

УДК 37.09

МЕТОДЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ В СИСТЕМАХ КОМПЬЮТЕРНОГО ТЕСТИРОВАНИЯ



Е.И. Лещевич
магистрант БГУИР,
инженер кафедры
электронной техники и
технологии БГУИР



П.В. Камлач
заместитель декана
факультета
компьютерного
проектирования,
кандидат технических
наук, доцент



И.И. Ревинская
аспирант бгуир,
ассистент кафедры
электронной техники
и технологии БГУИР



В.М. Бондарик
декан факультета
доуниверситетской
подготовки и
профессиональной
ориентации,
кандидат
технических наук,
доцент

Белорусский государственный университет информатики и радиоэлектроники,
Республика Беларусь
Email: e.leshchovich@bsuir.by, kamlachpv@bsuir.by

Е.И. Лещевич

Окончил Белорусский государственный университет информатики и радиоэлектроники. Магистрант кафедры инженерной и компьютерной графики БГУИР. Область научного интереса – методы обработки результатов компьютерного тестирования.

П.В. Камлач

Окончил Белорусский государственный университет информатики и радиоэлектроники. Доцент, кандидат технических наук, заместитель декана факультета компьютерного проектирования, доцент кафедры электронной техники и технологии БГУИР. Область научного интереса – проектирование медицинских электронных систем.

И.И. Ревинская

Окончила Белорусский государственный университет информатики и радиоэлектроники. Аспирант и ассистент кафедры электронной техники и технологии БГУИР. Область научного интереса – методы регистрации и обработки биомедицинских сигналов.

В.М. Бондарик

Образование: 1983-1988 – Минский радиотехнический институт, специальность «Конструирование и производство радиоаппаратуры», квалификация – инженер-конструктор-технолог. Область научного интереса – проектирование медицинских электронных систем, внедрение дистанционных образовательных технологий.

Аннотация. В работе рассмотрены методы компьютерного тестирования, методы оценки надежности результатов тестирования. Предложены система компьютерного тестирования и способ оценки знаний тестируемых, позволяющие оценивать вероятность использования внешних источников информации при ответах на вопросы теста.

Ключевые слова: надежность, тестирование, приложение.

Введение.

В современном образовании компьютерное тестирование широко используется как средство определения качества знаний у учащихся. В отличие от классических форм контроля знаний, компьютерное тестирование более эффективный инструмент определения результатов, так как тестирование повышает технологичность процедуры проверки: обеспечивает автоматическую проверку знаний, повышает объективность оценки, сокращает время проверки.

Технологии BigData находят широкое применение в разработке систем компьютерного тестирования, которые включают в себя применение баз данных, обработку результатов тестирования, статистический анализ полученных ответов, оценку достоверности полученных результатов. С увеличением количества вопросов, тестов и тестируемых людей, увеличивается время обработки тестов и возникает вопрос определения надежности результатов тестирования.

Система компьютерного тестирования – это информационная система, предназначенная для проверки знаний в рамках учебного процесса. Ее состав может быть определен, исходя из структуры любой информационной системы, например, включать следующие модули: система проведения тестирования, модуль аутентификации, база участников, протокол тестирования, модуль проверки, статистический анализ, шкала оценок, база тестируемых. Системы компьютерного тестирования позволяют освободить преподавателя от рутинной работы при проведении экзаменов и промежуточной оценке знаний в традиционном учебном процессе, а при обучении с использованием дистанционных технологий становятся основным средством контроля тестов.

Еще одними из основных преимуществ использования компьютерного тестирования являются возможность автоматизации обработки результатов, объективность контроля и быстрая проверка качества подготовки большого числа тестируемых по широкому кругу вопросов. Это позволяет определить разделы, которые представляют наибольшую сложность в изучении, и, возможно, корректировать процесс обучения в зависимости от результатов тестирования. Именно в автоматизированных системах тестирования в наибольшей степени проявляются многие преимущества тестового контроля знаний (оперативность, легкость сбора статистики и пр.). Однако, большим недостатком тестового контроля является недобросовестный подход тестируемых к выбору ответов на вопросы (использование сторонних источников информации). Для этого разработана система компьютерного тестирования со структурированной базой данных, которая позволяет определять использует ли тестируемый сторонние источники информации при прохождении тестов.

