

разработки системы проведения тестов применялись следующие средства разработки: Flash 5, PHP, HTML.

Программа наполнения и просмотра результатов проведения тестов предназначена для преподавателя. Она предоставляет возможность изменять уже созданные тесты, путем добавления или изменения вопросов, создавать новые тесты, а также просматривать результаты прохождения различных тестов студентами. Программа разработана с помощью среды Borland C++ Builder. Для подключения к базе данных использовалась технология ADO.

В системе используется база данных в формате Microsoft SQL Server. Она размещается на сервере и используется при проведении тестирования, добавлении/изменении/удалении тестов, и при просмотре результатов прохождения студентами различных тестов.

—  —

Методика оценивания в компьютерной системе тестирования знаний

Шкиль А.С., Чумаченко С.В., Напрасник С.В.

Харьковский национальный университет радиоэлектроники,

Харьков, Украина

E-mail: ri@kture.kharkov.ua

Abstract

The issues of knowledge evaluation in a learning testing system are considered. The developed system uses questions with several numbers of alternate answers in one seance. The formula of range limits calculation of rating scale is offered. It depends on a number of questions in a testing seance, a number of alternatives in each question and a number of marks in a rating scale. The factor of random selection of answers is taken into account in the proposed formula.

Одним из неотъемлемых компонентов дистанционного обучения является компьютерная система тестирования знаний [1]. Основными составляющими такой системы являются база тестовых заданий, база тестируемых с результатами тестирования и статистикой, студия разработки тестов, интерфейс общения с тестовой системой, а также правила допуска и процедура прохождения тестирования, способы оценивания и обработки статистики. С точки зрения технологий программирования система тестирования знаний должна обладать платформенной независимостью, работать в сетевом варианте и, желательно, использовать свободно распространяемые программные продукты.

Разрабатываемая в настоящее время на кафедре АПВТ ХНУРЭ система дистанционного тестирования знаний обучаемых (СДТ UNITEST) построена на

основе Web-технологий и может работать как в корпоративной сети (INTRANET), так и в глобальной сети INTERNET.

Использование и администрирование системы UNITEST полностью реализовано с помощью Web-интерфейса, что дает ряд существенных преимуществ:

- использование гипертекста обеспечивает связность разнородной информации;
- браузер (клиентская программа WWW-сервера) дает единый, достаточно простой, быстрый и привычный интерфейс пользователя;
- Web-интерфейс позволяет осуществлять удаленное администрирование.

Студия разработки тестов реализована полностью на Web-технологиях. С ее помощью возможно создание вопросов с использованием графики, таблиц, анимации, спецсимволов и т.д. В студии заложена возможность структурирования теста на темы и подтемы, что позволяет равномерно включать в сеансы тестирования вопросы из разных разделов предметной области. Тесты, созданные в студии, легко переносимы, что обеспечивается формированием XML-файла, который позволяет быстро создать тест на другом сервере или использоваться в качестве резервной копии.

Подсистема регистрации состоит из следующих компонентов.

- Редактор пользователей позволяет создавать группу тестируемых как через Web-интерфейс, так и при помощи plaintext файла. Один тестируемый может находиться одновременно в нескольких разных группах.
- Модуль допуска позволяет преподавателю выбрать из группы тестируемых тех, кто допущен к прохождению теста. Здесь же создаются пароли для прохождения теста. Метод, по которому создаются пароли, преподаватель определяет сам.
- Генератор паролей позволяет создавать для определенной группы тестируемых файл RTF-формата с паролями.

Подсистема оценивания и обработки статистики формирует оценку тестируемого за сеанс, нормирует ее относительно выбранной шкалы оценивания, формирует таблицы результатов тестирования для каждого пользователя и каждого вопроса. Выводится подробный отчет о том, как отдельный тестируемый или группа отвечали на заданные вопросы. Генератор ведомостей позволяет создавать стандартную ведомость с результатами тестирования для определенной группы в формате RTF.

Особенностью архитектуры клиент-сервер является использование выделенных серверов баз данных (БД), понимающих запросы на языке структурированных запросов SQL и выполняющих поиск, сортировку и агрегирование информации на месте без излишней «перекачки» данных на рабочие станции. Клиенты серверов БД получают последовательно и порциями только результаты запросов. База данных тестовой системы UNITEST создана при помощи СУБД MySQL. Используются поля фиксированной длины, что значительно повышает производительность системы.

При создании СДТ UNITEST были использованы такие программные продукты.

PHP – язык написания сценариев, внедренный в HTML. PHP является интерпретируемым языком программирования, поэтому не возникает проблем с переносом программ из одной операционной системы в другую.

MySQL является небольшой и быстрой реляционной СУБД. Ее преимуществами является многопоточность, поддержка нескольких одновременных запросов, записи фиксированной и переменной длины. Все данные хранятся в формате ISO8859_1.

