

Министерство образования и науки Республики Казахстан
Северо-Казахстанский университет
им. М. Козыбаева

УДК 621.391

На правах рукописи

КАЗАНБАЕВА АЛЬБИНА СОВЕТОВНА

**Разработка интеллектуальной системы оценки результатов
обучения.**

6D075100 «Информатика, вычислительная техника и управление»

Диссертация на соискание ученой степени
доктора философии (PhD)

Научные консультанты:
кандидат физико-математических наук, профессор
Куликов В.П.

кандидат технических наук, доцент
Крылова Е.М.

Республика Казахстан
Петропавловск, 2021 г.

СОДЕРЖАНИЕ

НОРМАТИВНЫЕ ССЫЛКИ	4
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	5
ВВЕДЕНИЕ	6
1 НЕПРЕРЫВНЫЙ КОНТРОЛЬ КАК НЕОБХОДИМОЕ УСЛОВИЕ КАЧЕСТВЕННОЙ ОЦЕНКИ УЧЕБНЫХ ДОСТИЖЕНИЙ	10
1.1 Основные термины и определения предметной области	10
1.2 Обзор и анализ существующих программ/систем тестирования	16
1.3 Методы(протоколы) и модели диагностических средств оценки результатов	20
1.4 Постановка задачи диссертационного исследования	30
1.5 Выводы по первой главе	31
2 МОДЕЛЬ УПРАВЛЕНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМОЙ ОЦЕНКИ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ	32
2.1 Анализ и сравнение результатов тестирования	32
2.2 Статистическая проверка, значимость и мощность выводов	46
2.3 Особенности работы интеллектуальной системы оценки результатов обучения	49
2.4 Оценивание результатов обучения методом адаптивной оценки знаний	52
2.5 Выводы по второй главе	58
3 ОЦЕНКА РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПОСРЕДСТВОМ ПРИМЕНЕНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ	59
3.1 Принципиальная структура интеллектуальной системы	59
3.2 Процедуры управления процессами интеллектуальной системы оценки результатов обучения	60
3.3 Банк тестовых заданий	67
3.4 Выводы по третьей главе	76
4 ПРОЕКТИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ОЦЕНКИ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ И ПРОВЕДЕНИЕ АПРОБАЦИИ	77
4.1 Принцип управления организацией системы оценки результатов обучения	77
4.2 Применение устного/письменного экзамена в интеллектуальной системе	80
4.3 Энтропийный подход как характеристика управления системой обучения и аттестации	85
4.4 Моделирование интерфейса и разработка оценочных заданий для работы в системе ИСОРО	94
4.5 Иллюстрация результатов функционирования ИСОРО	100
4.6 Выводы по четвертой главе	106
ЗАКЛЮЧЕНИЕ	107
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	109

ПРИЛОЖЕНИЕ А	115
ПРИЛОЖЕНИЕ Б.....	116
ПРИЛОЖЕНИЕ В	117

НОРМАТИВНЫЕ ССЫЛКИ

В настоящей диссертации использованы ссылки на следующие документы и стандарты:

1. «Закон Республики Казахстан от 27 июля 2007 года №319-III «Об образовании» (с изменениями и дополнениями на 19.04.2019 г.)»;
2. «Закон Республики Казахстан от 7 января 2003 года №370 «Об электронном документе и электронной цифровой подписи» (с изменениями и дополнениями от 13.11.2019 г.)»;
3. Государственный общеобязательный стандарт образования, «Приказ Министра образования и науки» Республики Казахстан от 31 октября 2018 года №604».
4. «Государственная программа развития образования в Республике Казахстан на 2020-2025 гг.».
5. «Об утверждении Государственной программы «Цифровой Казахстан», Постановление Правительства Республики Казахстан №827 от 12 декабря 2017 года».

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

БД – база данных;
БЗ – база знаний;
БП – база правил;
БТЗ – банк тестовых заданий;
БПО – базовое программное обеспечение;
ВОУД – внешняя оценка учебных достижений;
ВТ – вычислительная техника;
ВУЗ – высшее учебное заведение;
ГОСО – Государственный общеобразовательный стандарт образования;
ЕНТ – единое национальное тестирование;
ИСОРО – интеллектуальная система оценки результатов обучения
МОН РК – Министерство образования и науки Республики Казахстан;
ПЗНГ – проверка значимости нуль-гипотезы;
ПК – персональный компьютер;
РК – Республика Казахстан
СКЗ – система контроля знаний;
СТКЗ - система тестового контроля знаний;
СУБД – система управления базой данных;
ТЗ – тестовое задание;
ТКЗ – тестовый контроль знаний.

ВВЕДЕНИЕ

Актуальность работы. Современный «уровень развития Казахстана ориентирован на ускоренное продвижение государства в сообщество 50-ти наиболее конкурентоспособных стран мира» [1]. «В связи с этим необходимо придать импульс развитию системы высшего образования для подготовки специалистов новых отраслей в рамках указанной экономики. Для этого необходимо обновить профессиональные стандарты в соответствии с требованиями рынка труда и лучшим мировым опытом в области обучения на рабочем месте» [1].

Однако кризис, связанный с пандемией COVID-19 внес свои коррективы в процесс развития системы высшего образования. Ускорился переход к более широкому онлайн обучению, что потребовало разработки новых решений для онлайн управления и оценки результатов обучения. Проблемы возникшие в сфере образования связаны в первую очередь с обеспечением непрерывности всего образовательного процесса, а также, в частности, сохранения целостности и эффективности онлайн-тестирования. Кроме того, многие системы оценки результатов обучения, применяемые в вузах, стали недоступными или невозможными для эффективного применения. Таким образом, актуальность данного исследования обусловлена необходимостью в разработки моделей и методов, позволяющих адаптировать выставление оценок и аттестации и внедрить современные технологии в процесс контроля и оценки результатов обучения обучающихся.

В связи с широким внедрением онлайн обучения, системы контроля знаний приобрели особую актуальность разработки контролирующих систем, основанных на тестировании.

Целью диссертационного исследования является разработка интеллектуальной системы, способствующей диагностированию обученности как последствия достигнутых результатов обучаемости путем применения разработанных моделей, методов и алгоритмов.

Задачи исследования. В рамках сформулированной цели были поставлены и решены следующие задачи:

- изучить и провести анализ системы оценивания образовательных результатов обучающихся, в том числе – методик и моделей оценивания знаний;
- провести анализ приемов прокторинга;
- провести анализ и сравнение методов тестирования;
- изучить и исследовать оценивание результатов обучения методом адаптивной оценки знаний;
- разработать модель интеллектуальной системы оценки и управления результатами обучения;
- разработать алгоритм интеграции готовых систем прокторинга в структуру тестовой системы;
- разработать структуру интеллектуальной системы;

- оценить возможность применения энтропийного подхода для эффективного управления ИСОРО;
- разработать модуль интеллектуальной системы оценки результатов обучения;
- провести апробацию интеллектуальной системы оценки результатов обучения.

Объектом исследования является система контроля в формате онлайн и процессы управления оцениванием результатов обучения.

Предмет исследования: Модели, методы и алгоритмы в системе контроля активности обучающихся и оценки их результатов обучения.

Методы исследования: аналитические исследования, методы теории управления, системный анализ, проектирования интеллектуальных систем, групповых экспертных оценок, математической статистики.

Теоретическую основу исследования составили труды отечественных и зарубежных ученых:

Огромный вклад в становление и развитие управления учебным процессом корифеи постсоветской цивилистики: В.П. Беспалько, П.Я. Гальперина, И.А. Зимней, Б.Г. Ильясова, Т.П. Костюковой, О.Е. Лебедева, В.В. Мартынова, В.А. Сластенина, Л.Р. Черняховской, Б.Ф. Скиннера и т.д.

В числе современных ученых, занимающихся исследованием информационной поддержки и управления оценочных результатов, следует отметить труды: В.С. Аванесова, В.А. Векслера, В.А. Короза, А.Н. Майорова, Е.Г. Матвиевской, В.В. Овчинникова, I.I. Bejar, H. Wainer.

В диссертации использованы труды отечественных, зарубежных, а также ученых дальнего зарубежья: С.Д. Даниловой, Е.А. Конопко, В.П. Куликова, В.П. Куликовой, С.И. Макарова, В.В. Мартынова, Г.М. Мутанова, Л.П. Фандоровой, Е.В. Шевчук, А.В. Шпака, Н. Bride, P. Bue, P. Godefroid, J. Julliard, N. Tillmann, S. Wise.

Научная новизна заключается в следующих положениях и результатах:

- представлена модель оценивания результатов обучения методом адаптирования оценки знаний;
- предложен энтропийный подход как индикативная характеристика процесса обучения и аттестации;
- разработан метод управления процессами интеллектуальной системы оценки результатов обучения;
- разработан модуль интеллектуальной системы оценки результатов обучения (для применения устного/письменного контроля знаний в онлайн формате);
- представлен метод иллюстрации результатов функционирования интеллектуальной системы оценки результатов обучения.

Практическая и теоретическая значимость. Основная практическая значимость исследования заключается в том, что полученные результаты могут быть использованы при разработке интеллектуальных систем тестирования, для оценки учебных достижений, по дисциплине в целом или

же по отдельной тематике дисциплины, а также при организации междисциплинарного контроля (экзамена).

Выводы и результаты теоретического исследования могут быть использованы в управлении процесса определения уровня готовности обучающихся по различным разделам: дисциплина в целом, частично/тематически дисциплина, высшее учебное заведение, факультет, специальность, и т.д. служат базой для разработки системы оценки успеваемости.

Положения диссертации, выносимые на защиту:

- модель интеллектуальной системы оценки и управления результатами обучения;
- общий алгоритм интеграции готовых систем прокторинга в структуру тестовой системы;
- структура интеллектуальной системы;
- метод управления адаптивирования оценки результатов обучения на основе энтропии;
- схема устройства адаптивного теста;
- алгоритм модуля интеллектуальной системы оценки результатов обучения;
- метод представления иллюстрация результатов функционирования ИСОРО.

Апробация работы. Результаты диссертационного исследования докладывались и обсуждались на:

- Международной научно-практической конференции «Непрерывное образование в XXI веке: проблемы, тенденции, перспективы развития», (Россия, г. Шадринск, 2016 г.);
- Международном форуме «Innovation and Global Issues в области социальных наук 2017», (Турция, г. Патара, 2017 г.);
- Седьмой Международной научно-технической конференции «Энергетика, информатика, инновации – 2017», (Россия, г. Смоленск, 2017 г.);
- Proceeding soft he II International Scientific and Practical Conference «Topical issues of science and education», (2017, Warsaw, Poland);
- Четвертой «Международной научно-практической конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук», (Тольятти, 2018 г.)[2];
- Международная научно-методическая конференция «Современные тренды непрерывного образования в России», (Россия, г. Новосибирск, 2019 г)[11].

Публикации. Основные результаты исследования нашли отражение в 13 научных работах, в том числе, в 5 статьях, опубликованных в изданиях, рекомендуемых Комитетом по обеспечению качества в сфере образования и науки МОН РК, в 6 статьях – в международных научных журналах, а 2 статьи в международном научном журнале, имеющем ненулевой импакт-фактор

(индексированном в базе данных Scopus), а также в свидетельстве об авторском праве №11321 от 10 июля 2020 года (Приложение А)[3] и актом о практическом использовании результатов диссертационной работы (Приложение Б).

Личный вклад автора.

Основные результаты теоретических и экспериментальных исследований получены автором самостоятельно. В печатных работах, которые написаны в соавторстве, соискателю принадлежит значимая роль при обобщении и анализе полученных результатов.

Структура и объем диссертации. Диссертация состоит из вводной части, основной (четыре главы), заключение, список использованных источников и приложение. Работа представлена на 117 страницах, в том числе 53 рисунка, 17 таблиц и 79 наименований библиографических источников.

Во введении аргументирован выбор темы исследования, раскрыта ее актуальность, сформулирована цель исследования, определяющие ее задачи, представлены объект и предмет исследования, раскрыты научная новизна, практическая и теоретическая значимость работы, выделены основные этапы исследования.

В первой главе комплексно рассматривается современное состояние непрерывного, необходимого контроля качественной оценки учебных достижений для определения профессиональных компетенций. Прикладным аспектом использования полученного в процессе исследования результата является возможность применения в проектировании и разработке системы оценки учебных достижений. Проведен обзор и анализ актуальных на сегодняшний день систем тестирования.

Во второй главе описаны модели интеллектуальной системы оценки результатов обучения. Проведен анализ результатов тестирования, изложены особенности работы системы.

В третьей главе представлены методы и алгоритмы управления процессами интеллектуальной системы оценки результатов обучения. Предложен алгоритм разработки системы управления тестированием.

Четвертая глава посвящена проектированию инструментальных средств системы и отдельного модуля (для применения устного/письменного экзамена в формате онлайн) оценки результатов обучения, а также к системе обучения и аттестации применен энтропийный подход. Также представлены методы иллюстрации результатов функционирования интеллектуальной системы оценки результатов обучения.

В заключении представлены результаты исследований, включающие основные выводы по итогам диссертационного исследования.

1 НЕПРЕРЫВНЫЙ КОНТРОЛЬ КАК НЕОБХОДИМОЕ УСЛОВИЕ КАЧЕСТВЕННОЙ ОЦЕНКИ УЧЕБНЫХ ДОСТИЖЕНИЙ

1.1 Основные термины и определения предметной области

В этой главе приведены определения основных понятий, используемых в современной тестологии в системах оценки знаний, обобщенные и иногда переработанные, которыми будем оперировать в настоящей работе.

«Аутентификация – это процедура проверки подлинности пользователя с помощью определенного пароля и т.д.» [4].

«Чтобы определить чью-то подлинность, можно воспользоваться тремя факторами»[4]:

1. «Пароль – то, что мы знаем (PIN-код, определенное слово, графический ключ, код для замка)» [4];

2. «Устройство –то, что мы имеем (пластиковая карта, ключ от замка, USB-ключ) [4];

3. Биометрика – понимается как совокупность методов, способствующих выявлению и учету особенностей работы на клавиатуре и манипулятором мышь, так же особенностей устной речи диктора, для уверенного опознания авторства произносимых и набираемых текстов, попадающих в БД в процессе интерактивного общения в среде ИСОРО.

«Учитывая ступень доверия и политику безопасности систем, аутентификация может считаться односторонней или взаимной. Обычно такое выполняется с поддержкой криптографических методов».

«Аутентификацию не стоит путать с авторизацией (процедурой предоставления субъекту обусловленных прав) и идентификацией (процедурой определения субъекта по его идентификатору)» [4].

Аутентификация при поддержке электронной подписи» [4]:

«Закон Республики Казахстан от 7 января 2003 года №370 (с изменениями и дополнениями от 13.11.2019 г.) «Об электронном документе и электронной цифровой подписи» [5].

1. «Простая электронная подпись – электронная подпись, которая посредством использования кодов, паролей или иных средств подтверждает факт формирования электронной подписи определенным лицом»[4].

2. «Неквалифицированная электронная подпись – электронная подпись, которая:

- получена в результате криптографического преобразования информации с использованием ключа электронной подписи;
- позволяет определить лицо, подписавшее электронный документ;
- позволяет обнаружить факт внесения изменений в электронный документ после момента его подписания;
- создается с использованием средств электронной подписи»[4].

3. «Квалифицированная электронная подпись – электронная подпись, которая соответствует всем признакам неквалифицированной электронной подписи и следующим дополнительным признакам:

- ключ проверки электронной подписи указан в квалифицированном сертификате;
- для создания и проверки электронной подписи используются средства, получившие подтверждение соответствия требованиям, установленным в соответствии с законодательными нормативами» [4].

Многоцелевая проверка пароля - одна из методов аутентификации в системе компьютерных технологий где представляется ввод идентификатора пользователя, называемый обычно «логин»[4], а пароль-некоторой конфиденциальной информации. Допустимая (эталонная) пара паролей входа находится в специальной базе данных.

«Принцип использования однократных паролей, можно разбить на следующие моменты: с использованием генератора псевдослучайных чисел, который в свою очередь является корпоративным для объекта и системы; использование временных меток совместно с единой системой времени; использование случайной базы паролей, одинаковую для субъекта и системы.

1-метод, использует генератор псевдослучайных чисел с одинаковым значением ради субъекта и для системы. Пароль, сформированный сущностью, может быть передан в систему при постепенном употреблении односторонней функции или при выполнении нового запроса на основе оригинальной информации из предыдущего запроса.

2-метод использует кратковременные метки. Примером такой технологии представляется SecurID. Он основан на использовании аппаратных ключей и синхронизации времени. Аутентификация основана на генерации случайных чисел посредством регулярных промежутков времени. Уникальный секретный ключ хранится исключительно в БД системы и в аппаратном приспособлении субъект. Когда элемент запрашивает доступ к системе, ему предлагается ввести pin-код, а также случайно выданный номер, который в данный момент отображается на аппаратном устройстве. Конструкция сравнивает подтвержденный pin-код и секретный ключ субъекта из своей базы данных и генерирует случайное число на основе параметров секретного ключа из базы данных и текущего времени. Впоследствии проверяется идентичность сгенерированного номера и номера, введенного субъектом.

3-метод основан на единой базе паролей для субъекта и системы и высокоточной синхронизации между ними. Впрочем, каждый пароль в наборе может быть использован только единожды»[7].

Из этого можно сделать вывод, что, если злоумышленником будет перехвачен пароль, используемый субъектом, который больше не является актуальным.

В сравнении с использованием одного и того же постоянного пароля, с одноразовые пароли гарантируют высокую степень защиты данных.

«Методы аутентификации, основанные на измерении биометрических параметров человека, обеспечивают практически 100% идентификацию, решая проблемы с потерей паролей и личных идентификаторов. Примерами реализации этих методов являются системы идентификации пользователей, основанные на узорах радужной оболочки, отпечатках ладоней, формах ушей, инфракрасных капиллярных изображениях, почерке, запахе, тембре голоса и даже ДНК. Новое направление – использование биометрических характеристик в смарт-платежных картах, токенах pass и элементах сотовой связи. Например, при оплате в магазине владелец карты прикладывает палец к сканеру, чтобы убедиться, что карта действительно его собственная»[4].

«Геометрия рук – эти устройства используются, когда трудно использовать сканеры пальцев из-за грязи или травм. Биологическая повторяемость геометрии рук составляет около 2%»[4].

«Глаз – эти приборы обладают высочайшей точностью. Теоретическая вероятность появления двух ирисов равна 1 из 1078» [4].

«Системы распознавания лиц, основанные на таком подходе, позволяют идентифицировать человека при определенных условиях с погрешностью не более 3%. В зависимости от метода можно идентифицировать человека на расстоянии от полуметра до нескольких десятков метров. Этот метод удобен тем, что позволяет реализовать его с помощью стандартных инструментов (веб-камера и т. д.). Более сложные методы требуют более сложного аппаратного обеспечения. Некоторые (не все) методы имеют недостаток подмены: можно произвести идентификацию, заменив лицо реального человека его фотографией» [4].

«Голосовая верификация подходит для использования в телекоммуникационных приложениях. Необходимая 16-битная звуковая карта и конденсаторный микрофон стоят менее 25 долларов. Вероятность ошибки составляет 2-5%. Эта технология подходит для проверки голосовых каналов, она более надежна, чем личный номер, установленный на частоте. Сейчас мы разрабатываем зоны для идентификации человека и его состояния по голосу: возбужденный, больной, говорящий правду, не по себе и т. д.» [4].

«Ввод с клавиатуры дает возможность при вводе пароля отслеживать скорость и интервалы между щелчками» [4].

«Идентификация – это процедура распознавания сущности по ее идентификатору (другими словами, это определение имени, имени пользователя или номера)» [4].

Аутентификация проводится при попытке входа в систему (например, операционную систему или службу электронной почты).

ID/идентификатором может быть:

- номер
- номер паспорта
- электронная почта
- номер страницы в социальной сети и т.д.

База данных (БД) – это организованная структура, предназначенная для хранения, изменения и обработки взаимосвязанной информации, преимущественно больших объемов.

База знаний является одной из важной части интеллектуальной системы. Наиболее распространенным классом подобных программ являются экспертные системы. Они предназначены для поиска путей решения задач из определенной области предмета, основанных на записях баз данных и описании ситуации пользователем.

Простые базы знаний могут быть использованы для создания высокоуровневых систем сбора данных в учреждении: документов, руководств нормативной характеристики технической поддержки. Главной целью создания подобных баз данных – является помощником менее компетентным сотрудникам найти существующие описание решения проблемы.

«Интеллектуальная система (ИС) – автоматизированная система, основанная на знаниях, или комплекс программных, лингвистических и логико-математических средств для реализации основной задачи – осуществления поддержки деятельности человека и поиска информации в режиме продвинутого диалога на естественном языке. Разработка естественно-языковых интерфейсов. Проблемы компьютерной лингвистики разрабатываются в ИИ с 1950-х гг. Данное направление охватывает также исследования методов и разработку систем, обеспечивающих реализацию процесса общения человека с компьютером на естественном языке (так называемые системы ЕЯ-общения)»[46].

Тест (англ. - test, тест, исследование) – совокупность вопросов и заданий, выдвигаемых испытуемому для измерения (диагностики) его личностных характеристик. В общенаучном смысле это короткий стандартизированный тест, направленный на получение наиболее важной информации о характеристиках конкретного объекта за короткий промежуток времени с целью определения его наличия или степени выраженности определенного свойства или качества. Они также бывают не только психологическими [6, стр. 17].

В данном случае речь идет о техническом тестировании.

Тестовое задание является частью сложного (составного) теста, в котором испытуемый выполняет самостоятельные действия во время теста, а его результат записывается в первичный протокол в виде отдельного ответа. Иными словами, это отдельная проблема (вопрос), на которую испытуемому предлагается дать отдельный ответ. Тестовое задание – краткое (неполное) контролируемое высказывание: содержит представленные элементы этого логического высказывания. Наиболее важным форматом тестового задания являются: информативность (диагностическая), трудность, дискриминативность, достоверность, валидность, надежность.

«Психометрический характер теста – это особенности, содержащий определенные модели качества для любого измерительного

психодиагностического принципа работы (теста)»[8, стр. 7]. Почти все современные авторы ссылаются на ряд психометрических свойств теста. Три характеристики: надежность, валидность и репрезентативность. Одни авторы относят к дискриминации (способность теста различать испытуемых), другие - к надежности (способность теста противостоять искажениям).