Методы оценки надежности результатов тестирования.

Тестирование с помощью параллельных форм.

Наиболее популярные педагогические тесты имеют параллельные формы. По сути, в них раскрывается одно и то же содержание, но вопросы внешне разные. Для обоснования параллелизма тестов помимо смысловой и логической связи вводится условие необходимости равенства их средних арифметических и дисперсий [1]. Несмотря на то, что сохраняется то же условие для обоснования параллелизма отдельных вопросов, студентам сначала дается одна форма теста, а через некоторое время – другая. Если полученные результаты сильно коррелируют между собой, то можно утверждать об устойчивости знаний учащихся.

Раздельная корреляция.

Этот метод позволяет проверить надежность в одном тесте [2]. Весь тест, состоящий из заданий или утверждений, делится пополам. Обычная процедура для этого такова: результаты испытуемых суммируются отдельно в четных заданиях и в нечетных; суммы, полученные пополам, соотносятся. Для того чтобы узнать, каков коэффициент

достоверности теста в целом, значение коэффициента, полученного из двух половин, устанавливаются по формуле 1 Спирмена-Брауна:

$$r = \frac{k}{k-1} \cdot \left[1 - \frac{\sum p_i g_i}{St^2} \right], \quad (1)$$

где k – количество заданий в тесте,

$\sum p_i g_i$ – сумма дисперсий отдельных заданий [1],

St^2 – сумма дисперсий по всему тесту.

Полученный коэффициент корреляции свидетельствует о надежности теста в зависимости от надежности каждого отдельного пункта, поэтому его лучше назвать коэффициентом внутренней согласованности теста [3].

Надежность теста тем выше, чем меньше дисперсия составляющей ошибки измерения и больше общая дисперсия. Поэтому надежный тест лучше различает сильных и слабых учащихся: их оценки должны быть более дифференцированными. Сравнивая пятибалльную шкалу оценивания, например, с десятибалльной, как в Республике Беларусь, очевиден главный недостаток именно малого размаха и, соответственно, низкой дифференцирующей способности.

Другой причиной недостаточной надежности тестов обычно является малое количество заданий. В теории тестирования известна зависимость между количеством эквивалентных задач и надежностью. На этом основана еще одна формула (2) Спирмена-Брауна, позволившая оценить, насколько повысится надежность теста при увеличении количества заданий в n раз:

$$r = \frac{nr}{1+(n-1)r}, \quad (2)$$

где r – повышенная надежность расширенного теста;

n – кратность увеличения количества заданий в тесте.

Повторное тестирование.

Один и тот же тест проводится в одной и той же группе испытуемых дважды через определенный промежуток времени. Результаты первого и второго тестов соотносятся, полученный коэффициент указывает на достоверность, в данном случае на воспроизводимость результатов испытуемых. В качестве ошибочных компонентов здесь рассматривались изменения знаний учащихся (забывание, заучивание и т. д.), влияющие на достоверность (стабильность) результатов [3, 4].

Коэффициент надежности r , полученный методом повторных испытаний, правильнее называть коэффициентом стабильности или коэффициентом воспроизводимости, и рассчитывается по формуле (3):

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}} \quad (3)$$

где x_i, y_i – индивидуальный балл i -го испытуемого в первом и втором тестировании соответственно,

n – число вопросов в тесте.

По значению коэффициента надежности можно дать качественно-количественную характеристику связи: если $r > 0,8$, то надежность высокая, если $0,5 < r < 0,8$ – удовлетворительная надежность, $r < 0,5$ – низкая надежность.

Прокторинг.

Эта процедура наблюдения за ходом дистанционного тестирования (по-английски «проктор» – наблюдатель на экзаменах в вузе) [3]. Это понятие уже достаточно хорошо известно не только специалистам. Исследователи выделяют следующие системы прокторинга:

1. Пассивный мониторинг программного обеспечения на компьютерах испытуемых. Это делается путем мониторинга приложений, которые студенты используют на своих ПК, и переключения на другие службы во время экзамена.