Apache – самый распространенный в мире Web-сервер. Apache по своим функциональным возможностям и надежности не уступает коммерческим серверам, а широкие возможности конфигурирования позволяют настроить его для работы практически с любой операционной системой.

Все эти продукты являются OpenSource проектами, поэтому они бесплатные и свободные для распространения. Данные средства являются мультиплатформенными, т.е. они будут одинаково успешно работать под операционными системами Windows, UNIX и другими.

Структура системы тестирования является модульной, что позволяет подключать модули с дополнительными функциями, изменять режимы работы отдельных модулей, управляя системой настроек и не затрагивая при этом остальную ее часть. Например, с помощью настроек можно изменять язык интерфейса системы, процедуры допуска и прохождения тестирования, способ оценивания и диапазон шкалы оценивания и т.д.

В компьютерных системах тестирования знаний особое внимание должно уделяться объективному оцениванию результатов, так как, во-первых, тестирование обычно происходит в отсутствие преподавателя и составить объективную оценку об уровне знаний и личности обучаемого преподаватель не может, и, во-вторых, оспорить результат компьютерного тестирования обучаемому достаточно сложно. Существует достаточно много способов формирования оценки результатов тестирования. Одним из способов оценивания знаний может быть следующий.

В разработанной системе тестирования знаний каждый обучаемый проходит сеанс тестирования, состоящий из фиксированного и заранее определенного количества N вопросов разных типов. По умолчанию в настройках $N = 20$. Каждому вопросу ставится в соответствие максимальный балл за правильный ответ B_i^{\max} и фактически набранный балл за ответ B_i . В выборочных вопросах при разном количестве альтернатив фактически набранный балл за частично правильный ответ может быть меньше максимального. Оценка за сеанс тестирования определяется по формуле

$$B = \left(\frac{\sum_{i=1}^N B_i}{\sum_{i=1}^N B_i^{\max}} \right) \cdot 100\% ,$$

где N – количество вопросов в сеансе.

Далее полученный процент правильных ответов нормируется относительно выбранной шкалы оценивания, D – диапазон шкалы оценивания. Например, традиционная шкала оценивания, используемая в ВУЗах Украины, имеет диапазон $D = 5$. Чтобы набранный процент правильных ответов в

интервале от 0 до 100% перевести в численные значения диапазонов шкалы оценивания, необходимо определить границы C_k интервалов шкалы. Для первого диапазона $k=1$ нижняя граница интервала равна 0, а для последнего диапазона при $k=5$ верхняя граница будет 100%. В литературных источниках описываются разные способы выбора границ интервалов диапазонов шкалы оценивания [2,3], но обычно они не связаны с главной характеристикой любого выборочного типа вопроса – количеством предлагаемых альтернативных вариантов ответов m (в дальнейшем просто “альтернатив”). Выбирая случайно ответы на выборочные вопросы, тестируемый, например, при $m=2$ может набрать 50% правильных ответов, не имея знаний в предметной области.

В разработанной системе тестирования знаний для оценивания результатов принята «плавающая» шкала, которая зависит от числа альтернатив выборочных вопросов. Если рассматривать сеанс тестирования из N вопросов с m альтернативами ответов каждый как испытания Бернулли, то в каждом испытании вероятность наступления желаемого события (правильный ответ) равна $1/m$ и вероятность неудачи (неправильный ответ) $(1-1/m)$. Учитывая, что частота наступления удачи мало отличается от его вероятности, в шкале оценивания можно выделить интервал случайного попадания при наугад выбранных ответах и определить его верхнюю границу $R=1/m$. Таким образом, это будет первый интервал шкалы, соответствующий оценке «очень плохо», и не участвующий в фактическом оценивании знаний. Остальные диапазоны, определяемые как $(1-R)$, являются оцениваемыми диапазонами шкалы. Примем, что диапазоны оцениваемой части шкалы имеют равную длину, что соответствует равновероятному распределению баллов за ответ в оцениваемом диапазоне. Такой подход традиционно используется в простых схемах оценивания [2]. Таким образом, оцениваемая зона шкалы ($D-1$ интервалов) делится на $(k-1)$ отрезков, где k – номер интервала. Верхняя граница k -го интервала может быть вычислена по формуле:

$$C_k = \left[R + (1-R) * \frac{k-1}{D-1} \right] * 100\% = \left[\frac{1}{m} + \left(1 - \frac{1}{m}\right) * \frac{k-1}{D-1} \right] * 100\%. \quad (1)$$

Нормирование полученной оценки B за сеанс относительно выбранной шкалы оценивания осуществляется с использованием неравенства:

$$C_{k-1} < B \leq C_k. \quad (2)$$

Отметим, что $C_k = \frac{1}{m} * 100\%$ при $k=1$ и $C_k = 100\%$ при $k=D$, что соответствует выбранному принципу оценивания.

