учнів шкіл і студентів вищих навчальних закладів за допомогою методу комп'ютерного тестування.

ПККТ повинен забезпечити:

- оцінку знань абітурієнтів при вступних іспитах;
- проведення іспитів і заліків у період екзаменаційних сесій;
- проведення поточних контрольних тестів;
- навчальне тестування (з поясненнями правильних відповідей і посиланнями на джерела інформації);
- тестування у вилученому режимі (через Інтернет);
- ведення статистики, розрахунки основних статистичних показників для всіх структурних рівнів навчальних закладів;
- повний набір інструментальних засобів для створення якісних тестових завдань і тестів;
- імпорт тестів і стандартних тестових завдань інших розробників;
- експорт тестів і тестових завдань у текстові файли і файли стандартних форматів;
- простий і доступний інтерфейс;
- можливість вибору мови інтерфейсу;
- вибір видів і рівнів тестування;
- автоматичне обмеження за часом;

- можливість використання в питаннях тестових завдань зображень, звукового і відео матеріалу;
- гнучка система установки максимальної і мінімальної бальної оцінки за кожне питання;
- можливість вибору і настроювання сценарію тестування;
- збереження результатів процесу тестування;
- збереження і облік інформації про користувачів, тести, тестові завдання та результатів у базі даних.

Структура системи ПККТ повинна обов'язково включати наступний перелік підсистем, представлених на рис. 1.

ПККТ складається з 8 підсистем, кожна з яких виконує своє завдання, пов'язану з об'єктом на наведеному рисунку:

- підсистема «Адміністратор» управляє структурою всіх об'єктів системи (додавання, зміна, видалення та виведення у файл інформації про користувачів, навчальні заклади, підрозділи і студентів);
- підсистема «Редактор тестових завдань» створює базу даних тестових завдань і дозволяє редагувати вже наявні завдання (створення, зміна і збереження тестових завдань);
- підсистема «Конструктор тестів» дозволяє створювати тести на основі бази даних тестових завдань за зазначеними параметрами (створення тестів різної складності, корегування та збереження тестів у БД);
- підсистема «Тест-менеджер» керує процесом залікового тестування в реальному часі і у запланованому режимі (завантаження залікових тестів, запуск обраних тестів на виконання, спостереження за процесом тестування кожного студента з комп'ютера викладача);
- підсистема «Тестування» проводить залікові, тренувальні та навчальні тестування, а також дозволяє проводити вилучене тестування через глобальну мережу Інтернет (вибір залікового, тренувального або навчального режиму тестування);
- підсистема «Аналізатор» обробляє результати залікового тестування і визначає параметри тестів, а також створює звіти за результатами обробки (розрахунок параметрів тестів і результатів студентів, шкалювання результатів, подання результатів тестування);
- підсистема «Статистика» дає можливість одержати зведені дані про результати тестування, згруповані за структурними підрозділами навчального закладу: факультетами, групами, класами і т.п. (відбір і угруповання результатів тестування за тестами, структурними підрозділами і студентами з наступним збереженням);
- підсистема «Результати студента» дає можливість одержати інформацію про всі дані результати, що перебувають у базі, тестування кожного студента (відбір всіх результатів тестування студента по всіх тестах).