2. Активное ограничение программного обеспечения на компьютерах учащихся. Например, приложение может блокировать доступ к другим приложениям во время экзаменов.

3. Пассивное видеонаблюдение с использованием программного обеспечения, которое получает доступ к веб-камере учащихся для прямой записи всех действий [1]. Онлайн-контроль проходит без экзаменатора, без участия человека. Записывается видео и звук, автоматически выявляются подозрительные действия людей и фиксируются нарушения. При грубом несоблюдении правил (замена экзаменуемого, копирование заданий) программа автоматически блокирует доступ к системе.

4. Активное видеонаблюдение: реализовано аналогично пассивному, но с добавлением мониторинга в реальном времени. Один проктор (наблюдатель) может наблюдать за несколькими студентами одновременно. Методы оценки достоверности компьютерного тестирования осуществляются на основе стереотипных поведенческих реакций – паттернов. Это определенный набор, паттерн поведенческих реакций или последовательность стереотипных действий человека по отношению к какой-либо сфере, где человек применяет паттерны. Это метод не нашел широкого распространения в виду сложности технической реализации и высокой стоимости программного обеспечения.

Обработка результатов компьютерного тестирования.

В классическом тестировании нахождение итога осуществляется за два шага: 1 шаг – проверка результатов выполнения каждого задания (по дихотомической шкале); 2 шаг – суммирование результатов по всем заданиям, определение конечного балла [5].

Эти же шаги можно выделить и при компьютерном тестировании, хотя в их реализации появляется определенная специфика. Шаг 1. Проверка выполнения отдельных заданий. Как отмечалось выше, при разработке компьютерного теста в банк заданий закладываются не только тексты заданий, но и верные (эталонные) ответы. В процессе опроса система принимает ответ учащегося и сопоставляет с эталоном. Результат сопоставления может быть оценен по-разному. При дихотомической схеме тестируемый получает за выполнение тестового задания «1», если ответ полностью совпал с эталонным, и «0» при несовпадении или частичном совпадении. Однако в программах компьютерного тестирования легко предусмотреть вычисление доли правильности ответа, величина которой находится между 0 и 1; как и ранее, результат выполнения задания i студентом j будем обозначать x_{ij} , однако, теперь смысл этой величины – доля правильности. При этом может быть установлен обязательный минимум выполнения задания, ниже которого доля принимается равной «0».

Например, для выполнения задания требуется дать 5 ответов, а минимальная доля установлена 0,6; тогда при 5-ти верных ответах доля составит 1, при 4-х – 0,8; при 3-х – 0,6; при 2-х – 0,4, 1-м и невыполнении – 0. Понятно, что оценка по доле правильности более точно отражает учебные достижения тестируемого.

В общем случае минимальная доля правильности может быть настраиваемым параметром схемы оценивания, однако, с нашей точки зрения вполне уместно установить его (и «зашить» в программу) равным именно 0,6 – это соответствует мировой практике, когда экзамен считается сданным, если студент набирает не менее 60 баллов из 100.

Шаг 2. Подсчет итогового значения

а) Доля правильности выполнения теста R_j рассчитывается по формуле:

$$R_j = \frac{1}{N} \sum_{i=1}^N x_{ij}, \quad (3)$$

где N – число заданий в тесте;

x_{ij} – доля правильности выполнения задания номер j студентом номер i ;

R_j – доля выполнения студентом j всего теста.

При использовании данного варианта по умолчанию принимается, что значимость всех заданий (и, следовательно, их вклад в результат) одинакова.