Формула (1) верна для сеанса тестирования, в котором используются вопросы с одинаковым числом альтернатив, но в реальности обычно авторы предлагают вопросы с различным числом альтернатив. Не учитывая различной сложности вопросов с разным числом альтернатив общую оценку за сеанс можно получить путем усреднения оценок для групп вопросов с равным числом альтернатив, т.е.

$$B = \sum_j (B_j * h_j) / \sum_j h_j = \sum_j (B_j * h_j) / N,$$

где h_j – количество вопросов с j -ым числом альтернатив.

При выбранной методике оценивания усреднение оценок фактически сводится к усреднению границ диапазонов C_k в формуле (1). С учетом того, что в (1) k и D не зависят от количества вопросов с разным числом альтернатив, получим:

$$C_k = \left[1 / \left(\sum_j (m_j * h_j) / N \right) + \left(1 - 1 / \sum_j (m_j * h_j) \right) * \frac{k-1}{D-1} \right] * 100\%. \quad (3)$$

Выражение $m_{cp} = \sum_j (m_j * h_j) / N$ можно определить как среднюю

альтернативность сеанса тестирования, которая вычисляется для каждого сеанса в оперативном режиме.

В качестве примера рассмотрим следующую задачу: определить, какой оценке в пятибалльной шкале оценивания соответствует 75% правильных ответов при наличии в сеансе тестирования из двадцати вопросов 2 вопроса с двумя альтернативами, 3 вопроса с тремя альтернативами и 15 вопросов с четырьмя альтернативами.

Средняя альтернативность сеанса для этого случая будет:

$$m_{cp} = (2 * 2 + 3 * 3 + 4 * 15) / 20 = 73 / 20 = 3,65.$$

По формуле вычислим верхние границы диапазонов шкалы оценивания:

$k=1, C_k=27\%$; $k=2, C_k=45\%$; $k=3, C_k=63\%$; $k=4, C_k=81\%$; $k=5, C_k=100\%$.

Таким образом, 75% правильных ответов соответствует оценке 4 ("хорошо") традиционной пятибалльной шкалы.

Испытание предложенной методики оценивания результатов тестирования проводилось на факультете КИУ ХНУРЭ на трех студенческих потоках (1-й курс – 200 и 150 студентов, 3-й курс – 120 человек). Всего было протестировано более 300 студентов. Для каждой из трех учебных дисциплин в базе данных находилось от 150 до 200 вопросов типа «выбор одного из нескольких» и «выбор нескольких из нескольких». Сеанс тестирования – 20 вопросов, выбор вопросов для сеанса осуществлялся случайным образом. В ходе тестирования данная методика оценивания показала свое удобство и эффективность.

Не претендуя на полноту и универсальность, методика вычисления по формуле (3) границ интервалов диапазонов шкалы оценивания достаточно хорошо позволила решить главную, с точки зрения авторов, проблему оценивания сеанса тестирования из выборочных вопросов – получение обоснованной «отрицательной» оценки («очень плохо» и «плохо») при случайном выборе альтернатив тестируемыми в условиях отсутствия знаний.

Литература

1. Кривуля Г.Ф., Пиженко И.Н., Шкиль А.С. Проверка знаний при дистанционном обучении // Образование и виртуальность-2001. – Харьков-Ялта: УАДО, 2001. – С.212-219.
2. Тест как инструмент измерения уровня знаний и трудности заданий в современной технологии обучения: Учебное пособие / Н.М.Оленик. Донецк, ДонГУ, 1991. – 66с.
3. Антофий Н.М. Шкалирование результатов тестового контроля. // АСУ и приборы автоматки, 2002. Вып. 118. – С. 119-121.



Модель системы компьютеризированного адаптивного тестирования

Печенежский Н.А., Маслов А.А.

Харьковский национальный университет радиозлектроники,

Харьков, Украина

E-mail: mpechen@cc.jyu.fi

Abstract

The majority of present day testing systems are static by their nature in a sense of test battery structure, and therefore the provided tests are not personalized to an examinee. In this paper Item Response Theory based model for computerized adaptive testing is considered, and its advantages and limitations are discussed. Then we present an idea of integration of IRT based model with the methodology of expert system design.

Постановка проблемы

Образование в целом, как и высшее образование – это саморазвивающаяся система, имеющая внутренние механизмы непрерывного развития. Одним из самых важных механизмов такого развития является деятельность по педагогическому контролю качества образования.

Тестовый контроль – один из перспективных методов объективной оценки знаний и способностей учащихся. Однако необходимо признать, что тестирование все еще не играет значительной роли в совершенствовании системы образования, хотя его потенциальные возможности осознаны, а прогресс в вычислительной технике и телекоммуникациях за прошедшие 10 лет дал качественно новые технические возможности для его применения в учебном процессе.