Надежность теста – это «психометрический характер теста, который защищает от фальсификации с намерением (лжи, неискренности испытуемого) или ненароком»[8, стр.7] мотивационного искажения.

Достоверность тестового задания также выражает степень его устойчивости к подделке. Достоверность измеряется путем сравнения результатов по обычным указаниям и указаниям на фальсификацию или путем вычисления корреляций между ответами на этот пункт и баллами испытуемых по специальной шкале ложности.

Надежность – устойчивость тестового задания определяется как мера устойчивости результатов этого задания при повторном тестировании.

Сложность – определяется теоретически (заранее) по результатам анализа структуры и объема задания или практически по результатам применения на большой выборке студентов.

Навык – это действие, формируемое повторением, характеризующееся высокой степенью развития и отсутствием сознательной регуляции и контроля отдельных элементов (существуют перцептивные, моторные и интеллектуальные навыки; один из вариантов - способность к обучению. На формирование умений влияют такие факторы, как мотивация и усвоение предмета, уровень его развития (наличие определенных предшествующих знаний и умений), полнота понимания содержания материала, владение которым происходит при производстве умений и др.

«Сложность «тестовых заданий – это показатель, специфичный для тестов способностей и тестов способности. Величина (параметр) сложности обычно понимается как монотонно убывающая функция процента испытуемых, давших правильный ответ на данный заданный вопрос: Чем меньшей степени испытуемые выполнили задание, тем выше степень сложности заданий»[6, стр. 22]. Очень легкие или очень сложные предметы автоматически окажутся малоинформативными. Поэтому для большая часть тестов»[6, стр. 23], (за исключением специальных адаптационных тестов) оптимальным считается средний уровень сложности. Современная психометрия настаивает на том, что ТТГ можно измерить только экспериментальным путем подачи тестовых заданий испытуемым, на которых» будет проводиться тест.

Стандартизация тестов – это «общее число экспериментальных, методических и статистических действий, обеспечивающих создание строго определенных компонентов теста (инструкции, набор заданий, порядок обработки протоколов и подсчета баллов, метод интерпретации). В частном случае стандартизация относится к сбору характерных стандартов тестирования и построению эталонной шкалы тестовых отметок»[8, стр.78].

«Тестовый балл – это количественный показатель, который рассчитывается путем обработки тестовых ответов с помощью тестового ключа по индивидуальному протоколу тестового соединения, состоящего из отдельных тестовых элементов и указывающего степень выраженности измеряемых свойств тестовой диагностической шкалы. С самого начала следует различать «сырые» (первичные) и стандартные результаты тестов. Первоначальный расчет тестового балла по ключу не позволяет сделать диагностические выводы» [7] см. предмет для конкретной диагностической категории), так как вам все равно придется сравнивать этот тестовый балл с диагностическими стандартами.

Навыки – это общие способы действия, которые позволяют выполнять действия с заданным результатом в широком диапазоне различных, изменяющихся условий. Если тот или иной навык тесно связан с определенными условиями и может быстро разрушиться при его нарушении, то навыки более универсальны и устойчивы к изменяющимся условиям деятельности. Умение базируется, как правило, на общих образно-понятийных моделях проблемы или, теории систематического формирования психических актов на общей ориентировочной основе. В отличие от специальных способностей, навыки проверяются в психодиагностике не тестами способностей, а результатами тестов, хотя в некоторых случаях граница между специальными способностями и навыками, следовательно, между тестами специальных способностей и результатами тестов, кажется искусственной, фактически чисто терминологической [8, стр. 6].

Представление – способность воспроизвести информацию в общих (основных) чертах.

Уровень освоения информации – представление, знание, умение, навык.

Система тестового контроля знаний (СТКЗ) – совокупность различных ресурсов (организационных, кадровых, методических и технических), решающих задачу ТКЗ.

Организационные ресурсы – нормативные документы (приказы, распоряжения, указания и т.д.) вузовского и кафедрального уровня, определяющие условия ее конкретного использования (включение в расписание, оплата дополнительной нагрузки, учет результатов, обязанности студентов и преподавателей и т.д.).

Эксперты – специально подготовленные преподаватели, заинтересованные во внедрении данной системы, составляющие и оценивающие тестовые задания.

Методические ресурсы – указания для преподавателей и студентов, обеспечивающие разработку и использование ТКЗ в частности, по переработке (структурированию) учебного материала, по методике разработки тестов и по методике их использования в учебном процессе.

Технические ресурсы – оборудование и программные средства (компьютерные тесты).

Контролируемое логическое утверждение (ФРЕЙМ) – минимальная доза учебного материала, имеет идентификатор (как правило, название) и логические элементы.

«Компьютерные тесты – включают в себя сбор тестовой информации в диалоговом окне между обучающимся и компьютерной техникой/системой. Тесты, включающие компьютерную отделку данных, собранные на бланках, не будут являться компьютерными» [9].

1.2 Обзор и анализ существующих программ/систем тестирования

Изучение качества обучения, его оснащения и контроля в контексте вхождения высшего образования в Болонский процесс является одной из ключевых проблем современной педагогической науки. В настоящее время проведено достаточно много исследований по различным аспектам проблемы овладения обучением. Однако вопросы мониторинга результатов обучения остаются недостаточно разработанными. В «частности, применения компьютерных технологий в практике обучающихся еще не стало предметом фундаментальных научных исследований и разработок»[10].

На сегодняшний день в условиях карантина система образования незамедлительно нуждается в модернизации и готовности перехода обучения, а также проведение итоговой аттестации в онлайн-формате. Одни из учреждений спокойно отнеслись к такому переходу, а некоторые, увы, не готовы.

В таблице 1.1 представлен сравнительный анализ имеющихся программ/систем тестирования, используемые в данное время вузами, где отображены функциональные возможности.

Хотелось бы отметить, что функциональные возможности у программ/систем практически друг от друга мало чем отличаются. Каждая из систем допускает неограниченное количество пользователей, здесь вопрос состоит в том, чтобы выдать этим пользователям допуск редактирования своих тестовых заданий без сопровождения соответствующих документов. Также важную роль играет генерация тестовых заданий, где предложенная модель системы ИСОРО выделилась тем что, интеллектуально адаптируется на предыдущие ответы и выдает соответственно более сложный или наводящий/легкий вопрос.

Таблица 1.1 Анализ имеющихся программ/систем тестирования

ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ						
№	Наименование	Пользователь/доступ	Редактор ТЗ	Установка	Генерация ТЗ	Отчеты
1	INDIGO	Количество пользователей не ограничено. Открытый доступ для самостоятельного редактирования и введение ТЗ.	Неограниченное количество вопросов; запрет пропуска вопросов; запрет возврата к пройденным вопросам; запрет завершения тестирования до ответа на все вопросы	Установка на любой ПК	Автоматическая генерация по темам и дисциплинам.	Печать и экспорт протоколов и отчетов
2	Let's test	Количество пользователей не ограничено. При приобретении имеется возможность управления допуска.	Неограниченное количество вопросов; установка шкалы оценки	Не требует установки, работает на ПК, планшете, телефоне	—	Доступны в кабинете Администратора
3	Online Test Pad	Количество пользователей не ограничено.	Неограниченное количество вопросов; редактор в любой форме	Не требует установки, работает на ПК, планшете, телефоне	Автоматическая генерация	Табличный отчет с экспортом в Excel
4	ТЗС	Количество пользователей не ограничено. Допуск	Неограниченное количество вопросов;	Устанавливается на компьютеры лаборатории САПР	Подача ТЗ в установленном порядке	Отчет по датам и группам (Таблица)

		выдается администратором				
5	WebTest	Администратор	Неограниченное количество вопросов;	Не предусмотрена, вход через сайт вуза	Автоматическая генерация	Отчет по датам и группам (ведомость), по студенту для реализации апелляции.
6	Moodle	Количество пользователей не ограничено. Выдается администратором с правом редактирования своих дисциплин	Неограниченное количество вопросов	Не предусмотрена, вход через сайт вуза	Подача ТЗ в установленном порядке	Отчет по датам и группам
7	Microsoft Teams	Количество пользователей не ограничено. Выдается администратором	Неограниченное количество вопросов, ввод комментария	Не предусмотрено, регистрация	Подача ТЗ в установленном порядке	Отчет
8	ИСОРО	Количество пользователей не ограничено. Выдается преподавателю с правом администратора	Неограниченное количество вопросов	Не предусмотрена, вход через сайт вуза (ПК, планшет, телефон)	Адаптивная и интерактивная подача ТЗ (допускает работу экзаменатора во время тестирования и вмешательства при формировании вопроса/ответа)	Отчеты по группам, по студентам, в таблице, в виде граф, оценка на каждый модуль, а также итоговая

В таблице 1.2 представлен анализ сравнений преимуществ в использовании форм ТЗ, а также идентификация/аутентификация.

Таблица 1.2 Сравнение преимуществ, имеющихся в системах тестирования

№	Наименование	Применение ТЗ закрытой формы	Применение ТЗ открытой формы	Применение ТЗ установленные последовательности	Применение ТЗ соответствие	С применением устного/письменного ответа	Идентификация* Аутентификация** Скрытая Аутентификация***
1	INDIGO	✓	✓	✓	✓	-	✓
2	Let's test	✓	-	-	-	-	-
3	Online Test Pad	✓	✓	✓	✓	-	-
4	ТЗС	✓	-	-	-	-	✓
5	WebTest	✓	✓	-	-	-	✓
6	Moodle	✓	✓	✓	✓	-	✓
7	Microsoft Teams	✓	-	-	-	-	-
8	ИСОРО	✓	✓	✓	✓	✓	✓✓✓

Как видно из таблицы, каждая из систем имеет свои преимущества, критерии которых не отображены в данной таблице, поскольку акцент поставлен на формы тестовых заданий, а также применение идентификации и аутентификации во время прохождения тестового контроля.

Представленная модель системы ИСОРО и разработанный модуль дает возможность использования всех форм тестирования, а также прохождения идентификации, аутентификации и скрытой аутентификации при прохождении контроля.

Для усовершенствования данной модели есть предлагаемые ниже варианты с использованием следующих подходов: а) отслеживания в реальном времени процесса прохождения тестов (собеседования он-лайн) через контроль взаимодействия пользователя с клавиатурой/мышью; б) отслеживания в реальном времени процесса прохождения тестов (собеседования он-лайн) через контроль голоса/звуков пользователя (диктора/дикторов) с ведением протокола устных сообщений дикторов; в) отслеживания в реальном времени процесса прохождения тестов (собеседования он-лайн) через контроль отображаемых на экране/экранах окон по побочным звуковым эффектам эти возможности прошли апробацию в СКУ им. М. Козыбаева и отражены, в том числе, и в работах магистрантов. Остановимся на каждой по отдельности.

1. «Разработка нейросетевого модуля, принимающий данные от пользователя, являющегося классификатором системы идентификации по индивидуальным побочным данным» [12]. С помощью языка

программирования Python реализован модуль обработки данных, поступающих с клавиатуры (время удержания клавиши и время между нажатиями клавиш). Для записи полученных данных с клавиатуры создана база данных, в которой записаны: субъект, идентификатор субъекта, код клавиши, время нажатия клавиши и время отпускания клавиши. На основании данных полученных с клавиатуры идет статистическая обработка данных и анализ активности в динамике (во времени).

2. «Разработка технологии нейронных сетей, позволяющей идентифицировать личность студента по его речи, распознавать его устный ответ и переводить его в текст, используя при этом активную Web-страницу» [13]. Разработанная информационная система, отвечает всем требованиям и функциональным возможностям в рамках авторизации и сдачи устного экзамена с применением технологии нейронных сетей.

3. «Реализация алгоритмов компрометирования персональных данных через побочные акустические каналы периферийных устройств с помощью нейронных сетей» [14, стр. 86-91].

В работе автора Денисюк А.В. [14] техника «компрометации» позволяет системе ИСОРО вести протокол прокторинга в реальном времени за счет звукового канала (наличия микрофона у ПК или лучше параллельное применение сотового телефона для создания побочного звукового канала в ИСОРО) и специфической «раскраски применяемых в интерфейсе деталей отображения информации», то есть полностью контролируется работа с окнами- на мониторе ПК, где проводится работа в ИСОРО.

На сегодняшний день известно, что технология прокторинга развивается для реализации более качественной оценки результатов обучения. Также ранее перечисленные работы магистрантов могут помочь в развитии автопрокторинга, который позволит автоматически распознавать/идентифицировать личность, подавать сигнал при появлении посторонних звуков и предметов в помещении, а также выявления открытия дополнительного окна обучающимся при прохождении тестирования.

1.3 Методы (протоколы) и модели диагностических средств оценки результатов

«Образование, являющееся одной из важнейших сфер человеческой деятельности, по-прежнему активно влияет на все стороны общественной жизни. Для обеспечения конкурентоспособности страны на мировой арене необходимо не только правильно определить направление развития новой экономической структуры, но и найти адекватные механизмы подготовки кадров. Современный уровень развития государства, его экономический потенциал и степень сформированности человеческого капитала напрямую зависят от готовности системы профессионального образования эффективно готовить высококвалифицированных специалистов. В настоящее время

строительство нематериального капитала, а именно человеческого интеллектуального капитала, становится приоритетным направлением развития страны» [15, стр. 953], [16, стр. 89].

1. С развитием технологической базы компьютерное тестирование становится средством оценки уровня достижений испытуемого [**Ошибка! Источник ссылки не найден.**]. В этих условиях соответствующие тесты для оценки уровня сформированности компетенций становятся неотъемлемой частью процесса.

2. В рамках информационной системы управления процессами формирования новых навыков и умений осуществляется методическое, математическое, лингвистическое, программное, информационное и инженерно-техническое обеспечение всех аспектов процессов. Информационную систему управления необходимо рассматривать как функциональный комплекс, имеющий иерархическую структуру.

Процесс создания корректного комплекта диагностических средств оценки уровня результатов подготовленности специалистов проходит поэтапно.

- Выбор материала, подлежащего испытанию формируется «круг тем, которые должны быть включены в пакет, а также необходимое количество заданий и их тип, к которому должен быть представлен каждый аспект. Состав представленного материала разбивается на несколько частей, и процент каждого блока в среднем формируется таким образом, чтобы процент вопросов, генерируемых для каждого блока, согласовывался с большим весом модуля.

- Создание задания многообразных типов в тестовой форме, входящий в комплект для всего курса или для тестовой части, и объединяют их в тематические порядки.

- Отбор и корректировка тестовых заданий из набора, уточнение и определение классификаций к нему.

- Разработка набора тестовых заданий из контролируемых заданий, который должен быть из заданий в тестовой форме возрастающей сложности с учетом необходимого показателя дескрипторов» [18].

3. «Распознавание полноты получения информации с необходимостью требует существования диагностических заданий открытых и закрытых типов»[18].

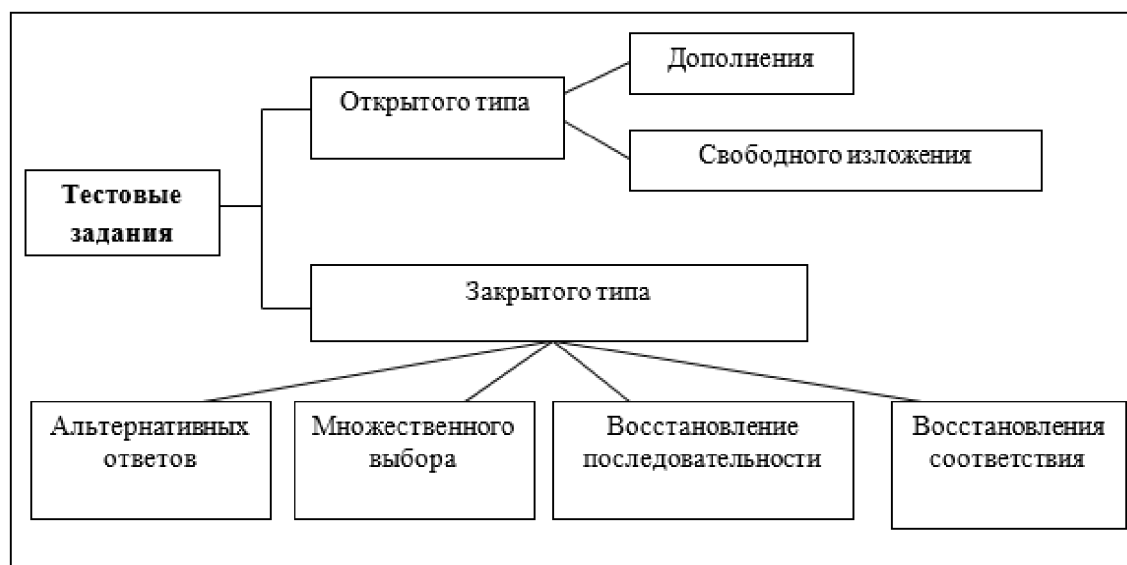


Рисунок 1.1 – Типы ТЗ

Для заданий открытого типа готовые ответы с выбором не даются. Требуется заключение, сформулированное испытуемым (интересна возможность протоколирования устного ответа, при оснащённости компьютера мультимедиа).

Задания, как правило, они имеют форму неполного утверждения, не содержащего одного или несколько основных компонентов. Основными компонентами могут быть: цифра, слово, буква или фраза, а также объекты – формулы и т. д. Формулировка задания обычно предполагает, что необходимо подчеркнуть ключевой элемент, например, поставить тире или многоточие.

Утверждение становится истинным утверждением, если ответ правильный, и ложным утверждением, если ответ неправильный.

Необходимость иметь все возможные варианты правильного ответа и отражать их в ключе создает определенные трудности в автоматизации процесса.

Тестовые задания закрытого типа:

а) «Если к заданиям даются готовые ответы на выбор (обычно один правильный и остальные неправильные), тогда эти задания называются заданиями с выбором одного правильного ответа»[6] или с одним выбором. При использовании этого типа мы придерживаемся правила: в любом задании с выбором правильного ответа этот правильный ответ должен быть единственным.

б) В качестве альтернативы используются задачи с множественным выбором. Эта форма исключает из общего списка ответов следующие варианты: «все ответы правильные» или «правильного ответа нет».

Существует не менее 4 и не более 7 вариантов выделения (дистракторов), что достаточно разумно, но не обязательно. Чем меньше дистракторов, тем больше вероятность угадывания правильного ответа. Слишком большое количество дистракторов делает задание громоздким.

в) Задания на определение правильной последовательности состоят из однородных элементов определенной группы и четкой формулировки критерия расположения этих элементов. Задание начинается со слов: «последовательность...», «определить порядок...» и т.д.

г) «задача установления соответствия состоит из двух групп элементов и четкой формулировки критерия выбора соответствия между ними»[7].

«Сходство устанавливается по принципу 1: М (одному из элементов первой группы соответствует М (возможно, один) элемент второй группы). Внутри группы имеется ввиду, что элементы схожие. Количество элементов во второй группе не могут повторять количество элементов в указанные в первой группой. Количество элементов могут быть ограничено совершенно произвольно, например, вторая группа – не более 10, первая группа – не менее 2. задание начинается со слова: «Соответствие...», «пропорциональность...» и т.д.[6, стр. 23].

Цифры и буквы используются в качестве идентификаторов (меток) элементов [6, стр. 23].

В Северо-Казахстанском университете имени Манаша Козыбаева на протяжении десятилетий, в области оценки знаний методом тестирования широко используется система тестирования WebTest собственной разработки на основе интернет-технологий.

Система полноценно функционирует для всех форм и видов аттестаций компьютерного контроля компетенций.

«Для проведения тестового контроля в оболочке WebTest предусмотрено несколько режимов тестирования» [11]:

1. «Режим обычного тестирования – это стандартный режим, который предусматривает при ответе на вопрос выбор одного из пяти вариантов ответов заданий закрытого типа (при четком тестировании) или нескольких (при, так называемом, нечетком тестировании)» [11];

2. «Режим тренинга – предусмотрен в системе WebTest для повышения качества подготовки студентов к внешней оценке учебных достижений, аттестации, абитуриентов к ЕНТ и т.д. Смысл его заключается в том, что при верном выборе варианта ответа в диалоговом окне загорается зеленый индикатор возле кнопки ответа. При неверном выборе варианта ответа одновременно загораются красный индикатор неверного ответа и зеленый индикатор верного» [11].

Для диагностики затруднений при организации компьютерного тестирования было проведено анкетирование среди профессорско-преподавательского состава Северо-Казахстанского университета имени Манаша Козыбаева.

Для анализа были выбраны несколько важных аспектов организации компьютерного тестирования:

- разработка тестовых материалов,
- разработка рисунков и формул для оболочки WebTest,
- порядок передачи тестов,

- работа с сопутствующей документацией.

Отвечая на первый вопрос, преподаватели указывали типы тестирования, где большинство преподавателей указали, что предпочитают использовать закрытый либо открытый тип тестирования. Это обусловлено тем, во-первых, что в университете используется система тестирования WebTest, которая распознает тесты закрытого типа, во-вторых, удобством создания такого типа тестирования, в-третьих, при использовании открытого типа тестирования преподаватель наглядно видит, что студент выполнил задание, зная на него четкий ответ, и работа была выполнена индивидуально.

Тест на установление соответствия и на установление правильной последовательности преподаватели предпочитают использовать гораздо реже. Возможно, это связано с трудностями разработки таких тестовых заданий и большими материальными затратами.

Оценивая преимущества компьютерного тестирования, используемые в современном учебном процессе, преподаватели отметили с их точки зрения девять важнейших, несущественными признаны показатели с весом (долей выборности) менее 10%.

Было опрошено 53 преподавателя, постоянно применяющих компьютерное тестирование в своей деятельности, при этом корректно заполненных анкет оказалось 45. Посредством интерактивного калькулятора [20,21] были рассчитаны объемы выборки с разными доверительными вероятностями и предельно допустимой ошибкой 5% (рис.1.2).

Расчет размера выборки

Доверительная вероятность ("точность")	☐ 85% ☐ 90% ☒ 95% ☐ 97% ☐ 99% ☐ 99,7%
Доверительный интервал ("погрешность" $\pm$ %)	5
Генеральная совокупность ("всего респондентов")	52
Очистить Рассчитать выборку	
Требуемый размер выборки	46

Расчет размера выборки

Доверительная вероятность ("точность")	☐ 85% ☒ 90% ☐ 95% ☐ 97% ☐ 99% ☐ 99,7%
Доверительный интервал ("погрешность" $\pm$ %)	5
Генеральная совокупность ("всего респондентов")	52
Очистить Рассчитать выборку	
Требуемый размер выборки	44

Рисунок 1.2 – Расчет размера выборки преподавателей с надежностью 95% и 90%

Ошибку не более 5% обеспечивают выборки в 46 анкет (с надежностью 95%) и 44 анкеты (с доверительной вероятностью 90%). Имеющиеся 45 анкет обеспечивают с надежностью 95% только 5,41% предельной выборочной ошибки (рис.1.3). Если уменьшить уровень доверия до 90% предельная ошибка уменьшается до 4,56%.