б) Усреднение долей выполнения заданий с учетом их трудности. Пусть назначенная трудность задания $p_i^{(D)}$ – именно она может служить весовым коэффициентом, отражающим значимость задания. Тогда долю выполнения теста можно найти как взвешенное среднее:

$$R_j = \frac{\sum_{i=1}^N x_{ij} p_i^{(D)}}{\sum_{i=1}^N p_i^{(D)}}. \quad (4)$$

Формулы (3) и (4), по сути, соответствуют формулам (1) и (2), но с нормировкой не на 1, а на максимально возможное количество баллов за тест (N). В формуле (4) – простое суммирование долей правильности выполнения всех заданий студентом j дает суммарный балл:

$$X_j = \sum_{i=1}^N x_{ij}. \quad (5)$$

Снова по умолчанию предполагается одинаковая значимость всех заданий (4) суммированием с учетом весов:

$$X_j = N \cdot \frac{\sum_{i=1}^N x_{ij} p_i^{(D)}}{\sum_{i=1}^N p_i^{(D)}} = N \cdot R_j. \quad (6)$$

Система компьютерного тестирования.

Разработанная система представляет собой приложение, которое открывается в окне веб-браузера. Интерфейс имеет достаточно простой и понятный дизайн. Студент запускает

приложение, вводит свои данные (логин, пароль) выбирает тест, который необходимо пройти и начинает работать. Окно программы представлено на рисунке 1.

Для того чтобы определить вероятность, когда тестируемый использует сторонние источники информации при ответах на вопросы теста, было целесообразно отслеживать поведение тестируемого путем фиксации данных о перемещениях по экрану и простоях мыши, а также времени ответа на вопросы теста.