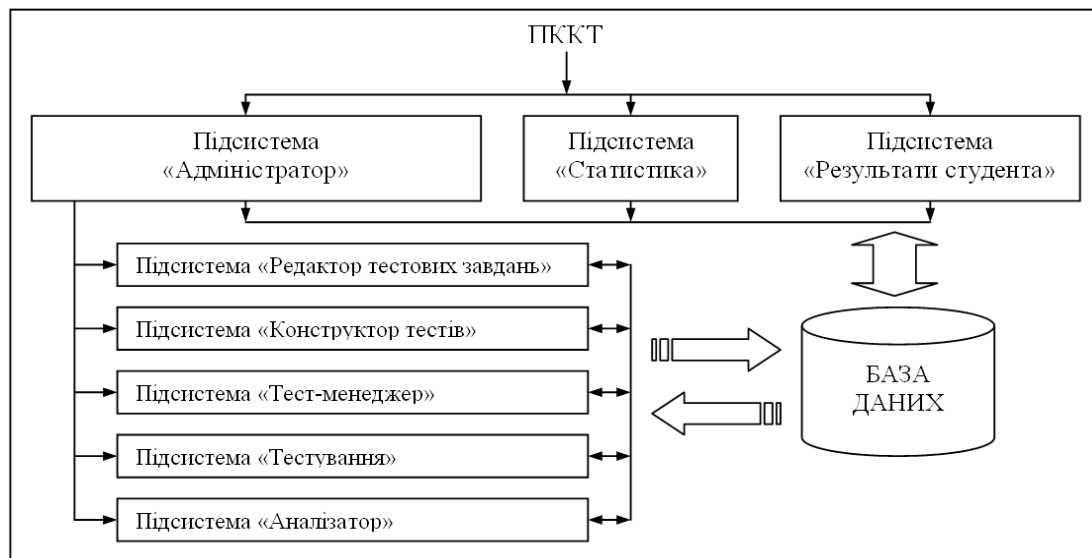


Рис. 1. Структура програми комплексного комп'ютерного тестування (ПККТ)

Висновки

Створення якісних з погляду математичних вимог тестів забезпечується технологічним циклом процесу, починаючи від створення тестових завдань і закінчуючи аналізом результатів тестування. Після визначення параметрів тестів цикл, у якому усуваються виявлені недоліки тестових завдань і тестів, повторюється доти, поки параметри тесту не будуть задовольняти пропонованим до нього вимогам.

Таким чином, автоматизована система контролю знань дозволить забезпечити всі необхідні функціональні можливості для розробки тестових завдань і тестів, організації проведення індивідуального і масового тестування, у тому числі в системі дистанційного утворення, адекватно оцінити рівень підготовки учнів, що пройшли тест, оцінити якість самого тесту, а також окремих тестових завдань.

Список літератури

1. Коноваленко О.Є. Концепція створення і використання інформаційного порталу тестування та оцінювання знань / О.Є. Коноваленко, В.О. Брусенцев, В.О. Ярута // Системи обробки інформації. – Х.: ХУПС, 2011. – Вип. 8 (98). – С. 307-309.

2. Коноваленко О.Є. Концепція використання багатоагентних систем управління в дистанційному навчанні / О.Є. Коноваленко, В.О. Брусенцев // Системи обробки інформації. – Х.: ХУПС, 2010. – Вип. 9 (90). – С. 153-156.

3. Бондаренко М.Ф. Оценивание тестовых заданий разных типов и определение их уровня сложности / М.Ф. Бондаренко, В.В. Семенец, Н.В. Белоус, И.В. Куцевич, И.А. Белоус // Искусственный интеллект. – 2009. – № 4. – С. 322-329.

4. Изотова Н.В. Повышение качества обучения в вузе средствами корректирующего контроля / Н.В. Изотова // Развитие внутривузовской системы обеспечения качества образования. – Армавир: АГПУ, 2004. – С. 96-100.

5. Чельщикова М.Б. Теория и практика конструирования педагогических тестов [Текст] / М.Б. Чельщикова. – М.: Логос, 2002. – 432 с.

Надійшла до редколегії 23.12.2011

Рецензент: д-р техн. наук, проф. В.А. Краснобаєв, Полтавський національний технічний університет ім. Ю. Кондратюка, Полтава.

ПРОБЛЕМА АВТОМАТИЗИРОВАННОГО ТЕСТИРОВАНИЯ И ОЦЕНИВАНИЯ ЗНАНИЙ

О.Е. Коноваленко, В.А. Брусенцев

Рассматриваются и анализируются проблемы внедрения современных автоматизированных систем тестирования. Предлагается подход к организации тестирования, связанный с применением методов интеллектуальных вычислений и инженерии знаний. Формулируются принципы интеллектуального тестирования, положенные в основу нового класса программных систем контроля и оценивания знаний.

Ключевые слова: тестирование, образование, дистанционное обучение, обработка информации.

PROBLEM OF THE AUTOMATED TESTING AND ESTIMATION OF KNOWLEDGE

O.E. Konovalenko, V.A. Brusencev

Are considered and analyzed problems of introduction of the modern automated systems of testing. Is offered the approach to the testing organization, connected with application of methods of intellectual calculations and engineering of knowledge. Are formulated the principles of intellectual testing taken as a principle new class of program monitoring systems and estimation of knowledge.

Keywords: testing, education, distance teaching, treatment of information.