Расчет доверительного интервала ("погрешность", $\pm$ %)

Доверительная вероятность ("точность")	☐ 85% ☐ 90% ☒ 95% ☐ 97% ☐ 99% ☐ 99,7%
Размер выборки, человек	45
Генеральная совокупность	52
Процент ответов	50
Рассчитать интервал	Очистить
Доверительный интервал ($\pm$ %)	5.41

Рисунок 1.3 – Расчет предельной ошибки, обеспечиваемой имеющимся пулом анкет с надежностью 95%

Анализ полученной информации дает возможность по результатам обработки 45 имеющихся корректно заполненных анкет делать статистически значимые выводы. Обобщающие результаты представлены в таблице 1.3.

Таблица 1.3 – Преимущества компьютерного тестирования

№	Преимущества компьютерного тестирования	Кол-во человек	%
1	Объективность	28	62
2	Валидность	33	73
3	Демократичность	37	82
4	Массовость и кратковременность	43	96
5	Технологичность	39	87
6	Проверка знаний в любое время	40	89
7	Сочетание с другими формами контроля	7	16
8	Всесторонность (полная проверка знаний)	16	36
9	Проведение промежуточного контроля по темам	44	98

Таким образом, абсолютное большинство (98%) преподавателей главным преимуществом считают проведение промежуточного тематического контроля, значит – и его соответствие важнейшим темам и модулям рабочей программы дисциплины. 96% опрошенных подчеркивают возможность за определенный установленный промежуток времени охватить итоговым контролем большое количество тестируемых. Всего 16% получило сочетание теста с другими формами контроля.

Результаты анализа наиболее существенных проблем в использовании компьютерного тестирования в учебном процессе по результатам анкетирования преподавателей представлены на рисунке 1.4.

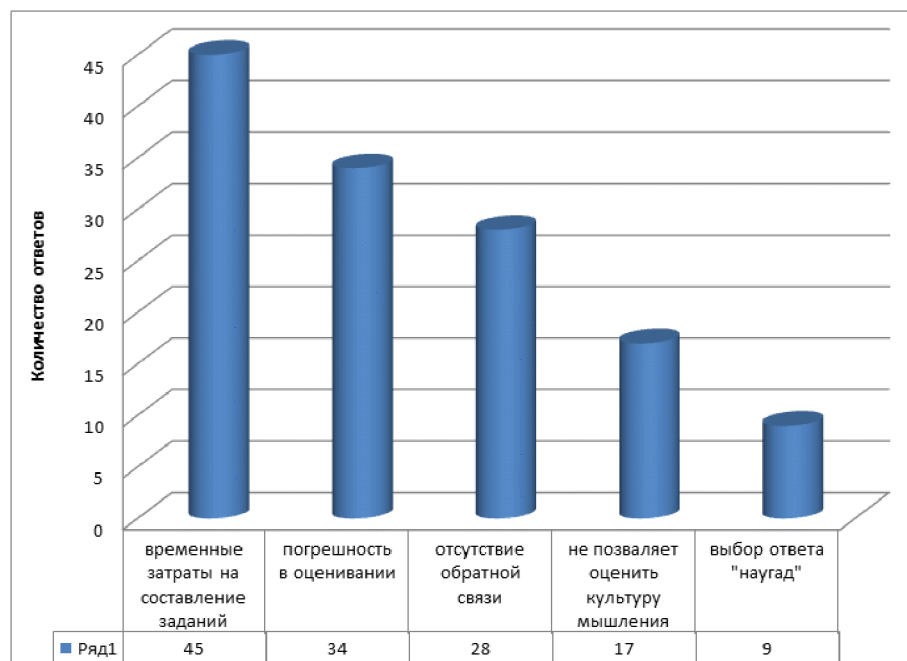


Рисунок 1.4 – Основные недостатки КТ

Как видно из диаграммы, все преподаватели выделили временные затраты на составление заданий как основную проблему тестирования. Это связано с отсутствием ресурса на сопровождение банка разработанных тестов по всему спектру учебных дисциплин, особенно для специальностей с набором в малокомплектные группы. Высок процент преподавателей (76%), обозначивших проблемой погрешности в оценивании результатов тестирования: тестирование позволяет оценивать знания студентов, а не их умения или навык решать практические задания (то есть не покрывается весь спектр HardSkills, а тем более не покрыт и так называемый спектр SoftSkills); невозможность оценивать сам процесс формирования ответа; категоричность оценки выполнения тестового задания. Более половины респондентов (62%) отметили отсутствие обратной связи, 38% – то, что тест не позволяет оценить культуры мышления студентов, 20% преподавателей считают главной проблемой возможность выбора ответа «наугад» [22].

Для более глубокого понимания преимуществ и недостатков компьютерного тестирования, проводилось анкетирование среди студентов 2 и 3 курсов. Достаточно стабильно около 2000 студентов в качестве контроля регулярно проходят компьютерное тестирование. Посредством интерактивного калькулятора (<https://socioline.ru/rv.php>) [20] был рассчитан объем выборки с доверительной вероятностью 95% и предельно допустимой ошибкой 5% (рис/ 1.5).

Расчет размера выборки

Доверительная вероятность ("точность") ☐ 85% ☐ 90% ☒ 95% ☐ 97% ☐ 99% ☐ 99,7%

Доверительный интервал ("погрешность" $\pm$ %)

Генеральная совокупность ("всего респондентов")

Требуемый размер выборки

Расчет доверительного интервала ("погрешность", $\pm$ %)

Доверительная вероятность ("точность")	☐ 85% ☐ 90% ☒ 95% ☐ 97% ☐ 99% ☐ 99,7%
Размер выборки, человек	322
Генеральная совокупность	2000
Процент ответов	50
Расчитать интервал Очистить	
Доверительный интервал ($\pm$ %)	5

Рисунок 1.5 – Расчет размера выборки учащихся при 5%-й значимости

Мнения студентов, на основании обработки 322 анкет, о преимуществах и недостатках компьютерного тестирования представлены в таблице 1.4.

Таблица 1.4 – Мнения студентов о преимуществах и недостатках тестирования

Критерии преимуществ и недостатков КТ		Компьютерное тестирование (322 студента)	
		Кол-во человек	%
ПРЕИМУЩЕСТВА	Объективность	322	100
	Демократичность	306	95
	Мобильность и кратковременность	322	100
	Технологичность	322	100
	Индивидуальное выполнение	322	100
	Психологический комфорт	245	76
	Умственный труд	167	52
	Выбор ответа наугад	74	23
НЕДОСТАТКИ	Ожидание результата	-	0
	Сбой в сети	306	95
	Недостаточная оснащенность аудиторий техникой	322	100

Анализ результатов тестирования показывает, что респондентов привлекает мобильность, технологичность, индивидуальное выполнение (100%). Отмечены спорные преимущества – «выбор ответа наугад» (23%) [23].

Был проведен анализ результатов опроса преподавателей и студентов, соотнесены их мнения по преимуществам и недостаткам компьютерного тестирования. Итоги анализа представлены в виде диаграммы на рисунках 1.6, 1.7.

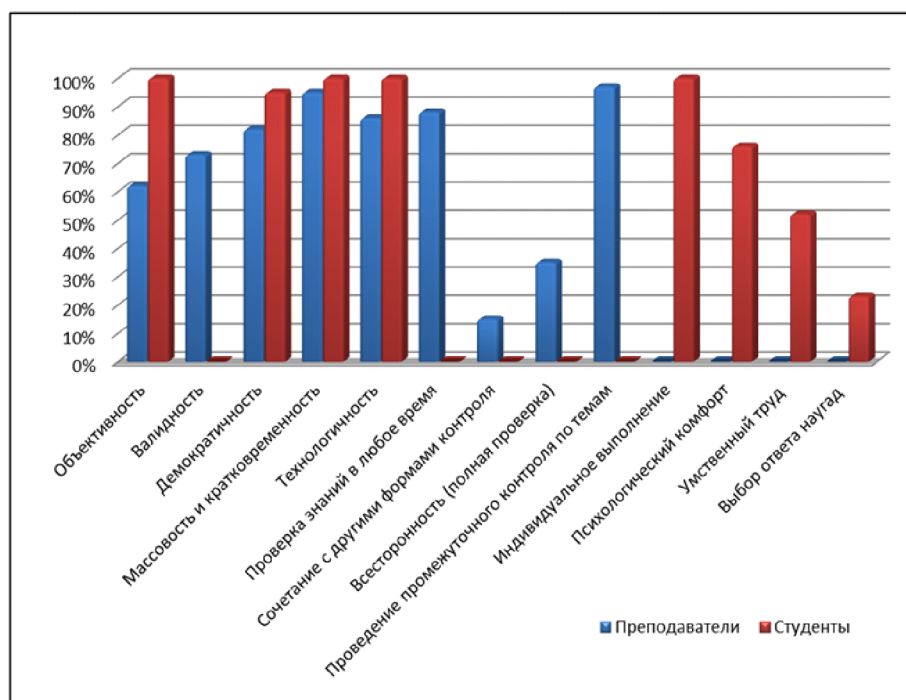


Рисунок 1.6 – Преимущества КТ в мнениях студентов и преподавателей

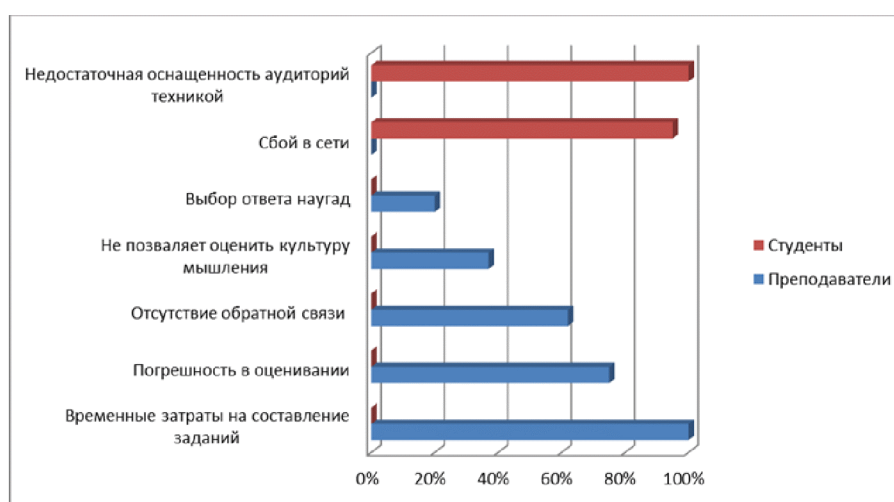


Рисунок 1.7 – Недостатки КТ в мнениях студентов и преподавателей

Все возрастающая трудоемкость деятельности определила необходимость изучения характера затруднений преподавателей при составлении тестовых заданий для контроля.

Согласно анкетным данным анализ полученных материалов позволил произвести обработку и выделить две основные группы затруднений (обобщающая таблица 1.5).

Таблица 1.5 – Затруднения преподавателей при составлении тестовых заданий

Группы Виды	Содержательная			Техническая		
	Содержание ТЗ	Разработка Вопроса к ТЗ	Разработка вариантов ответов	Технология ввода тестов в оболочку	Оформление рисунков	Оформление формул
Кол-во человек	37	4	29	41	23	15
%	82	9	64	91	51	33

Основные затруднения преподавателей связаны с технологией ввода тестов в оболочку (91%), разработкой содержания тестовых заданий (82%), разработкой вариантов ответов (64%) и оформлением рисунков (51%).

«Важной задачей было выявление этапов компьютерного тестирования и содержания каждого этапа: определение ключевой идеи, целеполагание, формирование структуры теста, определение формы тестовых заданий. Соблюдение последовательности этапов и всей процедуры компьютерного тестирования обеспечивает адекватность оценивания и определение уровня освоения компетенций (за некоторым исключением)» [24].

«Внедрение новых форм организации, распространение современных методов обучения, является неотъемлемым условием. Также существенной корректировке должны подвергнуться и подходы к оценке качества учебных достижений. Образовательной системе необходимо обновлять компоненты с учетом развития социальной сферы» [25].

«Основное, что в тестировании нехорошо – это постулирование того, чего реально не произошло (как минимум, по критерию устойчивости и наблюдаемости). Например, авторство теста и ответственность автора/авторов, предполагающая знакомство с тестологией и, следовательно, владение теорией статистического объяснения» [26].

Основные проблемы «закрытого» теста, «линейного» по выполнению: дублирование заданий; возможность угадывания; «потребность наращивания объемов и разнообразия тестового материала при имеющихся на сегодняшний день ограничительных условиях; многочисленные методические «ловушки»; потеря управления проверки (хотя бы отчасти) над «сложностью», «трудностью».

«Не решает проблему угадывания и переход к «открытой» форме тестового задания» [26, стр.195].

1.4 Постановка задачи диссертационного исследования

Исходя из вышеизложенного материала, определена научная и практическая цель данного исследования, заключающиеся в повышении контроля обучения с помощью системы управления, адаптивные компоненты, которые используют разработанные схемы и алгоритмы управления в системе оценки результатов обучения (рис.1.8), что способствует оперативному реагированию на возникающие проблемы.

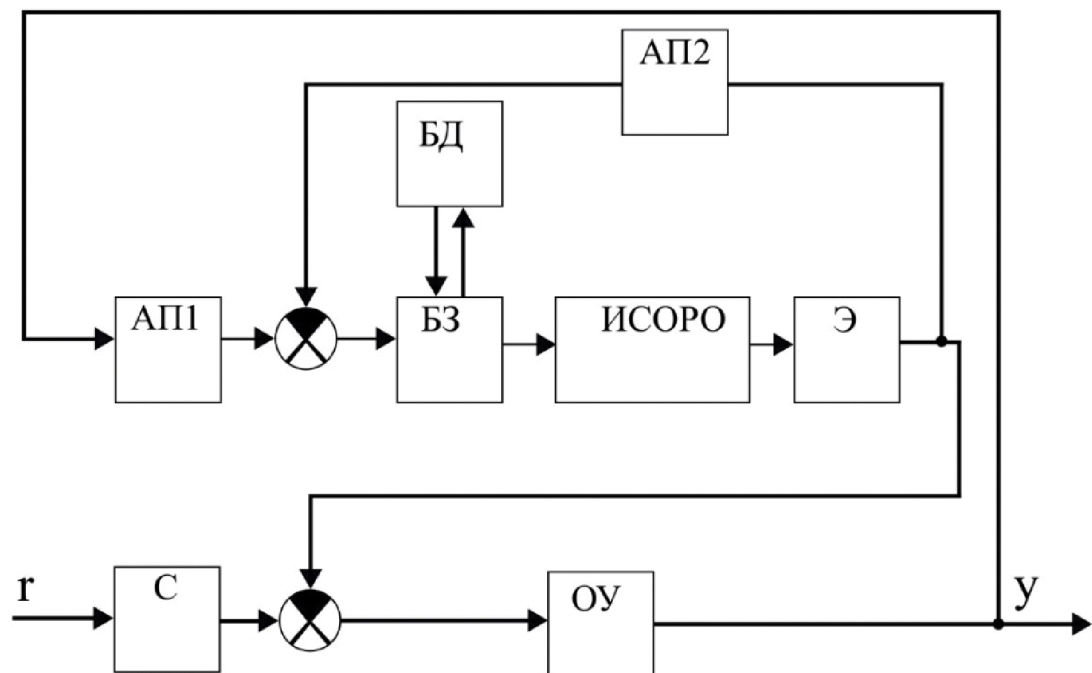


Рисунок 1.8 – Схема интеллектуальной системы управления обучением

Согласно предложенной схеме рассматривается контур управления, основанный на принципе отрицательной обратной связи, в котором управляющее воздействие корректируется в зависимости от выходного сигнала. Объектом управления (ОУ) является процесс обучения в вузе.

Функциональная схема состоит из ОУ – процесс обучения в вузе и управляющего устройства, который в свою очередь включает автопрокторинг двух видов, базы данных и знаний, модуля интеллектуальной системы оценки результатов обучения, а также принятие решения методом энтропийного подхода.

Автопрокторинг 1 (АП1) – идентификация на входе методом сравнения пароля, введенного пользователем с логином/паролем, сохраненный в базе данных, а также обрабатывает то что использует АП2. Автопрокторинг 2 (АП2) – аутентификация во время тестирования, в петле обратной связи, а именно распознавание звуков/голоса, клавиатурного ввода и через побочные акустические каналы периферийных устройств. Вся эта информация поступает в базу данных (БД). При переходе к блоку интеллектуальной системы оценки

результатов обучения (ИСОРО), блок (Э) использует энтропийную оценку состояния для базы знаний (БЗ) и БД принимает решение.

Информация с блока (Э) возвращается через сумматор на вход (ОУ) для обеспечения управления процессом.

Входные и выходные сигналы:

г – задающий входной сигнал;

с – регулятор характеристики (нормативные требования оценивания);

у – выходной сигнал(выход).

Таким образом, цель данного исследования разработка интеллектуальной системы, способствующей оценке результатов обучения путем применения:

- адаптивных технологий;
- разных версий заданий (закрытые, открытые, мультимедиа и прочие);
- явного (или неявного) наличия эксперта-экзаменатора в процессе проведения экзамена.

1.5 Выводы по первой главе

В практике современного образования нет универсального оценочного метода, который бы полностью решал проблему объективности в оценке профессиональной компетентности студентов. Одни методы оценки позволяют оценить результаты развития одних профессиональных компетенций, другие – наоборот [27].

Подводя итог, в результате можно вынести несколько определенных заключений:

а) Процессный контроль как необходимая, а также непрерывная часть образовательного процесса, должна быть систематически устроенным и производиться во всех его функциях не только для окончательного/итогового контроля. Методы, формы и средства контроля должны быть изменчивыми и гибкими. Только в таком случае контроль обеспечивает индивидуализацию процесса, присущую самим условиям учебного процесса: по отдельности каждой учебной группы, актуальному материалу, уровню готовности;

б) Контроль, а также оценка знаний и умений студентов должен проводиться систематически на всех уровнях обучения, и только при хорошо налаженном контроле и своевременной оценке результатов, можно говорить об эффективном обучении;

в) Помимо традиционных форм контроля знаний, необходимо внедрить и освоить контрольно-обучающие программные средства, так как использование компьютера для проведения оперативного контроля и записи результатов самостоятельной работы в онлайн режиме.

2 МОДЕЛЬ УПРАВЛЕНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМОЙ ОЦЕНКИ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

2.1 Анализ и сравнение результатов тестирования

Сегодня использование тестовых технологий как инструмента оперативного и итогового контроля качества обучения студентов повсеместно используется в сфере образования. Тестовый вид контроля получил широкое распространение, потому что он имеет ряд преимуществ перед другими методами контроля знаний и является обязательной составляющей многих педагогических инноваций.

Компьютеризация образования влечет изменение форм организации учебного процесса, среди которых особо можно выделить современные компьютерные системы адаптивного обучения и контроля. Неотъемлемой частью таких систем является тестовый контроль.

Для этого необходимо уметь оценивать качество всего теста и качество входящих в его состав заданий [28], соотнося итоги успеваемости обучающихся посредством тестирования и прочие их достижения (например, для потенциального использования мнения студентов в экспертизе).

Некорректное использование статистических методов, недостаточный уровень статистической грамотности, инерционность сообщества системы образования привело к закреплению традиции статистического оценивания, заблуждению о сущности информации, которую она предоставляет.

К проблемам статистического вывода относятся:

- рутинное и ошибочное применение проверки значимости нуль-гипотезы (ПЗНГ) в исследованиях;
- неадекватность ПЗНГ по отношению к целям и задачам исследований;
- неправильная интерпретация результатов ПЗНГ.

Ныне, как правило, исследования в социотехнических системах планируются [29]. И желательно планировать качественно. Организация и дизайн научных исследований – за рамками данного исследования. Основные возникающие вопросы:

- описательная статистика (как грамотно описать выборку);
- аналитическая статистика (как грамотно выявить свойства и перенести их на генеральную совокупность).

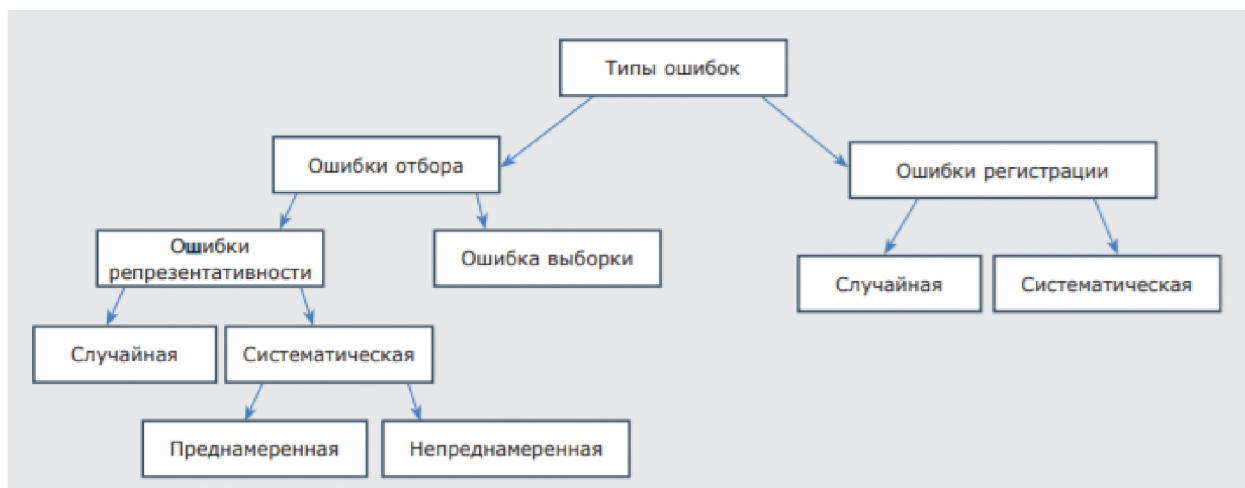


Рисунок 2.1 – Классификация ошибок

Проблемы статистическим выводам и бинарным решениям на основе некоего критерия, который с некоторой вероятностью может выдавать неверный результат, создают мифы и традиции статистического оценивания.