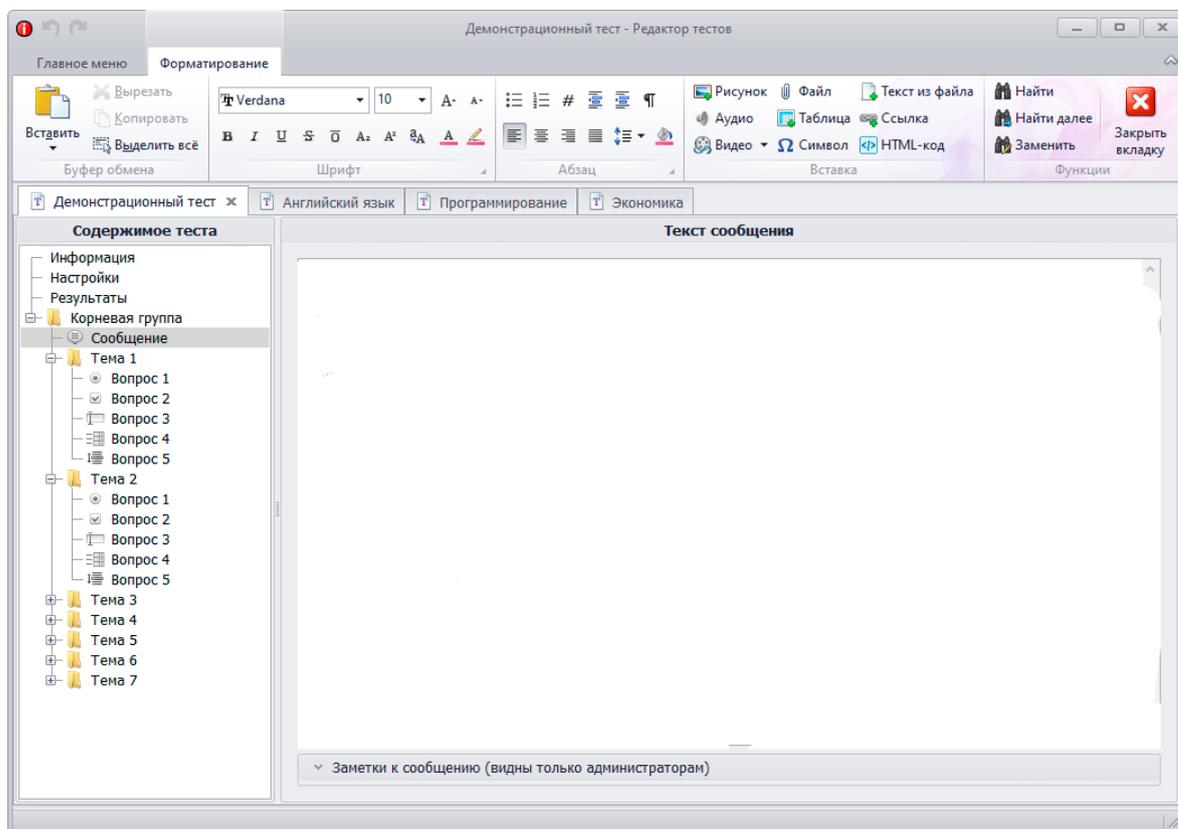


Рисунок 1. Интерфейс программы

В систему было встроено два таймера: первый отсчитывает время ответа на вопросы теста, а второй отсчитывает время бездействия мыши во время ответа на каждый вопрос. Принцип его работы заключается в проверке изменения координат мыши с интервалом 0,25 секунды. Если за данный период времени координаты местоположения мыши не изменились, то таймер начинает отсчет и запись времени бездействия. Как только местоположение манипулятора изменяется, полученный интервал (время бездействия) записывается в массив данных, и операция повторяется снова. По завершении прохождения испытуемым теста, формируется отчет с данными, собранными за это время, с целью определения вероятности использования испытуемым внешних источников информации.

Работа системы заключается в том, чтобы анализировать поведение испытуемого во время прохождения теста и проверять с какой вероятностью испытуемый использовал внешние источники информации при ответах на вопросы теста. Итоговая вероятность использования внешних источников информации записывается как процентное соотношение количества вопросов, на которые испытуемый предположительно отвечал с использованием внешних источников информации к общему количеству вопросов в тесте, с указанием номера темы, вопрос из которой вызвал затруднения при ответе.

Заключение.

Разработана система компьютерного тестирования, которая позволяет оценивать

знания тестируемых и вероятность использования внешних источников информации при ответах на вопросы теста. При этом вероятность использования сторонних источников информации определялась по времени бездействия мыши. Предложенная система будет совершенствоваться в дальнейшем, так как для объективной оценки достоверности результатов тестирования и оценки вероятности использования сторонних источников необходимо учитывать больше факторов.

Список использованных источников

[1] Methods of using information and communication technologies in the educational process: a study guide. Part 1: Conceptual foundations of computer didactics / Ural. state ped. un-t, Institute of Informatics and Inform. Technologies, Dept. new inform. technologies in education. - Yekaterinburg: [b. i.], 2013.

[2] Mayorov, A. N. Theory and practice of creating tests for the education system / A. N. Mayorov. - М. : Intellect-center, 2001. - 296 p.

[3] Kim, V. S. Testing educational achievements. Monograph / V. S. Kim. - Ussuriysk: UGPI Publishing House, 2007. - 214 p.

[4] Бараз, В.Р. Использование MS Excel для анализа статистических данных : учеб. пособие / В. Р. Бараз, В. Ф. Пегашкин; М-во образования и науки РФ; ФГАОУ ВПО «УрФУ им. первого Президента России Б.Н.Ельцина», Нижнетагил. техн. ин-т (филиал). - 2-е изд., перераб. и доп. - Нижний Тагил : НТИ (филиал) УрФУ, 2014. – 181 с.

[5] Аванесов, В. С. Основы научной организации педагогического контроля в высшей школе / В. С. Аванесов. – М. : НИТУ «МИСиС», 1989. – 167 с.

APPLICATION OF BIGDATA TECHNOLOGIES IN THE DEVELOPMENT OF COMPUTER TESTING SYSTEMS

E.I. LESHCHEVICH
*Master student of
BSUIR, engineer of the
Department of Electronic
Engineering and
Technology of BSUIR*

P.V. KAMLACH
*PhD, Deputy Dean of the
Faculty of Computer
Design*

I.I. REVINSKAYA
*Postgraduate student of
the BSUIR, assistant of
the Department of
Electronic Engineering
and Technology of
BSUIR*

V. M. BANDARIK
*Dean of the faculty of
pre-University training
and vocational
guidance, candidate of
technical Sciences*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus
Email: e.leshchevich@bsuir.by, kamlachpv@bsuir.by*

Abstract. The paper considers methods of computer testing, methods for assessing the reliability of test results. A system of computer testing and a method for assessing the knowledge of the tested are proposed, which allow assessing the likelihood of using external sources of information when answering test questions.

Keywords: reliability, testing, application.