«Одна из основных проблем тестирования: «Пределы роста объемов тестовой базы становятся неизбежной темой для обсуждения: - большая база в меньшей степени «боится разглашения», зато аргументированность введения очередного тестового задания становится явной проблемой; - метод «исчерпания» заданий в процессе составления «книжки тестов» обеспечивает неповторимость в текущих времени и объеме, но соблюдение пропорции сложность-трудность-разнообразие приводит к дальнейшему разрастанию базы и т.д. Очевидно, валидизация такой базы – это многочисленные аттестационные «прогоны», иначе статистика применения низка, и выводы по ней – неточны. Разрешающая способность заданий должна быть соотнесена со шкалой оценок, но кратность «прогонов» только возрастает. Из этого следует неутешительный результат – Ваши экзаменационные материалы, скорее всего, обладают непроверяемой разрешающей способностью и недоказуемой надежностью с неизбежной рекомендацией: на живых людях – не применять!» [26, стр.196].

Результаты тестирования представляют собой некоторую генеральную совокупность статистических данных. «Количественную оценку надежности выборочного показателя используют, для того чтобы составить представление о ведущей характеристике. Это реализовывают либо на основе выборочного показателя с учетом его случайной ошибки, либо на основе выдвижения некоторой гипотезы (о значении средней дисперсии, характере распределения, связи) в отношении свойств» [31] основной совокупности.

Значение «случайной ошибки репрезентативности зависит:

- 1) от объема выборки;
- 2) степени вариации изучаемого признака в основной совокупности;
- 3) принятого метода становления выборочной совокупности» [31].

В контексте проводимой работы формирование выборки не рассматривается, принято допущение об отсутствии систематических ошибок, связанных с нарушением правил отбора и/или реализацией выборки.

Сведение к минимуму ошибок выборок является одной из основных целей статистического анализа данных, полученных при исследовании образовательных процессов. При этом никакая статистическая обработка данных не в состоянии устранить неизвестную ошибку отбора или регистрации, которая может возникать на любом этапе исследования.

В «тестовой» жизни университета можно выделить два направления:

- массовое тестирование (республиканского уровня), тогда обработку данных, в том числе – для сравнительного анализа, можно проводить по методикам, не требующим предварительных знаний об исследуемом объекте (системе);

- регулярное тестирование, предполагающее разнообразие количества тестируемых, частоту применения теста и т.п., т.е. методики, требующие предварительных знаний об изучаемом явлении.

Для первого направления, согласно таблице К.А. Отдельновой (таблица 2.1), «три уровня точности результатов исследования с практической точки зрения условно делятся таким образом:

- «ориентировочное знакомство» соответствует пилотному исследованию;
- «исследование средней точности» обеспечивает результаты для опубликования в качестве научной статьи;
- результаты «исследования повышенной точности» позволяют сформировать окончательные заключения» [29, стр.97].

Таблица 2.1 – Определение объема выборки по методике К.А.Отдельновой

Уровень значимости (p)	Уровень точности		
	Ориентировочное знакомство	Исследование средней точности	Исследование повышенной точности
0,05	44	100	400
0,01	100	225	900

В исследованиях образовательных систем и процессов в качестве границы статистической значимости результатов вполне корректно использовать уровень значимости равный 0,05. Согласно методике К.А.Отдельновой (широко применяется в исследовании социологических систем) получение отдельных значимых результатов тестирования в отдельно взятом ВУЗе обеспечить можно.

Таблица, разработанная N. Fox (таблица 2.2) [30], предусматривает для определения необходимого объема выборки оценку величины допускаемой ошибки.

Таблица 2.2 – Определение объема выборки по методике N. Fox

Величина допускаемой ошибки, %	Объем выборки
10	88
5	350
3	971
2	2188
1	8750

Объединение двух методик дает информационный материал для разработки организационно-управленческих решений по проблемам учебного (образовательного) процесса.

Учитывая важность для социальных/социотехнических систем ориентироваться на мощность статистических выводов, не ограничиваясь их значимостью (надежностью), при планировании исследования/эксперимента для решения конкретных задач важно не только «подстраивать» статистические процедуры к схеме исследования, «но и установить оптимальное соотношение между мерой надежности полученных результатов и величиной предельной ошибки выборки. Таким образом, организация выборочного наблюдения предполагает определение объема выборки, необходимого для получения требуемой точности результатов с заданной вероятностью» [31] и приемлемой мощностью.

«Необходимый объем выборки представляет собой функцию вариации измеряемых параметров генеральной совокупности и точности оценки этих параметров, требуемой исследователем»[34]. В региональном ВУЗе чаще всего приходится довольствоваться частными результатами тестирования (конкретных сессий, специальностей, предметов и т.п.). В зависимости от конкретных условий применяется несколько расчетных схем определения необходимого объема выборки для значимых результатов на основе формул предельной ошибки выборки в соответствии с видом и способом отбора (таблица 2.3) [31].

Таблица 2.3 – Определение необходимого объема выборки

Вид изучаемых признаков	
Количественный	Качественный (пропорции от генеральной совокупности)
1. При неизвестном объеме генеральной совокупности (повторный отбор)	3. При неизвестном объеме генеральной совокупности (повторный отбор)
$n = \frac{t^2 \cdot \sigma^2}{\Delta^2}$	$n = \frac{t^2 \cdot P \cdot Q}{\Delta^2}$

2. При известном объеме генеральной совокупности (бесповторный отбор)	4. При известном объеме генеральной совокупности (бесповторный отбор)
$n = \frac{t^2 \cdot \sigma^2 \cdot N}{\Delta^2 \cdot N + t^2 \cdot \sigma^2}$	$n = \frac{t^2 \cdot P \cdot Q \cdot N}{\Delta^2 \cdot N + t^2 \cdot P \cdot Q}$

здесь:

1. t^2 – коэффициент, зависящий от выбранного исследователем доверительного уровня, «критическое значение критерия Стьюдента при соответствующем уровне значимости»[33] (табл. 2.4);

2. σ – стандартное отклонение признака, который будет изучаться в исследовании (часто – выборочная дисперсия);

3. Δ – предельно допустимая ошибка выборки (в предлагаемых исследованиях это, как правило, 5%); характеризует точность выборки с определенной вероятностью, величина которой зависит от значения коэффициента доверия t .

4. N – объем генеральной совокупности.

5. P – доля случаев, в которых встречается изучаемый признак (доля респондентов с наличием исследуемого признака,);

6. Q – доля случаев, в которых не встречается изучаемый признак ($100 - P$) (доля респондентов, у которых исследуемый признак отсутствует).

В отсутствие априорной информации значения P и Q по умолчанию принимаются за 0,5, поскольку точно неизвестны до проведения исследования (именно при таких значениях размер ошибки выборки максимален); далее ошибку можно уменьшить – при наличии дополнительных результатов исследования, т.е. поступления апостериорной информации.

Как видим, чем больше дисперсия оцениваемых параметров генеральной совокупности, тем больший объем выборки требуется для того, чтобы обеспечить требуемую точность.

Таблица 2.4 – часто используемые критические значения t -критерия

уровень доверия	0,683	0,950	0,954	0,990	0,997
t^2	1,00	1,96	2,00	2,58	3,00

Отметим, что в целях повышения мощности статистических выводов контролируется доверительный уровень до 95%.

Из формул таблицы 2.3. можно получить формулы расчета Δ – предельно допустимой ошибки выборки. «При использовании выборочного наблюдения характеристика его результатов возможна на основе сопоставления полученных пределов ошибок выборочных показателей с величиной допустимой погрешности. При решении задачи определения вероятности

того, что ошибка выборки не превысит допустимой погрешности, расчеты проводятся также на основе формул из таблицы 2.3» [31].

Значения σ – стандартное отклонение признака – можно получить только после проведения выборочного исследования, при необходимости заменяем на приближенные, определенные на основе каких-либо пилотных выборочных наблюдений, аналитических предыдущих обследований и т.п. «При любом виде выборки ее объем n рассчитываем по формуле повторного отбора, если же расчетное n превысит 5%, то корректируем значение по формуле бесповторного отбора» [32].

Проблемы с получением информативно-значимых данных иллюстрируют рис. 2.2, рис.2.3. и рис.2.4.

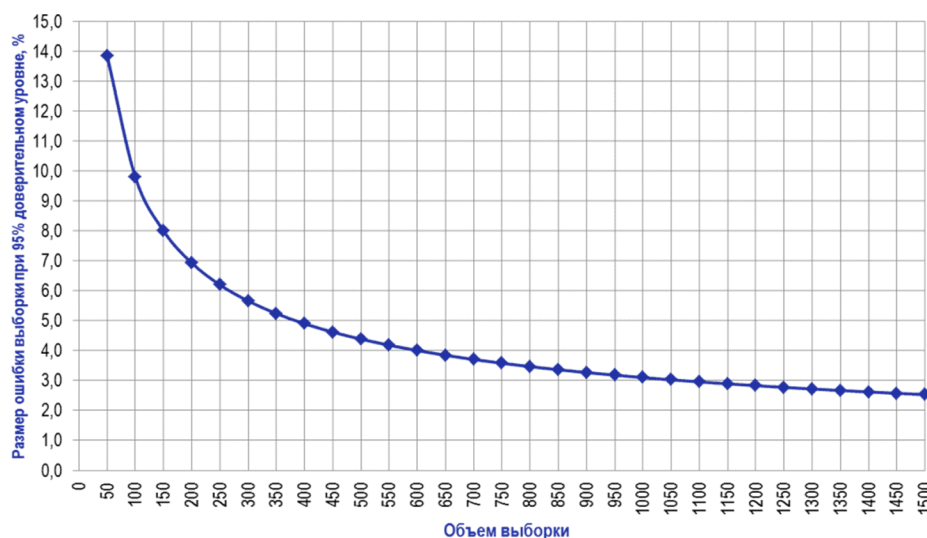


Рисунок 2.2 – Зависимость ошибки выборки от ее объема при 95% надежности

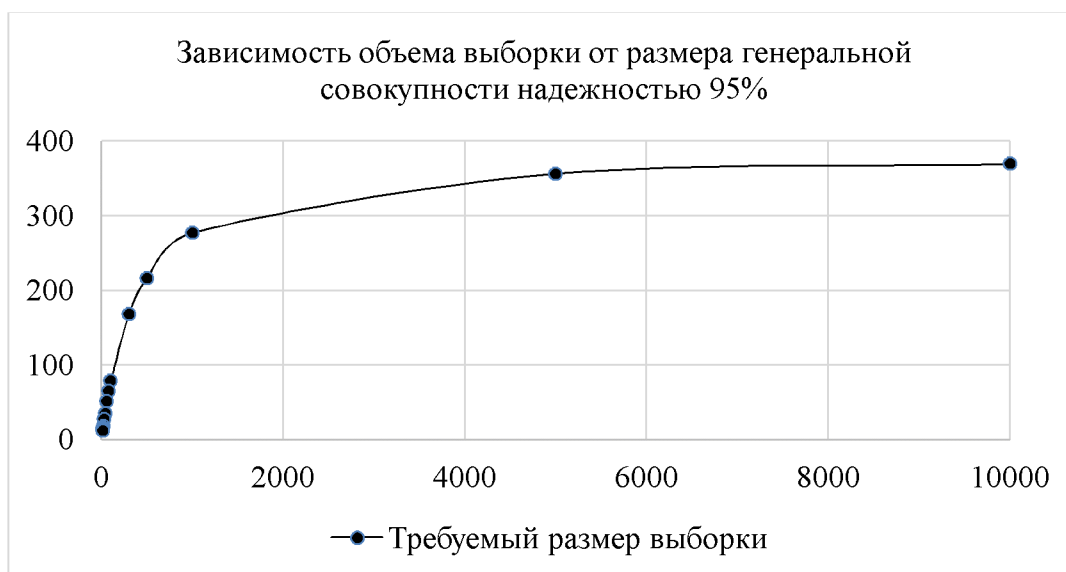


Рисунок 2.3 – Оценка размера выборки в зависимости от генеральной совокупности (до 10000 единиц) при 95% надежности.

Для большей иллюстративности фрагмент рисунка 2.3 представлен на рисунке 2.4.

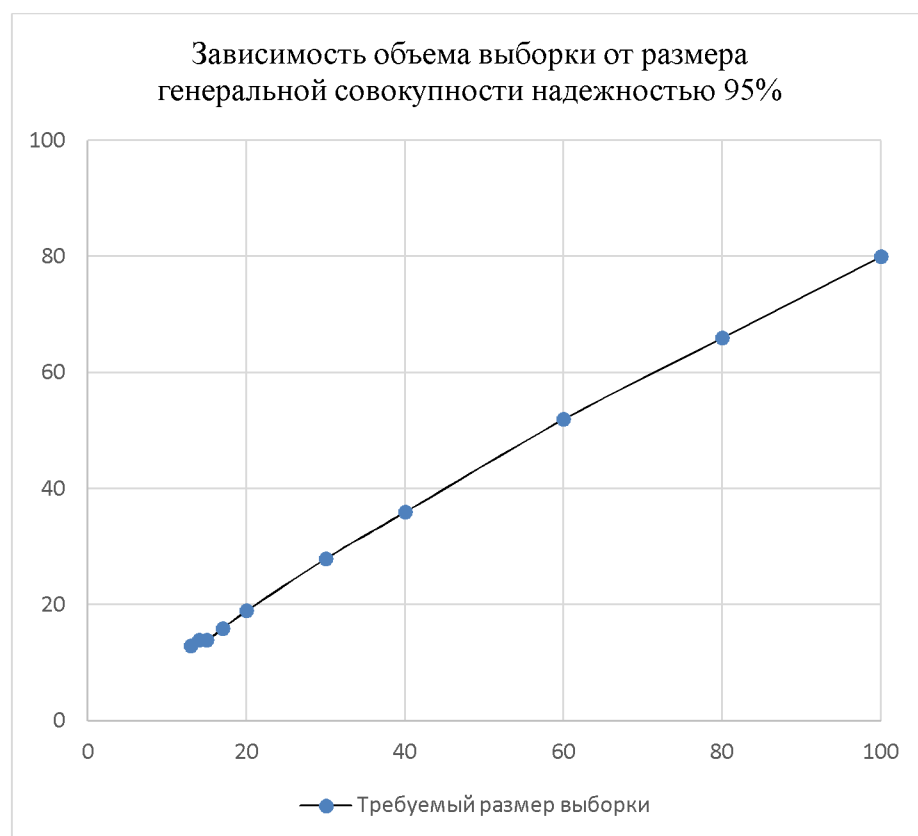


Рисунок 2.4 – Оценка размера выборки в зависимости от генеральной совокупности (до 100 единиц) при 95% надежности

«Если объем выборки меньше 10 единиц, то при ее описании следует указать весь перечень полученных значений, не прибегая к традиционным методам математической статистики, которые в данном случае являются неприменимыми»[33]. Таким образом, результаты тестирования малочисленных учебных групп (что не редкость в настоящее время) не помогут принять правильное организационное решение по поводу повышения качества подготовки специалиста.

«Эмпирическое правило о применимости центральной предельной теоремы и допущении нормальности распределения всех возможных выборочных средних работает, если объем выборки не менее 100. Если же меньше 100, то нужно иметь веские доказательства/обоснования нормальности распределения генеральной совокупности»[32], и только в этом случае можно полагать, что распределение, которому подчиняются выборочные статистики, является нормальным.

Исследование малых выборок в 10÷30 единиц требует особого подхода [33, 34], не всегда согласующегося с традиционными методиками.

Оценка параметра может быть точечной или интервальной. «Интервальный подход более оправдан при выборочном исследовании.

Точность интервальной оценки параметра, измеряемого при выборочном исследовании, определяется двумя показателями:

- интервалом, в котором ожидается обнаружить оцениваемый параметр;

- вероятностью обнаружения этого параметра в данном интервале.

Эти два показателя объединяет понятие доверительного интервала»[32].

Важность именно интервального оценивания демонстрируется следующим примером.

Экспертная группа в количестве четырех преподавателей оценивает трудность учебного модуля из 10 заданий. Имеется ли у экспертной группы единоголосное мнение? Применение дисперсионной модели дает такие результаты (таб. 2.5 рис. 2.5).

Таблица 2.5 – Результаты дисперсионной модели

Преподаватели m	Усредненная мера трудности $\bar{t}(m)$	Ошибка $\Delta_{\alpha=0.05}$	$\underline{\bar{t}}_m^0$	$\overline{\bar{t}}_m^0$
1	0.50	0.18	0.32	0.68
2	0.40	0.14	0.26	0.54
3	0.62	0.21	0.41	0.83
4	0.48	0.15	0.33	0.63

Фактически наблюдаемое значение статистики $F = \frac{s_1^2}{s_2^2} = \frac{0.75}{0.08} = 9.38$.

Критическое значение F-статистики Фишера на уровне значимости $\alpha = 0.05$ при степенях свободы $k_1 = 3$ и $k_2 = 30$ равно $F_{\alpha;k_1;k_2} = 2.92$. Так как верно $F > F_{\alpha;k_1;k_2}$ нулевая гипотеза отвергается, т.е. с надежностью 0.95 можно утверждать, что мера трудности модуля зависит от мнения конкретного преподавателя. И, следовательно, экспертная группа не может прийти к единому мнению.

Но рассмотрим доверительные интервалы для оценок генеральных средних по малой выборке с надежностью 95% (уровнем значимости 5%): $\underline{\bar{t}}_m^0 = \bar{t}(m) - \Delta_{\alpha}$ и $\overline{\bar{t}}_m^0 = \bar{t}(m) + \Delta_{\alpha}$ - соответственно левая и правая границы интервалов (рис.2.5) [0,33; 0,63], [0,41; 0,83], [0,32; 0,68], [0,26; 0,54].

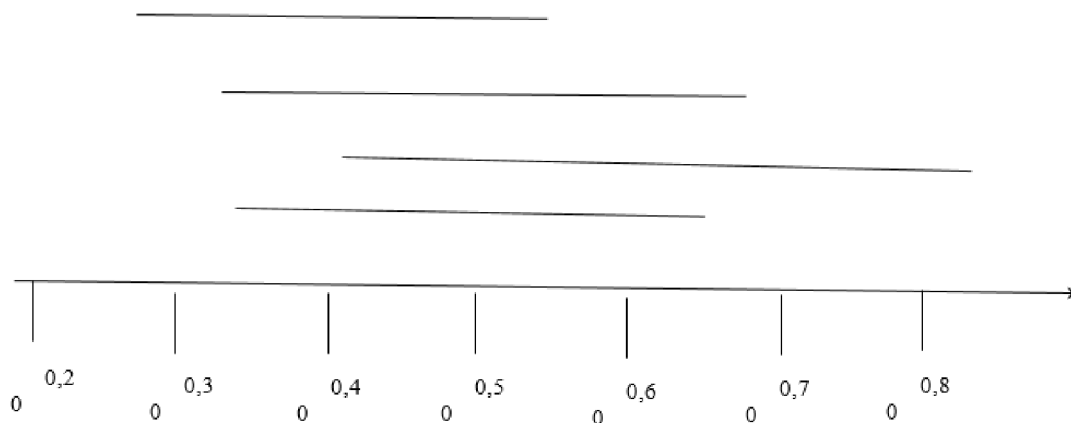


Рисунок 2.5 – Доверительные интервалы

Доверительные интервалы имеют частичные перекрытия, следовательно, несмотря на имеющуюся связь между мерой трудности и конкретным членом экспертной группы, нельзя утверждать, что экспертная комиссия не в состоянии выработать единого решения.

«Анализ полученных данных часто включает в себя и анализ подвыборок. Объемы их, естественно, меньше основной выборки, поэтому ошибка для выводов по подвыборкам больше, чем ошибка по выборке в целом»[34].

Например, при проектировании объема выборки с надежностью 95% с учетом необходимости анализа подвыборок, надо учитывать ошибки выборки/подвыборки (таблицы 2.6 (фрагмент) и 2.7). Здесь числа условные; без предварительного исследования считаем долю исследуемого признака/качества равной 50% (при поступлении информации доля исследуемого признака корректируется); генеральная совокупность – учащиеся ВУЗов РК – значительно больше выборок, соответственно, имеем:

Ошибка выборки= $1,96 \cdot \sqrt{0,5 \cdot 0,5/n}$.

Расчет проведен посредством интерактивного калькулятора (<https://scanmarket.ru/blog/vyborka-razmer-ne-glavnoe-ili-glavnoe#calc2>) [21].

Таблица 2.6 – Проектирование объема выборки с надежностью 95% с учетом анализа подвыборок (фрагмент)

Выборка/подвыборка		Объем	Ошибка %	Доверительный интервал
Университет		2000	±2,19	(47.81%;52.19%)
Факультет1 (40%)	Факультет2 (32,5%)	800	±3,46	(46.54%;53.46%)
	Факультет3 (27,5%)	650	±3,84	(46.16%;53.84%)
	Специальность31 (54,5%)	550	±4,18	(45.82%;54.18%)
	Специальность32 (45,5%)	300	±5,66	(44.34%;55.66%)
		250	±6,20	(43.80%;56.20%)

		Очное32 (64%)	160	±7,75	(42.25%;57.75%)
		Заочное32 (36%)	90	±10,33	(39.67%;60.33%)

Таким образом, для принятия решений по университету в целом (ошибка составляет 2,19%) и по факультетам (ошибка не превышает принятых 5%) выборки хватает. Но при делении, например, выборки Факультет3 на подгруппы (подвыборки) по специальностям, выводы по каждой подгруппе уже недостаточно точны. Для количественных характеристик групп по формам обучения ошибка полученных показателей будет достаточно велика.

Таблица 2.7 – Разница в ошибке выборки по подвыборкам при разных объемах выборки с 95%-й надежностью

	доля %	Выборка 2000, ошибка 2,19%		Выборка 4500, ошибка 1,46%		Выборка 2000 + дополнит.	
		Объем подвыборк и	Ошибк а %	Объем подвыбор ки	Ошибк а %	Объем подвыбо рки	Ошибк а %
Специальность11	20	400	4,90	900	3,27	400	4,90
Специальность12	11,5	230	6,46	518	4,31	230+154	5,00
Специальность13	8,5	170	7,52	383	5,01	170+214	5,00
.....	...	...	...	...	...	...	...
Специальность31	15	300	5,66	675	3,77	300+84	5,00
Специальность32	12,5	250	6,20	563	4,13	250+134	5,00

Итак, при выборке в 2000 объема данных не хватает для значимых выводов по специальностям. Можно предложить компромиссные решения:

- увеличение выборки, например, до 4500 респондентов;
- добор респондентов по некоторым специальностям.

Возможны ли подобные компромиссы в условиях ВУЗа?

Если для исследования некоторого признака принята допустимая ошибка выборки 5%, то для выводов по специальностям при университетской выборке (с соответствующими пропорциями) в 4500 единиц вполне приемлемы для принятия решений.

Минимальный объем выборки, обеспечивающий допустимую ошибку выборки 5% равен 384 единицам. Следовательно, по отдельным специальностям, согласно таблице 2.7, необходим добор до 200-220 респондентов.

Таким образом, при разработке решений необходимо учитывать возможности/ресурсы ВУЗа (временные, материальные, финансовые и т.п.) чтобы получить информацию с приемлемой ошибкой.

В отсутствие достаточной статистической базы для исследования вопросов, связанных с оцениванием достижений учащихся (в частности – тестированием), формируются квазислучайные выборки, а также выборки на основе суждения эксперта[35]. Результаты и их значимость еще предстоит исследовать.

В текущую систему ИСОРО (показано в пункте 4.4, главы 4) расчеты по минимальному/оптимальному расчету необходимого объема выборки включены в виде интерактивных калькуляторов. В настоящее время корректировка значений P и Q (доли исследуемого признака) проводится посредством скрытых Марковских моделей по восстановлению значений переходных вероятностей и/или Байесовской процедуры по мере поступления новых данных. Применимость этой процедуры оправдана, т.к. на принципе Байеса основана оценка упрощенной системы испытаний надежности, контроля качества и т.п., когда при невозможности (нецелесообразности) тотальной проверки испытание сводится к замене испытания одного фактора испытанием другого.

Дискретные расчеты проводятся по обобщенной формуле для серии s экспериментов, когда только в m ($m \leq s$) из них наступило событие A (некий результат) [36]:

$$P(H_i | A^*) = \frac{P(H_i) \cdot P(A | H_i)^m \cdot (1 - P(A | H_i))^{s-m}}{\sum_{j=1}^n P(H_j) \cdot P(A | H_j)^m \cdot (1 - P(A | H_j))^{s-m}} \quad (1 \leq i \leq n) \quad (1)$$

здесь:

A^* - сложное событие, результат серии s опытов, причем исходы отдельных экспериментов представляют собой взаимно независимые события;

« $P(H_i)$ - априорные вероятности гипотез H_i (некоторых взаимоисключающих событий, содержательно описываемых условиями проведения эксперимента);

$P(A/H_i)$ - условные вероятности события A при каждой гипотезе (с одной и только с одной из гипотез событие A происходит совместно, априорное предположение – с каким именно. Иными словами, гипотезы – частные случаи события A);

$P(H_i/A)$ - апостериорная вероятность какой-либо гипотезы при условии, что событие A произошло»[76].

«Информативным будет эксперимент, результат которого существенно изменяет априорные вероятности гипотез. Перед постановкой сложного и/или дорогостоящего эксперимента всегда имеет смысл оценить его информативность на основе имеющихся данных об априорных и условных вероятностях»[76].

В дальнейшем разрабатываемая интеллектуальная система предполагает наличие нейросети глубокого обучения.

Для оценивания тестовых заданий предложены следующие методы [28, стр. 77].

Для начала все тестовые задания были структурированы в виде графа (дерева), где вершинами являются номера вопросов теста. Каждое дерево состоит из вопросов отдельного раздела курса информатики. Сложность каждого вопроса определяется его весом, который находится как длина пути

до него от корня соответствующего дерева минус единица. На рисунке 2.6, корень дерева имеет вес равный единице. Чем сложнее вопрос, тем длиннее путь и тем больше вес вопроса.

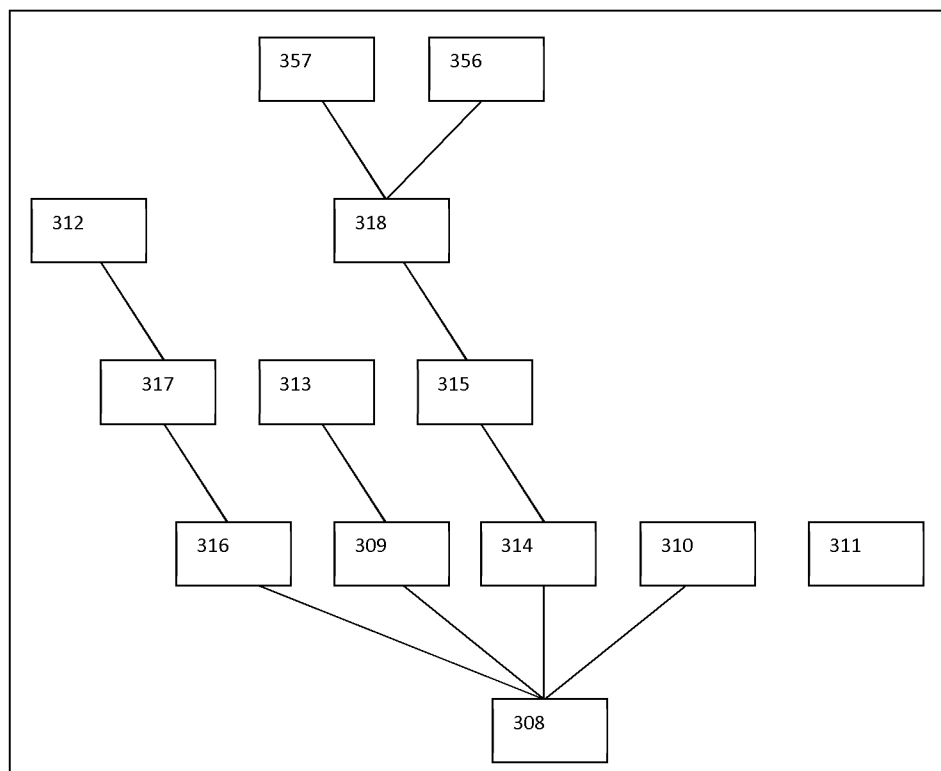


Рисунок 2.6. – Структурирование тестового задания в виде графа

Таким образом, каждому вопросу ставится в соответствие число - его вес, равное длине пути до него от корня дерева без единицы.

Первый из предложенных методов оценивания заключается в том, что за каждый правильный ответ испытуемый получает 1(единицу), за неверный ответ ноль.

Второй метод оценивания заключается в том, что за каждый правильный ответ испытуемый получает количество баллов, равное весу этого вопроса, за неправильный ответ – «0» баллов. Назовем этот метод «линейным методом».

Третий метод, который условно назовем «логарифмическим методом» заключается в следующем: за каждый правильный ответ тестирующийся получает количество баллов, равное натуральному логарифму веса вопроса, за неправильный – «0» баллов.

Согласно четвертому методу оценивания тестовых заданий каждый правильный ответ приносит испытуемому количество баллов равное квадрату веса вопроса, неправильный – «0» баллов. Его назовем «квадратичный метод».

Пятый метод оценивания, который назовем «методом квадратного корня», за каждый правильный ответ приносит тестирующемуся количество баллов, равное значению квадратного корня из веса вопроса, за неправильный – «0» баллов. Методы оценивания с количеством баллов за правильный и неправильный ответы (табл. 2.8).

Таблица 2.8 – Методы оценивания с количеством баллов

Метод оценивания	Количество баллов за правильный ответ	Количество баллов за неправильный ответ
Существующий метод	1	0
Линейный метод	k	0
Логарифмический метод	Ln(k)	0
Квадратичный метод	k ²	0
Метод квадратного корня	$\sqrt{k}$	0
Примечание: k – вес вопроса		

Проверяется допущение о подчинении результатов тестирования нормальному распределению – очень сильное предположение классической тестологии:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\alpha)^2}{2\sigma^2}} \quad (2)$$

где

α - математическое ожидание,

σ - среднее квадратическое отклонение нормального распределения.

Согласно критерию согласия χ^2 -квадрат Пирсона (менее мощный тест из имеющихся, зато более толерантен к различным шкалам измерения) результаты тестирования, с использованием приведенных в таблице 2.9 методов оценивания тестовых заданий имеют распределение близкое к нормальному.

В таблице 2.9 предоставлены значения математического ожидания, дисперсии и среднеквадратического отклонения результатов, полученных с использованием каждого из приведенных методов оценивания.

Таблица 2.9 – Математическое ожидание дисперсии и среднеквадратического отклонения результатов

Метод оценивания	Математическое ожидание	Дисперсия	Среднеквадратическое отклонение
Существующий метод	49	146,25	12,09
Линейный метод	46,6	162,69	12,76
Логарифмический метод	47,5	145	12,04
Квадратичный метод	44,3	238,76	15,45
Метод квадратного корня	48,2	151,01	12,29

Из данной таблицы видно, что, самая большая дисперсия и самое меньшее значение математического ожидания в случае квадратичного метода

оценивания. Это значит, что средний балл, получаемый за тест меньший, а степень рассеивания баллов большая.

Контроль знаний тестовым методом вместе с многочисленными плюсами (удобством и высокой скоростью проверки и т.п.) определил целый ряд проблем. Не секрет, что зачастую составляются такие тесты, которые недостоверно оценивают знания студентов. Это чревато тем, что в категорию отличников могут попасть слабые студенты [37], а инверсия результатов, как показатель необъективности, это как раз то, чего классическая система тестирования предполагает избежать.

2.2 Статистическая проверка, значимость и мощность выводов

Проблема статистического вывода в социальных/социотехнических – заблуждения исследователей по поводу проверки статистической значимости нуль-гипотезы (ПЗНГ) [38] а именно:

- рутинное и ошибочное применение ПЗНГ в исследованиях;
- неадекватность ПЗНГ по отношению к целям и задачам исследований;
- неправильная интерпретация результатов ПЗНГ.

Краткое обобщение основных ошибок в интерпретации результатов ПЗНГ, представленных в обзоре Р. Клайна [39], приведено (https://studme.org/1260102029269/psihologiya/problemnye_kotorye_kazhdyy_raz_nuzhno_reshat_zanovo).

При разработке и принятии решений необходимо учитывать возможные ошибки, допускаемые при интерпретации статистической оценки событий (таб.2.10) [40]:

1) «Ошибка первого рода (false positive, «ложная тревога») состоит в том, что будет отвергнута правильная гипотеза.

2) Ошибка второго рода («пропуск цели») состоит в том, что будет принята неправильная гипотеза.

Правильное решение может быть принято также в двух случаях:

1) Гипотеза принимается, причем и в действительности она правильная.

2) Гипотеза отвергается, причем и в действительности она неверна»[33, 76].

Надо отметить наличие «помех» в выводах как результат эмоционально-содержательной трактовки терминов «ложноположительный/ложноотрицательный», которые не имеют отношения к желательности/нежелательности самого события.

Таблица 2.10 – Ошибки критериев

Результат применения критерия	Верная гипотеза	
	Н0	Н1
	Различий нет в действительности	Различия есть в действительности

	(результаты случайны и не имеют отношения к популяции)	(результаты неслучайны и отражают свойства популяции)
Решение Н0 Различия не выявлены с помощью критерия	Истинно отрицательный результат (правильное решение, верно принята Н0, интерпретация совпадает с реальностью), $1-\alpha$	Ложноотрицательный результат (ошибка II рода, неверно принята гипотеза Н0, интерпретация не совпадает с реальностью), β
Решение Н1 Различия выявлены с помощью критерия	Ложно положительный результат (ошибка I рода, неверно отвергнута гипотеза Н0, интерпретация не совпадает с реальностью), α (уровень значимости)	Истинно положительный результат (правильное решение, верно отвергнута Н0, интерпретация совпадает с реальностью), $1-\beta$ (мощность/чувствительность критерия)

«Уровень значимости α – это максимальная приемлемая вероятность ошибки I рода; того, что значение критерия окажется не меньше критического значения при условии справедливости нулевой гипотезы. Иными словами, α -уровень показывает вероятность ошибочного заключения о существовании различий, которых в действительности нет. Вычислительные среды часто автоматически рассчитывают достигаемый уровень значимости (p-value) – наименьшую величину уровня значимости, при которой нулевая гипотеза отвергается для данного значения статистики критерия. Как правило, естественно стремление снизить уровень значимости. Ошибка I рода находится под контролем исследователя»[76].

«Ошибка II рода β характеризует мощность или чувствительность критерия, которая равна $1-\beta$. Мощность – это вероятность отклонить нулевую гипотезу, когда она действительно ложна (вероятность не совершить ошибку II рода), т.е. это вероятность обнаружить реальные статистически значимые различия. Мощность нелинейно возрастает как функция размера эффекта, слабо контролируется (скорее – наблюдается с ожиданием отклика на коррекцию ошибки I рода). В практике системных исследований принято считать, что мощность, т.е. вероятность обнаружения значимого эффекта в случае его наличия, должна быть не менее 70÷80%»[76].

«При планировании эксперимента (активного или пассивного) мощность должна быть разумно высокой, чтобы обнаружить разумные отклонения от нулевой гипотезы. В противном случае, эксперимент не следует проводить»[33]. Поэтому для моделируемой интеллектуальной системы оценки результатов обучения декларируется следующее: ошибка II рода должна быть столь малой, насколько это возможно при фиксированном уровне ошибки I рода. Это обеспечивает максимально высокую мощность критерия.

Оценка разумных пределов мощности при оценке достижений учащихся, в том числе – тестированием, на данный момент актуальна и еще на стадии

исследования. На данный момент надо учитывать α -уровень значимости, устанавливаемого исследователем, направленности гипотезы и уровня значимости p -level; размера эффекта (в результате обучения сам эффект и его «величина» на стадии дискуссий) и его вариативности в популяции; количества испытуемых, т.е. размера выборки (что на текущий момент является большой проблемой). В плане понижения вариативности (например, более четкой дифференциации учащихся по критерию «качества» подготовки специалиста посредством оценки именно тестированием) в моделируемой системе декларируется блок статистической обработки с «тонкими» процедурами контроля, надежными методами и методиками получения и обработки данных, в том числе - повышающих мощность статистических выводов. А это, в свою очередь, требует внимательного отношения к экспериментальной схеме. В том числе:

- адекватного и корректного измерения переменных и фиксации в соответствующих шкалах;
- функционального контроля и управления условиями образовательного эксперимента (в том числе – тестирования: формальной/неформальной обстановки, наличия фактора соревновательности или других стимулов и т.д.);
- учета факторности (мультивариативности) эксперимента (например, статистические методы обработки могут предполагать/исключать взаимодействие тестируемых, независимость тестовых вопросов и т.п.);
- организации корректной схемы экспериментального контроля, что обеспечит внутреннюю валидность эксперимента-тестирования, «очищение» от факторов несистематической изменчивости, систематических и сопутствующих смещений (здесь проблемой является не только малые выборки результатов тестирования, но и факторы времени и последовательности).

Применительно к проблемам тестирования предполагается наличие вариативности расчетной схемы в зависимости от условий. Чем сильнее эффект, тем проще его обнаружить. И мощность можно и нужно рассчитывать только для конкретной альтернативной гипотезы, т.е. для конкретного эффекта. Проблема малых выборок контрольно-измерительного мероприятия (КИМ) демонстрирует известный факт: вероятность обнаружения небольшого эффекта, да еще при использовании ненадежных методик, изначально мала, тогда как в схожем исследовании, но с **большим** количеством результатов тестирования, этот эффект мог бы быть обнаружен.

Например, в контексте оценивания учебных достижений, при сомнении в достаточном уровне компетентности, учащемуся может быть выдано дополнительное задание/тест. Ошибка первого рода – когда плохие результаты задания, вызванные случайными причинами (учащийся болен, рассеян и т.п.) объявят неслучайными и учащегося, имеющего потенциальную возможность обучения, отчислят (либо оставят на летний семестр и т.д.). Ошибка второго рода – если неслучайные низкие результаты выполнения задания (тестовые показатели) объявят случайными (учащийся был рассеян,

плохо себя чувствовал и т.п.) и дадут еще один шанс продолжить обучение без изменения условий. «Цена» ошибки первого рода – репутация несостоятельности специалиста при потенциальной возможности достичь необходимого уровня компетенции. «Цена» ошибки второго рода – упущенная возможность/необходимость смены профессии, карьерного роста в другой сфере деятельности и т.п.

Вероятность «ложной тревоги», т.е. ошибки I рода, коррелирует с вероятностью «пропуска цели», т.е. ошибки II рода. Иными словами, чем более чувствительна система, тем больше «рисковых» событий она диагностирует – значит, предотвращает (на всякий случай отвергает). Но при росте $(1-\beta)$ -мощности системы неизбежно растет и α -вероятность ложноположительной реакции. А это может привести к тому, что побочный вред от очень чувствительной системы (пропустившей событие) будет превышать пользу.

«Так как с ростом вероятности ошибки первого рода обычно уменьшается вероятность ошибки второго рода, и наоборот, настройка принимающей решение системы должна представлять собой компромисс. Где именно находится точка получаемого такой настройкой баланса, зависит от оценки последствий при совершении обоих видов ошибок» [79].

Существует ряд факторов, которые непосредственно влияют на мощность критерия:

- Некоторые статистические тесты по своей природе имеют большую мощность, чем другие. Информация, которую приносят параметрические тесты вместе с их допущением, помогает повысить эффективность тестов. Параметрические тесты часто устойчивы к отклонениям от своих предварительных условий. При этом нулевые гипотезы различаются между параметрическим тестом и его непараметрическим аналогом. В частности, нулевая гипотеза для параметрического теста содержит конкретное параметрическое предположение о распределении статистики теста (которая и рассчитывается по-разному для двух тестов. Несмотря на то, что два значения α могут быть равны, но рассчитываются на основе разных тестовых статистических данных, которые имеют разные распределения при разных нулевых гипотезах.

- «Чем больше объем выборки, тем меньше β , и соответственно, выше мощность. Однако увеличение числа наблюдений связано с финансовыми и временными затратами.

- Чем меньше α , тем ниже мощность.

- Вариабельность наблюдений (величина экспериментальных эффектов);

- чем больше отношение величины различий к стандартному отклонению, тем выше мощность.

- Уровень ошибки в экспериментальных измерениях. Ошибка измерения интерпретируется как «шум», который может скрыть эффект в реальных экспериментах. Все действия, улучшающие точность и надежность измерения, могут увеличить статистическую мощность»[41].

Ошибку второго рода вычисляют для каждого вида проверки гипотез порознь. Интерактивная возможность для заданного значения α построить семейство кривых – операционных характеристик, которые иллюстрируют зависимость ошибки второго рода от Δ и n , предоставлена [41].

2.3 Особенности работы интеллектуальной системы оценки результатов обучения

«В условиях рыночной экономики, как уже отмечалось, модифицируется необходимость и уровень условий потребителей образовательных продуктов, что устанавливает потребность долговременного улучшения образовательного процесса как для эффективной работы на рынке образовательных услуг, аналогично для потребностей рынка труда. Одним из путей решения этой проблемы является внедрение передовых коммуникационных информационно-образовательных технологий, позволяющих образовательным учреждениям идти в ногу со временем и предоставлять студентам новые возможности и повышенную гибкость в процессе получения качественного образования. Увеличивающаяся надобность образовательного процесса в эффективных цифровых образовательных ресурсах определена неимением групповых программных решений, способных увеличить эффективность образования» [43].

В настоящее время преимущественно многообещающими направлениями по части компьютерного образования представляются адаптивные и мультимедийные технологии (адаптивные системы тестирования, адаптивные гипермедиа-системы и др.). Цифровые образовательные ресурсы, созданные на принципах адаптации и разнообразных вариантах обратной связи, основательно увеличивают эффективность и расширяют способности прогрессивного компьютерного преподавания.

«Возможные информационные системы широко употребляются для управления подготовкой специалистов, а также для обеспечения учебного процесса. Кроме этого, обширный эксперимент педагогических инноваций, авторские школы и инновационные педагоги, итоги психолого-педагогических исследований в развитии образовательных процессов безостановочно требуют обобщения и систематизации» [43].

«Адаптивное тестирование обуславливается как совокупность процессов формирования, представления и оценки итогов адаптационных тестов, увеличивающих действительность измерения по сравнению с традиционным тестированием посредством оптимизации подбора характеристик заданий, их количества, очередности и скорости представления применительно к характеристикам учебных предметов» [43].

«Адаптивное тестирование возможно сопоставить с устным экзаменом. После основного эффективного/успешного (неуспешного) ответа обучающегося, преподаватель задает наиболее сложный (более легкий)

вопрос, и, если обучающийся с ним справляется (не может справиться), преподаватель дает оценку 5 или 2. Чтобы гарантировать такой же высокий уровень надежности в случае «средних» людей, необходимо задавать больше вопросов, чем выявлять крайние случаи. Итак, смысл адаптивности охватывается в том, что тест адаптируется к уровню студента, что ускоряет локализацию его практического тестового балла» [43, стр. 29].

В данной работе на рисунке 2.7, представлен алгоритм управления интеллектуальной системой оценки результатов обучения, который позволяет проведение адаптивного теста в комплексе с применением устного/письменного ответа.

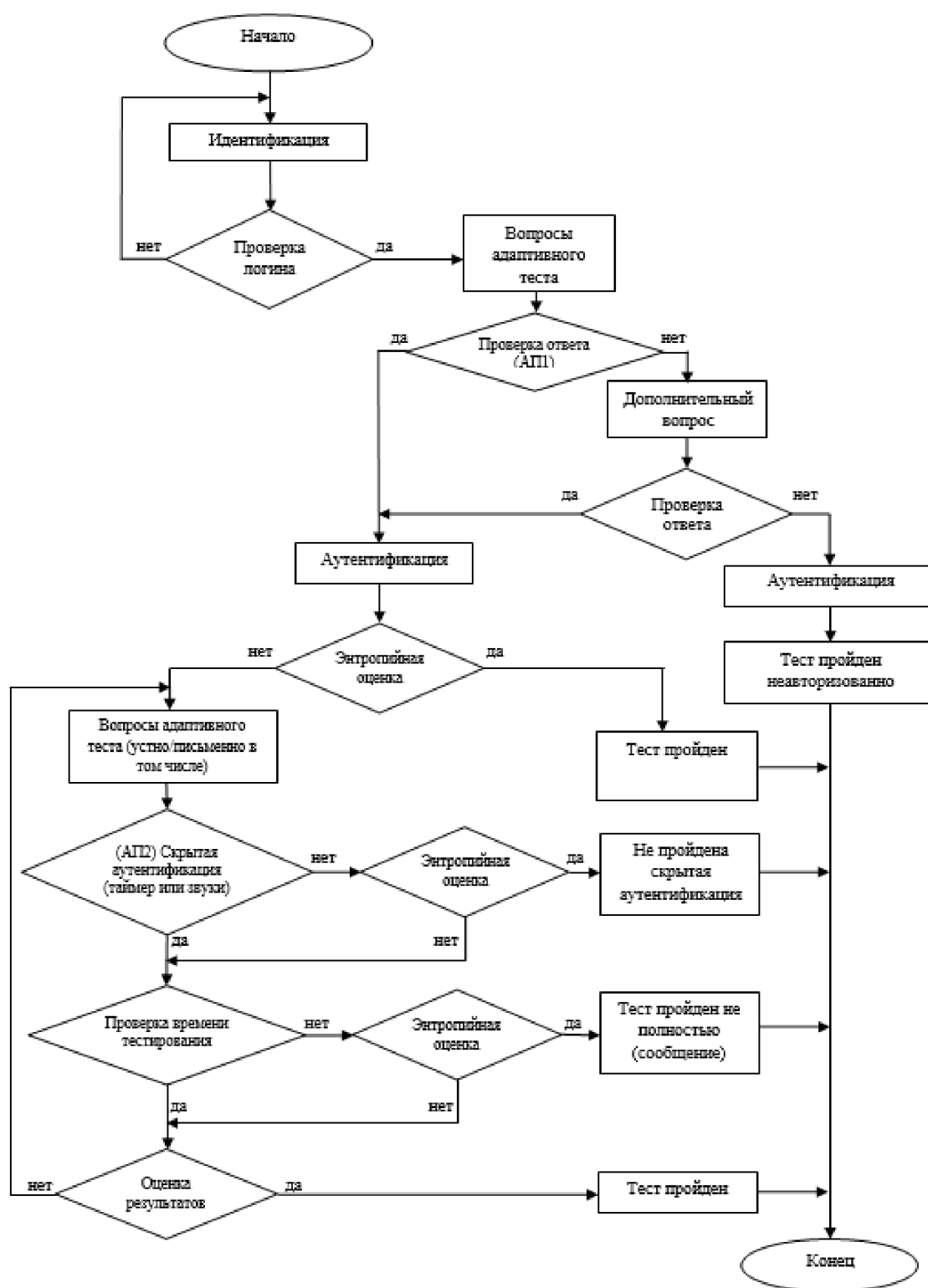


Рисунок 2.7 – Общий алгоритм управления интеллектуальной системы оценки результатов обучения

Прохождение определенных методов идентификации и проверки подлинности происходит на фоне применения достаточно длинного имени пользователя и ввода достаточно длинного пароля, которые создаются автоматически при регистрации студента в этой системе.

Предлагается задавать в начале, как вариант оценки знаний, задание в виде вопроса открытого типа (с помощью введения ответа устно или ввода клавиатурным способом), который одновременно используется для дальнейшей проверки аутентичности, если ответ будет правильным.

Аутентификация включает:

- контроль темпа ввода текстового сообщения с клавиатуры;
- или распознавание голоса диктора, если ответ озвучивался голосом;
- потенциально возможно распознавание лиц (при наличии видео- или веб-камеры).

Положительная аутентификация предваряет энтропийную оценку логики опроса и открывает доступ к адаптивным тестам (в том числе устного/письменного). Адаптивный процесс тестирования включает в себя как закрытые тестовые вопросы (с возможностью, или даже требованием ответа голосом), так и вопросы на соответствие, вопросы на определение последовательности, вопросы с несколькими выборами ответов и т.д.

Схема выделяется неординарностью структурно-логической модификации оценки взаимодействия и является качественно новым словом в применении в образовательном процессе. Это характеризует как адаптивный алгоритм тестирования, учитывающий модульную организацию содержания образования, так и все многообразие уровней сложности материала и подстраивается под естественное течение процесса взаимодействия.

2.4 Оценивание результатов обучения методом адаптивной оценки знаний

«Имеющийся на практике опыт показал, что одной из важнейших составляющих модернизации образовательного процесса является разработка объективных методов мониторинга знаний и оценки качества образования. Диагностика знаний имеет три основные функции:

1. Он контролирует процесс обучения. Контроль дает возможность увидеть ошибки, оценить результаты обучения, выполнить коррекцию знаний и умений.

2. Воспитывает познавательную мотивацию и стимулирует творческую активность. Правильно организованная система контроля позволяет пробудить чувство ответственности за результаты обучения (получить более высокие баллы), способствует стремлению к целенаправленной работе по исправлению ошибок;

3. Это средство обучения и развития (результаты применения представлены в работах)» [44].

«Организация тестовой базы данных основана на следующих принципах:

- тестовая база (иначе говоря, базовый тест, Банк тестовых заданий или базовый набор тестовых заданий) – это общий набор заданий для оценки утраты знаний, полностью отражающий содержание предметной области. В

этом случае тест представляет собой выборку заданий из тестовой базы данных, сформированной в зависимости от ответов испытуемого (студента). В конце теста на основе информации, полученной в результате анализа ответов на заранее объявленные тестовые задания, делается вероятностный вывод об уровне знаний;

- достоверность содержания базового теста была проверена экспертом, и в этом нет никаких сомнений;

- каждая задача тестовой базы данных имеет несколько атрибутов (параметров), основным из которых является сложность тестовой задачи;

Параметры тестовых заданий на начальном этапе определяют эксперты, в дальнейшем, по мере тестирования и накопления статистических данных, значения параметров тестовых заданий уточняются с помощью IRT –моделей» [45]. При этом оценивание знаний также осуществляется в соответствии с положениями математической теории оценки качества тестовых заданий [45, стр.78].

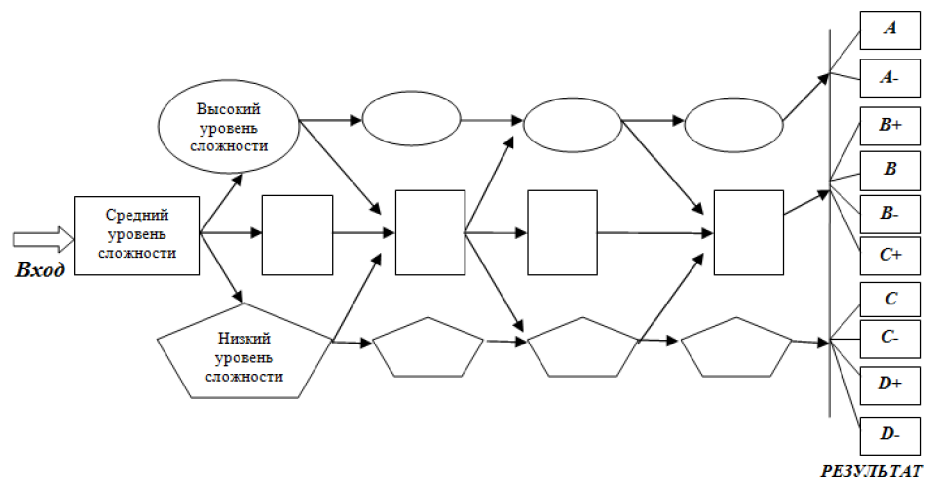


Рисунок 2.8 – Схема адаптивного теста

Согласно принципам организации тестовой базы тестовые задания являются «кирпичиками», из которых складывается тест – измеритель уровня знаний, и обладают рядом параметров, основным из которых является трудность.

При формировании адаптивных тестов в практических приложениях, как правило, используется модель Раша либо модель Бирнбаума [46].

Тем не менее, наиболее адекватно тестовое задание описывает четырехпараметрическая логистическая IRT-модель

$$P_j(\theta) = \gamma_j + (\lambda_j - \gamma_j) \frac{e^{a_j(\theta - \beta_j)}}{1 + e^{a_j(\theta - \beta_j)}} = \gamma_j + \frac{\lambda_j - \gamma_j}{1 + e^{-a_j(\theta - \beta_j)}}, \quad (3)$$

с учетом

$$|\Psi(1,7x) - \Phi(x)| < 0.01, \quad (4)$$

примет вид

$$P_j(\theta) = \gamma_j + (\lambda_j - \gamma_j) \frac{e^{1.7a_j(\theta - \beta_j)}}{1 + e^{1.7a_j(\theta - \beta_j)}} = \gamma_j + \frac{\lambda_j - \gamma_j}{1 + e^{-1.7a_j(\theta - \beta_j)}} \quad (5)$$

Учитывая то, что содержательная валидность базового теста, а именно содержательная валидность тестовых заданий, проверена экспертами и не вызывает сомнений, можно полагать $\lambda_j = 1$. Таким образом, равенство (5) примет вид

$$P_j(\theta) = \gamma_j + (1 - \gamma_j) \frac{e^{1.7a_j(\theta - \beta_j)}}{1 + e^{1.7a_j(\theta - \beta_j)}} = \gamma_j + \frac{1 - \gamma_j}{1 + e^{-1.7a_j(\theta - \beta_j)}} \quad (6)$$

что соответствует «трехпараметрической логистической IRT-модели и выражает вероятность $P_j(\theta)$ правильного решения испытуемым с уровнем знаний θ задания j , обладающего дифференцирующей способностью a_j и трудностью β_j , γ_j - вероятность угадывания верного ответа» [28].

Рассмотрим пример самого рядового тестирования, когда используется тест закрытого типа. Это может быть экзамен в сессию, вступительное тестирование или, например, ВОУД (Внешняя оценка учебных достижений). Видимые различия в объеме (количестве) тестовых заданий на тест, времени на заполнение и форме подачи (типе издания – бумажный или электронный). Поскольку консервативное исполнение на бумажном носителе вряд ли переводимо на «адаптивные рельсы», будем считать все рассматриваемые далее варианты тестирования имплементированными электронными версиями [28, стр. 63].

Опишем видимые (или предполагаемые в наличии) признаки типичного неадаптивного теста. Закрытая форма, объемы заданий (вопросов) от десятков до сотен, вопросы содержат от нескольких до десятка вариантов выбора (задания на соответствие или восстановление последовательности, как и открытая форма, принципиально мало что меняет, впрочем) [47]. Комплект тестовых заданий собирается «случайным» с вероятным контролем некоторых качеств – примерного уровня (баланса) трудности, сложности, валидности, надежности, точности, адекватности и возможно еще ряда подобного свойства характеристик. Наличие конкретной характеристики доказывается (обосновывается) экспертами и статистикой массового применения [48]. Как правило, осуществляется контроль времени отводимого на применение тестового комплекта, либо в форме «в целом» на комплект, либо по каждому заданию в тесте в отдельности.

Поясним, почему сильные вариации в оформлении комплекта не принципиальны. Открытый тест, по завершении либо потребует контроля и оценки «в ручном» режиме экспертом, либо развитого ИИ, идеально имитирующего интеллект человека-эксперта. Иначе открытое тестирование сводится к сверке ответов с более-менее «продвинутом» шаблоном и все-таки

потребуется оформление ответа на достаточно формальном языке (тезаурусе) и знакомства тестируемого с этим тезаурусом и шаблоном. Что, видимо, сведется к выбору в пределах десятка альтернатив (с вариациями), то есть неоформленному, но подразумеваемому «закрытию» теста по форме. Тест оформленный как задание на установление соответствия (многих ко многим) преобразуется к тесту на соответствие один-к-одному с закрытым выбором истина-ложь в смысле соответствия. То есть это стандартная форма закрытого теста, но собранная в компактное представление в виде единственного задания на соответствие. Аналогично преобразуется и тест на восстановление последовательности, преобразование приводит к комбинаторному взрыву в случае демонстрации «всех» последовательностей (соответствий) в отдельных заданиях с выбором истина-ложь, но его легко избежать, сохраняя только «типичные» ошибки выбора последовательности (соответствия). Собственно, и все различные по форме «закрытые» тестовые задания также приводимы к бинарной форме с выбором истина-ложь по отношению к завершённому связному предложению из вариантов связки вопрос-ответ исходного тестового задания [48, стр. 7].

Снижение затрат времени, особенно экспертов, формируемых в группы, можно решать переходя в тестирование адаптивное, которое, как отмечалось выше нереализуемо на бумажной основе, но активно используется лидерами технологий (например, Microsoft).

Что дает адаптивное тестирование? Сокращение объемов подготавливаемого тестового материала, так как адаптивная схема имеет возможности более тонкой «настройки», учитывающей поведение тестируемого в процессе проведения теста, что сокращает как количество вопросов, генерируемых в тестовом комплекте, так и время, отводимое на собственно тестирование [48, стр. 9].

Рассмотрим адаптируемое тестирование подробнее.

Нулевой вариант. Адаптивное тестирование спровоцированное самим тестируемым. Предположим наличие студента, пришедшего на экзамен, с текущим рейтингом по предмету такого уровня, что максимум, на который он может рассчитывать по предмету, по итогам экзамена – «не F», даже при идеальном ответе на тестовый комплект. Такой студент может пролистать весь тест, обнаружить половину заданий, которые ему кажутся «простыми», «знакомыми» или просто известными в силу каких-то не обсуждаемых здесь причин, отметить в них ответы и – завершить тест, в «лучшем случае» протыкав-пометив наугад оставшуюся незнакомой половину тестового комплекта. Итог такого «адаптивного» оценивания повлияет абсолютно минимально на самого студента, но на качество тестовой базы, в статистическом смысле обработки его результатов, этот результат не повлияет и самое разумное, отказаться от заметной части тестовой базы в силу ее «общеизвестности».

Простейший вариант. Адаптивность реализована индивидуальным подбором времени на ответ каждого тестового задания. Очевидно, назначение

самого минимального времени на выбор варианта ответа, компьютерное тестирование «адаптирует» поведение экзаменуемого в том смысле, что он успеет ответить только на «знакомые» ему тестовые задания. Конечно, его либо стоит о таком выборе предупредить, либо он какую-то часть заданий потратит на осознание этого факта.

Вариант на основе простейшего. Реализуется невозможность вернуть и перевыбрать отмеченный ранее выбор в тестовом задании, вплоть до реализации движения в тестовом комплекте «только вперед». Время ограничивается только на тест в целом, а объем тестовых заданий, предлагаемых тестируемому, определяется только емкостью базы заданий. Часть испытуемых, вероятно, пойдет в этом случае, по нулевому варианту, особенно если «в зачет» идут только задания с совершенным выбором.

«Классическая» адаптация. Все задания, даже внутри одной серии, скорее всего «невозвратной» реализации, как в простейшем случае. Временные ограничения либо на тест целиком, либо в пределах каждой серии одного уровня, сам уровень может подниматься ужесточением контроля времени. Первые несколько тестовых заданий предлагаются из заданий классифицируемых в базе тестовых заданий как задание на «С». Верный выбор в большинстве из них, переключает задания на уровень «В», если выбор по-прежнему преимущественно верный, то на задания уровня «А». В случае сохраняющейся тенденции экзаменуемый получит А, иначе – А-. Если тенденция «сломалась» раньше, то оценка В. Если первая серия уровня С «не задалась», предлагается уровень «D», повтор неудачи и в этой серии вопросов приводит к завершению на уровне оценки F. Средний уровень серии, выбранной первой, предполагает сокращение объема тестового комплекта для всех тестируемых, поскольку «экстремалам» не придется проходить тесты всех уровней, а «середничкам» все уровни и предъявлять не придется. Таким образом, адаптивная схема допускает тестовые задания с параметризацией, поскольку ограничивает объемы по сравнению с простыми надерганными наугад тестами из большой базы. Тратится много ресурсов на разработку относительно компактной базы алгоритмов и шаблонов, стратифицированных по уровням. Неясно как прогнозируются уровни вопросам включаемых в уровень, если это не экспериментально подобранные тайминги. Более того, шансы угадать ответы для медианной серии уровня «С» критичны для всего хода тестирования[49].

Вариация предыдущей адаптивной схемы. Серия, переключающая уровень состоит из одного тестового задания, правильный ответ повышает уровень вопроса, неправильный выбор – понижает, объем заданий максимальный, если средняя оценка уровня правильных ответов нестабильна и выставляется как усредненная в любом случае, впрочем, если серия сопровождает некоторый уровень осреднения с малым отклонением, то тест завершается раньше. Алгоритм завершения можно предложить следующий – на протяжении первой трети [50] тестовых заданий фиксируется лучший уровень (как вариант, средний), далее, по достижению (превышению) этого же

уровня оценки тестирование завершается с достигнутой оценкой, иначе, тест завершается исчерпанием времени тестирования (всего запланированного ранее объема тестового комплекта) и усредненным по всему тесту результатом (или медианной оценкой, как более устойчивой к случайностям). Сложности назначения уровня тестовым заданиям те же, что и ранее.

Вариация предыдущей адаптивной схемы. Можно предлагать тестируемому, на определенном этапе (или постоянно), выбирать уровни задания, если своими ответами тестируемый выбор не хуже подтверждает более какого-то количества раз, тестирование прекращать на достигнутом уровне. Эта версия предполагает движение от самых простых заданий.

Вариации адаптивной схемы можно продолжать, но главное тут – разметка содержимого тестового комплекта по уровням. Можно предложить строить тестовую базу адаптивного тестирования по максимальной аналогии с устным экзаменом. Пусть движок базы допускает участие тестируемого и эксперта-экзаменатора в процессе тестирования «интерактивно-двусторонне», то есть: эксперт присутствует на тестировании как часть процесса. Движок позволяет ему вмешиваться в процесс, переназначая тестовое задание для экзаменуемого, в том числе назначая вновь уровень заданию [6, стр.56].

Протокол далее будет использован для настройки системы принятия решений при подаче тестовых заданий в автоматическом (или на ранних стадиях автоматизированном, но при участии эксперта) режиме. Сам протокол будет просматриваться и утверждаться для аппликации в базе тестов минимум еще одним независимым экспертом. Стартовать тестирование может с базы тестов минимального наполнения и разметки по уровням, просто чтобы понизить трудозатраты эксперта в интерактивном режиме по выбору и разметке (переразметке) уровней задания. Задания имеет смысл готовить на простейших уровнях в виде закрытых бинарных тестов (обсуждали выше по тексту), затем в версии с выбором единственного варианта из многих, затем неединственного выбора из многих, затем, с учетом присутствия эксперта в процессе – допускать открытые варианты ответов (в дальнейшем их можно «закрывать», включая в множественный выбор с помощью эксперта прямо в интерактивном режиме). Понятно, что такой интерактивный режим тестовой базы должен предполагать совершенно другой регламент учета времени теста и строго последовательную сдачу экзамена тестируемыми, чтобы не превращать эксперта в «многостаночника». Возможно, стоит начинать такие эксперименты с малокомплектных групп [50, стр.9].

Открытые и полуоткрытые версии тестовых заданий могут преобладать на ранних этапах разработки адаптивных тестов в интерактивной среде, в том числе и с целью накопить статистику интервалов времени между нажатиями клавиш клавиатуры и мыши по тестируемым и экспертам для косвенных методов опознания работающего на ПК пользователя, исключая, таким образом подлог и соучастие.

2.5 Выводы по второй главе

В настоящее время проверка знаний методом тестирования наряду с множеством преимуществ (удобство и высокая скорость проверки и т.д.) демонстрирует ряд проблем. В частности, часто разрабатываются тесты, неверно оценивающие знания студентов.

В связи с этим был проведен анализ нескольких методов тестирования, сравнение их с уже применяемым в СКУ подходом и между собой. Приведены методы оценивания с количеством баллов, а также основных описательных статистических характеристик результатов, полученных с помощью вышеперечисленных методов оценки.

Подчеркивается проблема статистического вывода в социальных / социотехнических заблуждениях исследователей о проверке статистической значимости нулевой гипотезы, а также приведены критерии ошибок.

Построен алгоритм управления интеллектуальной системой оценки результатов обучения, позволяющий решить задачу применения энтропийной оценки для адаптивного тестирования в сочетании с использованием устного / письменного ответа.

3 ОЦЕНКА РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПОСРЕДСТВОМ ПРИМЕНЕНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ

3.1 Принципиальная структура интеллектуальной системы

Отметим, что внедрение инновационных образовательных технологий в непрерывном образовании становится все более актуальным и эффективным инструментом профессионального развития. Тесты могут использоваться на любом этапе обучения. Одни из них предназначены для оценки готовности обучающихся к освоению нового учебного курса, другие для выявления конкретных пробелов в знаниях. Также тесты используются с целью планирования необходимой целенаправленной корректирующей работы, для прогнозирования дальнейшего процесса обучения и для определения компетенций специалистов организаций. Основное тестирование включает проведение входных, текущих, рубежных контролей, промежуточной аттестации по каждой учебной дисциплине, в том числе и внешней оценки учебных достижений (ВОУД), государственной итоговой аттестации.

Максимально «широкое внедрение информационных технологий в образовательную систему и решение проблемы качества образования требует изучения и разработки моделей, соответствующих процессам образовательной деятельности высших учебных заведений. Повышение качества образования является основным направлением развития системы образования. Одной из важнейших задач управления качеством высшего образования является задача мониторинга и оценки качества образования. В современном информационном обществе автоматизированное тестирование может стать основным инструментом» [51], [75, стр. 115].

В данной работе представлены результаты изучения проблемы развития оценки результатов обучения методом тестирования. «Анализ систем управления знаниями студентов показал, что разработка систем в основном базируется на эмпирической базе. Без должного научно-методического обоснования, без опоры на теоретические модели готовности специалиста»[51]. «Объективные потребности вузовской практики, при реализации программы повышения качества подготовки специалистов необходимо разработать научно обоснованную систему мониторинга знаний обучающихся. Сегодня существует техническая проблема, заключающаяся в совершенствовании процесса автоматизированной оценки результатов тестирования обучающихся. Решение, которое повысит эффективность контроля знаний, обеспечит объективность оценки выполнения теста как на уровне усвоения знаний тестовых» [75] материалов, так и на основе теста полностью. Для решения технической задачи предлагается метод оценки знаний студентов, основанный на использовании адаптивного подхода и вопросно-ответных отношений как результата взаимодействия с потоком документов характеризуемых параметрической энтропией.

3.2 Процедуры управления процессами интеллектуальной системы оценки результатов обучения

В работе [45, стр. 11] представлен метод генерации тестов, основанный на вычислении достижимого и связанного недо-приближения абстракции системы событий. Проблемы формального описания процессов системы обучения и аттестации методом тестирования были рассмотрены в исследованиях [52], [53], [54], [55], [56]. В некоторых работах [57] предложен метод визуализации и интерпретации результатов мониторинга сохраняемых знаний. Обоснована возможность его использования для анализа проблем обучения на индивидуальном и групповом уровнях.

Согласно изученному материалу об использовании различных методов при разработке, а также проведения компьютерного тестирования приведены результаты которые улучшат способность выявлять несложные прохождения теста и выполнять индивидуальную проверку баллов [57, стр. 504], [58, 59].

«Существует несколько подходов к оценке ТЗ, которые рекомендуются ведущими специалистами в данной области. Данная разработка базируется как на международном опыте тестирования, так и на российском и казахстанском (ими разработаны специальные рекомендации, отличающиеся от международных по подготовке тестов, целям и методике оценки результатов). В.С. Аванесов (ведущий специалист в области теории педагогических измерений) «предлагает в каждом последующем задании за правильный ответ давать один балл, за неправильный ноль [8, стр.7]. В то же время он отмечает существование «более продвинутых» схем оценки, например, в тестовых центрах западных стран. Задание сокращается в соответствии с его вкладом в общее изменение тестовых баллов испытуемых, то есть балл в каждом конкретном случае будет разным и зависит от конкретного набора вопросов. В конце теста автор предлагает зафиксировать баллы за угадывание (это актуально для вопросов с одним правильным вариантом). В случае вопросов с несколькими правильными вариантами он считает нецелесообразным»[8] вносить коррективы в оценку.

Некоторые не решенные проблемы, связанные с оценкой обучающегося методом тестирования рассмотрены в работе [60]. В современных исследованиях [61] сравниваются различные алгоритмы классификации для прогнозирования и улучшения успеваемости обучающихся. Полученные результаты помогают вовремя прогнозировать итоговую успеваемость обучающихся и разрабатывать мероприятия, которые могут улучшить их итоговую оценку.

В ходе имитационного исследования был предложен адаптивный тест [62], ориентированный на усилия, который позволил ослабить неправильную ориентацию на предмет, вызванную немотивированным прохождением теста. Некоторыми исследователями разработаны нечеткие правила изменения уровня сложности при адаптивном тестировании [63, 64, 65,]. Авторы исследования которых [66] посвященное процессу проектирования и

разработки модуля компьютеризированного адаптивного тестирования, интегрированного в UZWEBMAT. UZWEBMAT является экспертной системой, поддерживающей адаптивную и интеллектуальную индивидуальную среду электронного обучения. Он был посвящен преподаванию вероятностного предмета.

«Классическая теория испытаний имеет ряд спорных допущений, а результаты ее применения имеют серьезные практические недостатки. В частности, было отмечено, что при оценке знаний студента с помощью тестов различной сложности можно получить различные представления о достижениях студентов. Статистика, рассчитанная в рамках классической теории тестов, позволяет получить относительное положение каждого испытуемого в стандартной выборке. Однако они не могут быть использованы для объективной оценки значений параметров, характеризующих уровень знаний испытуемых и сложность тестовых заданий. Вопрос «какова объективная оценка уровня подготовленности студента по предмету?» в классической теории тестов остается открытым. Он попытался ответить на этот вопрос в рамках другого методологического подхода к созданию педагогических тестов. Также можно интерпретировать результаты тестов в рамках так называемой response theory (IRT)»[7. стр. 115], но, как показала практика, эта теория также имеет ряд серьезных недостатков[67].

Недостатком имеющихся теорий является тот факт, что, так или иначе, все эти теории предполагают обязательное наличие «пилотажного тестирования» и большого количества «испытаний» для каждого задания. В реальной жизни не каждое высшее учебное заведение себе это может позволить. А особенно это неприемлемо для тестирования студентов по предметам, содержательная часть которых меняется год от года (например, информатика и т.п.). Учитывая, что по спец предмету на межсессионном контроле тестируется явно недостаточное (для определения качества теста по предложенным теориям) количество студентов, тесты недостаточно объективно оценивают знания студентов.

Разумеется, полученный «случайным» образом тестовый комплект из унифицированного к бинарному представлению тестового пула заданий будет в разы более громоздким. Но это нивелируется отчасти отбрасыванием «нетипичных» ошибок и параметризацией оформления тестового задания. То есть, счетные задания могут генериться из формул на основе датчика случайных чисел. Чертежи (рисунки) также восстанавливаться из параметризованных случайных параметров работающих алгоритмов. Так же алгоритмически можно задавать варьируемые тексты. Очевидно, затраты на так подготовленную тестовую базу вырастут кратно, в том числе и затраты времени экспертов в связке с программистами. Видимо, эта проблема одна из тех, что мешает такому прочтению компьютерной версии тестирования [47, стр. 5].

Одним из вариантов преодоления данных затруднений была разработка автоматизированного тестирования на уже имеющейся в высшем учебном

заведении (ВУЗ) собственной оболочки WebTest. Все это позволяет утверждать, что целесообразным является проведение исследования, посвященного методу оценивания результатов автоматизированного тестирования.

Система оценок, когда знания характеризуются такими общими терминами, как «хорошие», «прочные», «глубокие» во многих случаях является неудовлетворительной. Четырехбалльная рейтинговая система также несколько меняет свое положение, если она основана на субъективных представлениях о знании и основана на интуитивных количественных критериях. Такая система удовлетворительно выполняет свои функции при оценке знаний человеком (преподавателем). Но при использовании для целей обучения технических средств обучения и контроля такая система оценок принципиально неприменима, так как технические устройства не обладают интуицией. Машина может работать только с формализованной и количественно определенной информацией [67, стр. 799].

Обычно знания определяются (с определенной степенью точности) путем постановки основных «контрольных заданий обучаемому и сравнения ответов обучаемого с теми, которые считаются заведомо правильными (эталонными). Понятие «элементарный контрольный вопрос» в данном случае является относительным. Она элементарна только тем, что в пределах рассматриваемой области не делится на более простые. Однако его содержание может быть довольно сложным. Понятие «ответ студента», рассматривается в широком смысле. В общем, это ответ на запрос. Она может быть детерминированной или вероятностной и может быть выражена не только словами, но и действиями, такими как выполнение определенных операций»[36].

Введем понятия «элементарное знание» и «общее знание». Под элементарным знанием понимается такое знание, которое относится к одному элементарному вопросу. Она определяется путем сравнения информации ответа с справочной информацией. В соответствии с общими знаниями мы будем рассматривать знания, относящиеся ко всем вопросам, задаваемым в рамках данной темы или курса. При этом вопрос о том, какое количество элементарных заданных «вопросов необходимо поставить, чтобы надежно оценить уровень общих знаний, а также вопрос о том, какой сложности должны быть при этом вопросы, относятся к области конкретных методик и в данном случае не рассматриваются. Полагаем, что разработанная система (модель) позволяет задать такое (достаточно большое) количество вопросов, которое во всех случаях обеспечит достаточно достоверную идентификацию уровня общих знаний по теме, курсу или его частям. То есть модель всегда удовлетворяет условию, где $Q \geq q$ - количество вопросов»[32], разрешенных моделью (возможности модели); q - количество вопросов, необходимых для достоверного определения уровня общих знаний (психолого-педагогические условия).

«В качестве количественного критерия знаний можно использовать степень соответствия T_i информации, содержащейся в ответе студента на i -й вопрос, той информации, которая считается правильной по отношению к данному вопросу (эталонной)

$$T_i = f(X_i, X_i'), (19)$$

где X_i – код обратной связи по i -му вопросу; X_i' – код справочной информации, соответствующей i -му вопросу. На сегодняшний день этот подход / метод широко используется. Предложен метод, в котором при определении уровня общих знаний учитывается не только степень соответствия информации, содержащейся в ответе студента, справочной информации, но и количество допущенных ошибок. В качестве критерия ошибки можно использовать степень различия между F_i и тех же типов информации:

$$F_i = f_l(X_i, X_i') (8)$$

Величина T_i , определяемая по формуле (19), будет рассматриваться как оценка знаний по i -му вопросу, величина F_i , определяемая по формуле (20) – ошибка, характеризующая незнание i -го вопроса. Величина балла T_i связана со степенью соответствия в прямо пропорциональной зависимости: чем выше степень соответствия, тем выше балл. Величина ошибки F_i обратно пропорциональна степени соответствия: чем ниже степень соответствия, тем выше ошибка. Такое отображение количественных критериев, характеризующих базовые знания, хорошо согласуется с интуитивным представлением о соотношении оценок и ошибок (чем больше ошибка, тем ниже оценка, и наоборот)» [37]. Общий балл S_i за i -й первому вопросу можно определить по формулам (21) и (22).

$$S_i = T_i (9)$$

$$S_i = T_i - F_i (10)$$

В первом случае она может быть получена на основе анализа оценки знания i -го вопроса (традиционное «четкое» тестирование), во втором, предлагаемом нами – с учетом допущенных ошибок (как разность оценок знания и ошибки i -го вопроса).

«В соответствии с положениями математической теории оценки качества тестовых заданий можно рассматривать уровень знаний θ как генеральную долю правильных ответов на все мыслимые тестовые задания, отображающие предметную область. Таким образом, по данным тестирования можно найти

точечную $\tilde{\theta}_k$ и/или интервальную $\left[\tilde{\theta}_k^1; \tilde{\theta}_k^2 \right]$, статистическую оценку θ .

Как известно из курса математической статистики, точечной оценкой генеральной доли является выборочная доля» [59, стр. 345] , то есть

$$\tilde{\theta}_k = p \quad (11)$$

где p — процент правильных ответов испытуемого / студента? Как определено в (23). Эта оценка объективна, последовательна и эффективна, но она является лишь приблизительной величиной , и точность аппроксимации можно считать достаточной для практических выводов только в том случае, когда испытуемому/студенту было предложено большое количество тестовых заданий ($k \rightarrow \infty$). Для тестов с небольшой длиной вопрос о точности оценки

может быть решен следующим образом. Задав вероятность δ , определим $\tilde{\theta}_k^1$ и $\tilde{\theta}_k^2$ так, чтобы выполнялось соотношение

$$P\left(\tilde{\theta}_k^1 \leq \theta \leq \tilde{\theta}_k^2\right) = \delta \quad (12)$$

Иными словами, найдем интервальную оценку θ с надежностью δ .

Традиционно границы доверительного интервала $\tilde{\theta}_k^1$ и $\tilde{\theta}_k^2$ для генеральной доли определяют, как решения уравнений

$$\sum_{m=kp}^k C_k^m \left(\tilde{\theta}_k^1\right)^m \left(1-\tilde{\theta}_k^1\right)^{k-m} = \frac{1-\delta}{2} \quad (13)$$

$$\sum_{m=0}^{kp} C_k^m \left(\tilde{\theta}_k^2\right)^m \left(1-\tilde{\theta}_k^2\right)^{k-m} = \frac{1-\delta}{2} \quad (14)$$

При $k > 20$ можно использовать приближенные формулы.

Одним из важнейших характеристик тестовых заданий является трудность. В классической тестологии существуют только статистические методы определения трудности заданий. Однако на практике, особенно для вступительного тестирования, попытка доопытного установления уровня трудности особенно актуальна, т.к. не для каждого задания из банка ТЗ (из-за принципа неразглашения ТЗ до вступительных экзаменов) имеется возможность определить трудность статистически. Но с другой стороны, невозможно вести речь о составлении сбалансированных вариантов вступительных ТЗ без учета такого качества ТЗ, как трудность.

Конечно, трудно добиться полного совпадения значений априорной и статистической трудности, т.к. качество определения трудности зависит не только от того, насколько хороша и продуманна система определения

трудности ТЗ, предложенная в настоящей работе, но и от того, насколько профессиональны и будут эксперты, работающие с ней. Но так или иначе, система будет в процессе работы самосовершенствоваться, а именно:

1. Корректировать априорные оценки трудности имеющимися статистическими оценками;

2. Оценивать, насколько корректно мнение эксперта и выдавать соответствующую информацию (предупреждать) об этом.

Для попытки априорного определения трудности первым делом необходимо установить факторы ($x_1, \dots, x_n$), влияющие на трудность, и установить их весовые коэффициенты. Априорной трудностью ТЗ будем считать величину, определяемую соотношением:

$$AT(x_1, \dots, x_n) = \sum_{i=1}^n T(x_i) \quad (15)$$

где $T(x_i)$ – количественное значение фактора трудности i для ТЗ.

Совместно с опытными преподавателями-экспертами в области вступительных тестовых заданий отображенную в (Таб. 3.1), была разработана система факторов, влияющих на трудность задания, для априорного определения трудности тестового задания по любой дисциплине:

Таблица 3.1 – Факторы, влияющие на трудность задания для априорного определения ТЗ

Критерий трудности	Качеств.хар-ка критерия	Итоговый ранг
время, необходимое для восприятия задания	большое	0,06667
	среднее	0,065
	малое	0,03
время, необходимое для решения (мин)	большое	0,06667
	среднее	0,055
	малое	0,01
трудность (по Аванесову)	уровень 1	0,11111
	уровень 2	0,09
	уровень 3	0,06
	уровень 4	0,03
	уровень 5	0,08889
трудность (по Беспалько)	уровень 0	0,08889
	уровень 1	0,04444
	уровень 2	0,06
	уровень 3	0,055
содержание	уровень 4	0,11111
	однородное (1 тема)	0,07
смысл	разнородное (неск-ко тем)	0,08
	очевидный	0,065
	неочевидный, без подтекста	0,06
	с подтекстом	0,085
по количеству элем. действий	на сообразительность	0,08
	1	0,04
	2 или 3	0,08
	>3	0,07
Требуется ли дополнительные построения	строго да	0,051
	в зависи-ти от уровня абитуранта	0,055
	строго нет	0,06667
время, необходимое для решения эксперту	>60с	0,11111
	40-60	0,1
	20-40	0,03
	5-20	0,06
	<5	0,03
может ли показаться очевидным на первый взгляд или "сбить с толку"	да	0,06
трудность темы	нет	0,02222
	высокая	0,06667
	средняя	0,03
по звеньям структуры данных	низкая	0,04
	мировоззренческий минимум	0,06667
	базовые знания	0,02222
	программные знания сверх базового уровня	0,08
содержание ТЗ	сверхпрограммные знания	0,08889
	вычислить	0,06667
	упростить(применение формул,законо-	0,04
	текстовая задача	0,06
Решается ли устно (без использования бумаги и записи)	строго да	0,06667
	в зависи-ти от уровня абитуранта	0,04
	строго нет	0,06

По системе априорного определения трудности ТЗ измерялась трудность 20 вступительных ТЗ по математике и сравнивалась со статистической трудностью, измеренной на около 500 абитуриентов. Результаты изображены на рисунке 3.1. Ошибка измерения этого опыта составила плюс минус 7%, что можно считать допустимым.

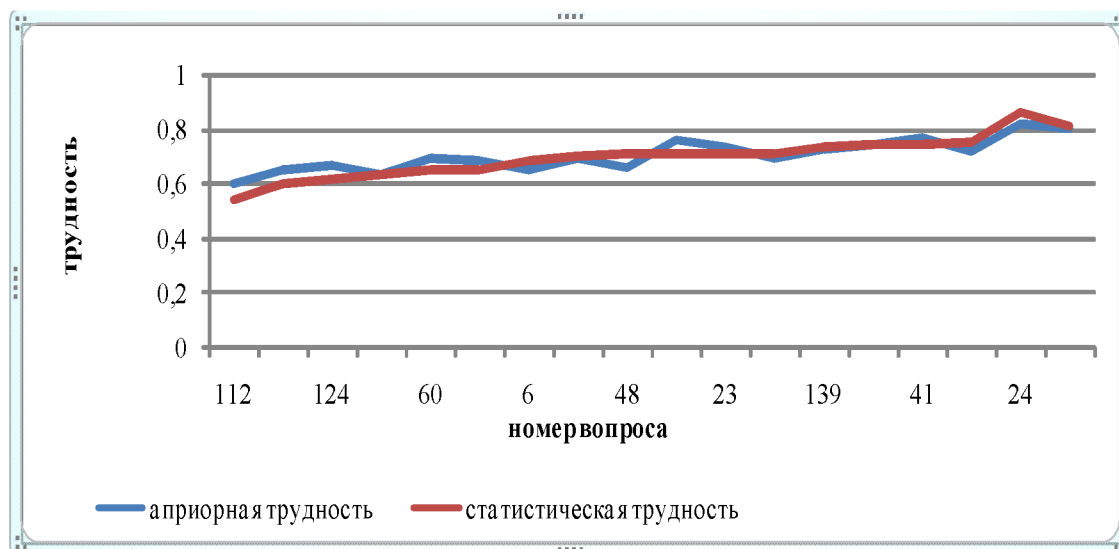


Рисунок 3.1 – Априорное определение сложности/трудности тестового задания

«Предложенный в данной работе метод позволяет выделить группу испытуемых, угадавших с высокой вероятностью, и группу «неправильных/плохих» вопросов, на которые большое количество обучающихся ответили случайным методом»[68].

Как видно в таблице 3.2 и на рисунке 3.2, при использовании данной методики результаты выравниваются.

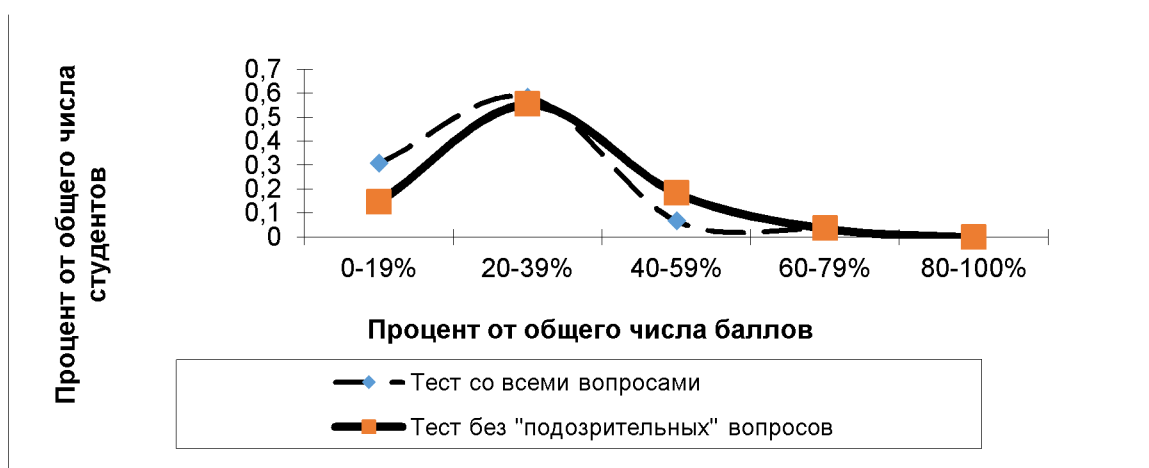


Рисунок 3.2 – Диаграмма результата тестирования

Таблица 3.2 – Результаты теста с исключением подозрительных вопросов

	Тест со всеми вопросами		Тест без "подозрительных" вопросов	
0-19%	9	31%	4	15%
20-39%	17	59%	15	56%
40-59%	2	7%	5	19%
60-79%	1	3%	1	4%
80-00%	0	0	0	0
	29		25	

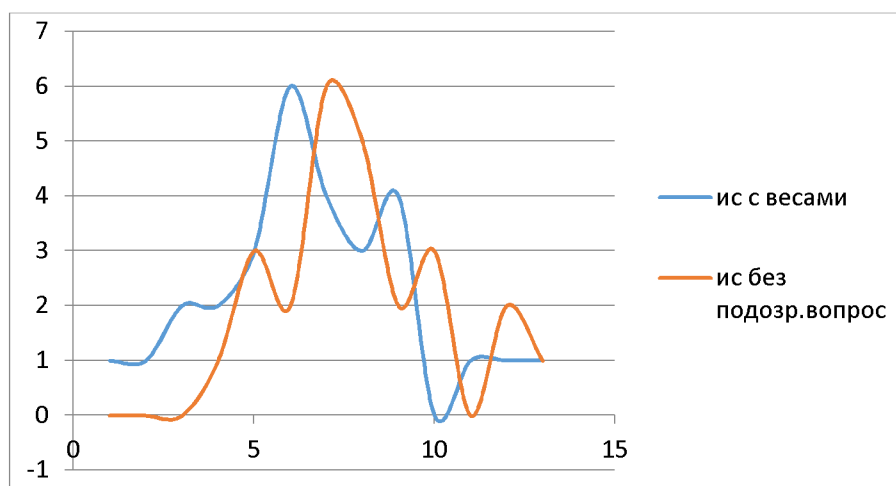


Рисунок 3.3 – Шкала оценивания тестовых заданий

Более подробную картину можно увидеть на рисунке 3.3, где фактически шкала оценивания «удлинилась» наполовину и приобрела более «нормальный» характер[67].

3.2 Банк тестовых заданий

В настоящей работе основным принципом является разделение создания банка тестовых заданий и процесса составления тестов (банка тестового материала) и тестирования. Банк тестовых заданий должен функционировать, как отдельная законченная система, и она может быть в дальнейшем использована различными системами тестирования. Главное назначение банка тестовых заданий является организация ввода, хранения и обработки большого количества задач и вопросов по разным предметам и представление пользователям широких возможностей по многоаспектному применению и использованию имеющейся информации.

Подготовительный этап организации тестирования включает в себя создание (в течение всего учебного года) банка тестовых заданий и банка тестового материала в виде электронной базы данных (БД). Тестовые задания классифицируются по темам и по уровням трудности, которые на начальном этапе проставляют эксперты (преподаватели высокой квалификации). Каждое

тестовое задание должно удовлетворять определенной БП для отбора тестовых заданий, требования к которой разработаны в главе 2. Банк сбора данных экспертов содержит в себе информацию о рекомендациях экспертов (преподавателей высокой категории), касающихся качеств тестового материала, разбиения на темы, уровни трудности и т.п. «Выверенный» экспертами тестовый материал помещается в банк тестовых заданий, задания в котором классифицированы по указанным критериям.

Первое требование, которое накладывается на любой тест – это полнота охвата всех разделов предмета. Для удовлетворения этому требованию, целесообразно разбить все вопросы на несколько тем и сгруппировать вопросы по темам. Элементами разбиения при этом могут служить, как и не связанные между собой разделы предмета, так и группа вопросов, удовлетворяющих одинаковым свойствам. Например, если мы хотим какой-то вопрос включить во все билеты будущего теста, то достаточно рассмотреть этот вопрос как отдельную тему, ибо в каждый билет войдет, хотя бы по одному вопросу из каждой темы. Подробнее схема создания банка тестовых заданий отображена на рисунке 3.4 и выглядит следующим образом:

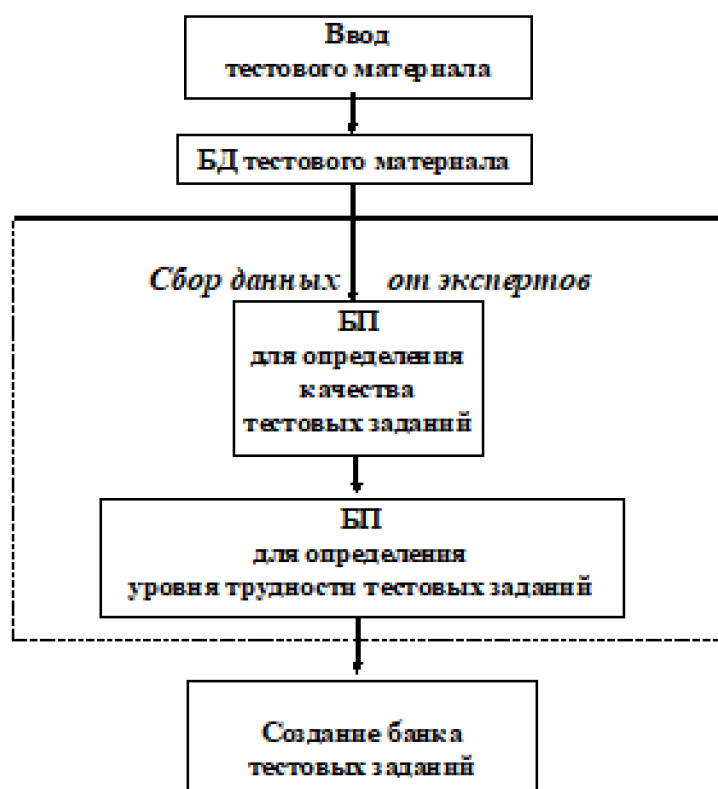


Рисунок 3.4 – Схема создания банка тестовых заданий

Выше описанный блок автоматизированной системы управления тестированием может подготавливаться в течении всего учебного года. При разработке модели автоматизированного банка тестовых заданий использована технология объектно-ориентированного проектирования. Главным объектом всей системы является задание. Каждое задание

характеризуется категорией (четкое – нечеткое), качеством («уд», «хор», «отл»), предметом, темой, модулем, которым оно принадлежит, трудностью, временем, необходимым для ответа, автором и экспертом, который его выверял и оценивал. Средним временем, трудностью и количеством ТЗ также характеризуются УЭ, модуль, предмет (для более легкой качественной и количественной оценки банка тестовых заданий). Содержание задания и ответов может содержать текстовую, графическую, формульную информацию и рисунки. При этом содержание задания может быть создано непосредственно в системе или импортировано из внешней системы. Каждому заданию может соответствовать произвольное количество ответов. При этом каждое задание может иметь один или несколько правильных ответов. Количество правильных ответов определяется категорией задания.

Генерация тестов из тестовых заданий может осуществляться автоматически, при этом пользователь может задать требуемые объем и трудность теста, а также наложить другие ограничения. Перед началом генерации теста система предлагает задать структуру теста и количество требуемых вариантов.

Структура теста:

1. Название предмета.
2. Общее количество ТЗ в тесте. (N)
3. Количество используемых модулей в тесте. (M)
 - 3.1. Название модуля
 - 3.2. Количество используемых тем в тесте. (L)
 - 3.2.1. Название темы.
 - 3.2.2. Количество ТЗ (K) (На основе предлагаемой категории тестового набора и рассчитывается по энтропийной оценке для такой категории, может корректироваться экспертом-составителем).
 - 3.2.3. Среднее время для одного ТЗ (T)
 - 3.2.4. Средняя трудность ТЗ (S)

База правил для генерации тестов данной темы с указанными K, T.

1. Выбрать из БТЗ (банка ТЗ) к тестовым заданиям и присвоить этому массиву имя ТЕСТ
2. Найти T_n – среднее время для одного ТЗ в ТЕСТЕ
3. $D = 0$
4. Если номер темы > 1 , то $T = T + A$
5. Если $T_n > T$, то переход к 13 иначе к 6
6. Если $T_n = T$, то переход к 16 иначе к 7
7. Удалить из массива ТЕСТ задание с мин временем ($t_{\text{минтест}}$), $\text{ТЕСТ} = \text{новыйТЕСТ}$
8. Копировать в новыйТЕСТ из БТЗ ТЗ, ближайшее по значению времени ($t_{\text{бтз}}$) к $t_{\text{минтест}}$, $t_{\text{бтз}} > t_{\text{минтест}}$, перейти к 11
9. Удалить из массива ТЕСТ задание с max временем ($t_{\text{мактест}}$), $\text{ТЕСТ} = \text{новыйТЕСТ}$

10. Копировать в новыйТЕСТ из БТЗ ТЗ, ближайшее по значению времени ($t_{бтз}$) к $t_{минтест}$, $t_{бтз} < t_{минтест}$, перейти к 11
11. Найти Тновое в новом ТЕСТЕ
12. $D_{новое} = T - T_{новое}$
13. Если $D_{новое} * D \leq 0$ то перейти к 14 иначе перейти к 5
14. Если номер темы = К, то если $D > 0$, $D = D_{новое}$, ТЕСТ= новыйТЕСТ, иначе 15
15. Если $|D_{новое}| - |D| \leq 0$, то 16 иначе с погрешностью $D = D_{новое}$, ТЕСТ= новыйТЕСТ
16. $A = -D$
17. С погрешностью D составлен ТЕСТ

Для генерации теста с заданной трудностью используется подобная база правил, с изменениями в п.8 и 10 (добавляется условие замены тестового задания с одинаковыми значениями времени, если таковых нет, то проверяется следующее по порядку задания). В режиме проведения тестирования все ранее сгенерированные тестовые наборы, предложенные однажды опрашиваемому, учитываются с энтропийной оценкой и используются для прогноза результата.

Блок оценки результатов тестирования

База правил оценки уровня знаний тестируемых

Блок выявления «гадаемых» ТЗ:

1. Если трудность вопроса от 0,14 до 0,26 то включить вопрос в разряд подозрительных и перейти к 2;
2. Если по критерию Фишера распределение «подозрительного» вопроса схоже с совокупностью смоделированных случайных вопросов, то считать вопрос гадаемым иначе считать вопрос обычным.

Блок оценки четкого тестирования

1. Если требуется оценить тест «нетрадиционными» методами, то перейти к блоку оценки нечеткого тестирования иначе оценивать традиционным методом и перейти к 2;
2. Если требуется оценить тест, исключая «гадаемые вопросы», то перейти к блоку выявления «гадаемых» ТЗ и перейти к 3 иначе оценивать традиционным методом и перейти к 4;
3. Если вопрос гадаемый, то исключить вопрос из дальнейшего рассмотрения перейти к 4;
4. Оценить тест традиционным методом по шкале 1 или 0.

5. Трудность вопроса
$$d_i = \frac{w - \sum_{q=1}^w s_{iq}}{w}$$
 (s_{iq} – оценка студента q на i вопрос, w – количество студентов в группе). Данная характеристика в дальнейшем привязывается к предполагаемой категории тестируемого (для экспертизы).

Блок оценки нечеткого тестирования

1. Если вопрос гадаемый, то исключить вопрос из дальнейшего рассмотрения перейти к 2;
2. Если известна априорная информация о степени подготовленности тестируемых (эксперт может выделить категорию «отличников», то перейти

к блоку «методика с отличниками» *иначе* перейти к блоку «универсальная методика» (учитывается категория тестируемого).

Блок «Универсальная методика»

$$t_k = \frac{a_k}{\sum_{k=1}^m a_k}$$

1. Вес правильного ответа t_k ($a_1 \dots a_m$ – число выборов правильных ответов).

$$f_j = - \frac{\prod_{j=1}^n b_j}{b_j \sum_{j=1}^n \frac{\prod_{j=1}^n b_j}{b_j}}$$

2. Вес неправильного ответа f_j *если* $\forall b_j \neq 0$, *то* *если* $b_j = 0$, *то* $f_j = 0$

$$f_j = - \frac{\prod_{j=1, j \neq k}^m b_j}{b_j \sum_{j=1, j \neq k}^m \frac{\prod_{j=1, j \neq k}^m b_j}{b_j}}$$

если $b_k = 0$, $k \neq j$ *то* ($b_1 \dots b_n$ – число выборов неправильных ответов).

$$r_i = 1 + \frac{\sum_{q=1}^w s_{iq}}{w}$$

3. Вес вопроса (s_{iq} – оценка студента q на i вопрос, w – количество студентов в группе). Используется для генерации «адаптивного» тестового набора в схеме неизвестной категории тестируемого с привлечением оценки объема набора энтропийной оценкой.

Блок «Методика с отличниками»

1. Ввод данных («отличников» по мнению экспертов);

2. Если студент, введенный в категорию «отличников», набрал менее 50% от максимума, то исключить его из категории «отличников» и выдать соответствующее сообщение эксперту;

3. Решается задача ЛП

$$\sum_{i=1}^l r_i = \sum_{i=1}^l \sum_{j=1}^5 a_{ij} \cdot x_j \rightarrow \max$$

при ограничениях

$$x_j \in [-1; 0], \text{ если } x_j \text{ – неправильный ответ}$$

$$x_j \in [0; 1], \text{ если } x_j \text{ – правильный ответ.}$$

($a_{i1}, \dots, a_{i5}$) – это множество выборов ответов "отличника" i на вопрос k ;

4. *Если* эталон единственный, *то* перейти к 7 *иначе* перейти к 5;

5. Перейти к блоку «универсальной» методики;

6. Выбрать из множества N получившихся по методики с отличниками эталонов $X_n = (x_{1n}, \dots, x_{5n})$ «ближайший» к эталону, получившемуся по

универсальной методике $X_y = (x_{1y}, \dots, x_{5y})$, такой, что $\sum_{i=1}^5 (x_{li} - x_{yi})^2$ была наименьшей (в данном случае если ветвление в этом месте не «ломает»

энтропийную оценку объема тестового набора, а БД содержит предыдущие итоги, подтверждающие категорию тестируемого, то с учетом коррекции веса (7.) переход к расчету оценки для тестируемого сочетающей пройденное начало и обнаруживаемый остаток);

$$7. \text{ Вес вопроса } r_k = \frac{\sum_{i=1}^l r_i^k}{l}.$$

Анализатор предназначен для обеспечения обратной связи в системе. Внутренняя обратная связь – это информация, которая поступает от студента к тестирующей программе в ответ на его действия при выполнении тестов. Она предназначена для самокоррекции тестирующей программой своей базы данных и базы правил. Информация внешней обратной связи в рассматриваемой системе поступает к педагогу и используется им для коррекции содержательной стороны тестов или самого курса лекций, а также деятельности студента. Схема анализатора тестовых заданий отображена на рисунке 3.5 и выглядит следующим образом:



Рисунок 3.5 – Схема анализатора тестовых заданий

База правил анализатора

Если тестирование четкое, то выполнить блок 1

Если тестирование нечеткое, то выполнить блок 2.

Блок 1.

Если выявлены гадаемые вопросы, то выдать соответствующую информацию эксперту. Если имеется априорная информация о тестируемых, то выдать эксперту информацию о степени валидности (вес вопроса по методике с «отличниками») и рекомендовать вопросы, имеющие коэффициент валидности менее 0,3, исключить из банка ТЗ или откорректировать. Если априорная трудность в банке ТЗ не равна апостериорной с точностью до 0,01, то откорректировать значение в банке ТЗ (и всех параметрических форм ТЗ). Если апостериорная трудность в банке ТЗ не равна апостериорной трудности для настоящего этапа тестирования с точностью до 0,01, то откорректировать значение в банке ТЗ (и всех параметрических форм ТЗ) с учетом накопившейся статистики. Если априорная оценка необходимого для решения задания время в банке ТЗ не равна апостериорной с точностью до 1 секунды, то откорректировать значение в банке ТЗ (и всех параметрических форм ТЗ). Если апостериорная оценка необходимого для решения задания время в банке ТЗ не равна апостериорной для настоящего этапа тестирования с точностью до

1 секунды, то откорректировать значение в банке ТЗ (и всех параметрических форм ТЗ).

1. Если коэффициент надежности теста удовлетворяет эксперта, то запомнить структуру теста.

Под структурой теста понимается: количество заданий (оптимальный вариант – параметрических) определенного уровня трудности и временной характеристики для каждого модуля, предмета.

Блок 2.

Если выявлены гадаемые вопросы, то выдать соответствующую информацию эксперту и исключить их из дальнейшего рассмотрения. Если имеется априорная информация о тестируемых, то выдать эксперту информацию о степени валидности (вес вопроса по методике с «отличниками») иначе выдать эксперту информацию о трудности (по универсальной методике) и рекомендовать вопросы, имеющие коэффициент валидности менее 0,3, исключить из банка ТЗ или откорректировать. Если апостериорная трудность в банке ТЗ не равна апостериорной трудности для настоящего этапа тестирования с точностью до 0,01, то откорректировать значение в банке ТЗ (и всех параметрических форм ТЗ) с учетом накопившейся статистики. Если апостериорная оценка необходимого для решения задания время в банке ТЗ не равна апостериорной для настоящего этапа тестирования с точностью до 1 секунды, то откорректировать значение в банке ТЗ (и всех параметрических форм ТЗ).

2. Если коэффициент надежности теста удовлетворяет эксперта, то запомнить структуру теста.

Под структурой теста понимается: количество заданий (оптимальный вариант – параметрических) определенного уровня трудности и временной характеристики для каждого модуля, предмета.

Контроль знаний является неотъемлемой частью педагогического процесса, особенно если мы хотим рассматривать этот процесс как технологический, т. е. обеспечивающий при определенных условиях гарантированное достижение поставленной цели. Хотя существуют отдельные программные средства, ориентированные на компьютеризацию контроля, тем не менее, эти средства не во всех случаях являются адекватными для решения соответствующих задач вступительного и межсессионного тестирования, решениями которых занимается представляемая в настоящей работе АСУТ (подсистема ИСОРО).

АСУТ можно разбить на следующие части: автоматическая и автоматизированная генерация тестовых заданий по описанию системы знаний, генерация теста из тестовых заданий, проведение компьютеризированного опроса по тесту и обработка результатов тестирования. Для эффективного использования педагогического тестирования на практике необходимо иметь возможность с минимальными затратами времени генерировать достаточное количество вариантов тестовых заданий. Генерация тестов из тестовых заданий может осуществляться

автоматически, при этом пользователь может задать требуемые объем и сложность теста, а также наложить другие ограничения.

Компьютеризованный опрос может осуществляться по тестам, сгенерированным непосредственно перед тестированием, что реально позволяет обеспечить каждого тестируемого собственным вариантом (в случае «компьютерного теста»). Обработка результатов педагогического тестирования является важной частью контроля знаний и рассматривается как элемент итеративного процесса разработки тестов, т.е. использования результатов тестирования для модификации теста с целью улучшения его апостериорных оценок. АСУТ была использована для генерации тестов по различным учебным дисциплинам для вступительных и межсессионных экзаменов в Северо-Казахстанском университете имени М. Козыбаева и продемонстрировала свою высокую эффективность [67, стр. 785].

Подробнее процесс тестирования выглядит следующим образом. Из банка тестового материала ЭС отбирает задания, удовлетворяющие определенной базе правил и оценивает уровень пригодности и уровень трудности. Задания, которые не удовлетворяют базе правил, в банк ТЗ не вносятся. Далее энтропийная оценка формирует тест с необходимым набором предметов, заданным количеством вопросов по каждой теме в предмете и заданным уровнем трудности вопросов и временным характеристикам. После обработки результатов тестирования анализатор выдает рекомендации по оптимизации теста. Блок базы знаний и данных модели состоит из банка тестовых материалов, банка тестовых заданий. Банк тестовых заданий может в любое время пополняться новыми заданиями. Взаимодействие блоков отражено на рисунке 3.6 линиями.

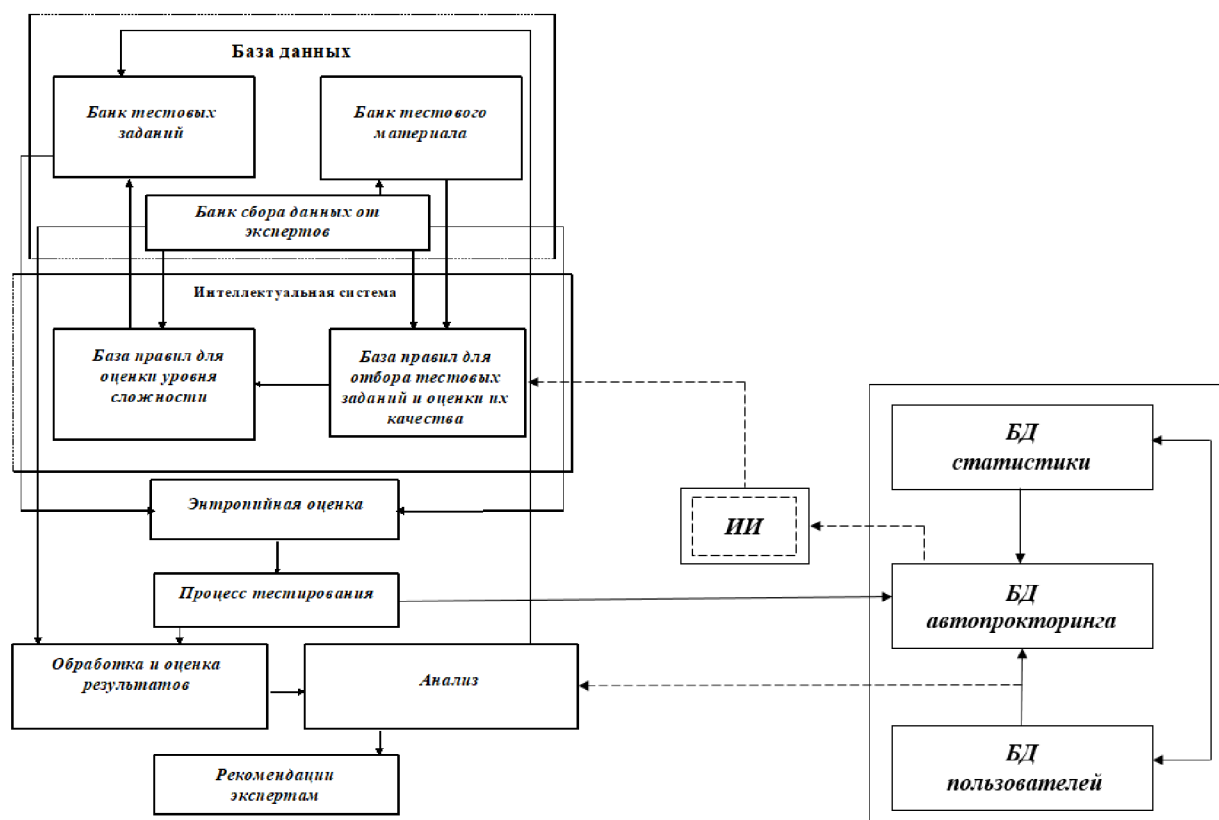


Рисунок 3.6 – Структура автоматизированной системы управления тестированием

Блок экспертного опроса состоит из базы правил (БП) для отбора тестовых заданий в банк; базы правил (БП) для оценки сложности тестовых заданий и ответов и блока сбора данных от экспертов (куда могут входить данные об априорных оценках отдельных качеств заданий, данные о тестируемых, данные о требуемых параметрах теста). Блок экспертного опроса (БЭО) предназначен для генерации и настройки блока оценки результатов тестирования, блока принятия решений (БПР) операционной части самого БЭО. Энтропийная оценка тестов непосредственно занимается формированием теста (учитывая поступающую из БЭО информацию о необходимом наборе предметов, количестве вопросов по каждой теме в предмете и общем уровне трудности теста) [67, стр. 809].

В блок принятия решений системы входит анализатор результатов тестирования, который выдает рекомендации по оптимизации теста (например, касающиеся определения качеств ТЗ, корректирует атрибуты заданий в банке ТЗ, запоминает структуру «удачно сформированного» по мнению экспертов тестов) и блок оценки результатов тестирования, который в зависимости от целей тестирования (межсессионное или вступительное) принимает решение о качествах ТЗ, о присваивании ТЗ оптимальных весов, разбалловок, или об исключении ТЗ из рассмотрения. Применение системы управления тестированием позволяет проводить экзамены на более высоком

технологическом и техническом уровне, повышает оценочную способность теста [37, стр. 369].

3.4 Выводы по третьей главе

В связи с использованием тестовой системы контроля знаний приобрели особую актуальность разработка контролирующих систем, основанных на тестировании, разработка методики составления тестов. Под оптимизированной обработкой данных тестирования понимается организация ввода результатов тестирования в базу хранения, сопряжение этой базы с базой тестовых материалов, обработка результатов, тестируемых и получение отчетных данных с минимальными затратами всех видов ресурсов.

Исследование заключалось в адаптировании основных критериев оценки тестового материала и тестового задания, разработке обобщенных формализованных подходов к оцениванию разнотипных вопросов четких и нечетких тестов, статистическом анализе получаемых результатов. По результатам исследования была создана автоматизированная система управления тестированием, позволяющая получать априорную информацию о тестовом задании, создавать тесты, проводить тестирование, автоматически получать его результаты. В программу заложены принципы, описанные в работе.

4 ПРОЕКТИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ОЦЕНКИ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ И ПРОВЕДЕНИЕ АПРОБАЦИИ

4.1 Принцип управления организацией системы оценки результатов обучения

«Используя кейс-технология для создания функциональной модели (IDEF0), реализующей процессно-ориентированное и структурное моделирование, предлагается контекстная диаграмма, которая является вершиной древовидной структуры диаграмм и представляет собой наиболее общее описание системы (рис.4.1).

Данная технология позволяет начать исследование с общего обзора системы, разлагая ее на части, тем самым сформировав иерархическую структуру системы с большим количеством уровней, а также продемонстрировать, как каждый процесс преобразует входные данные в выходные, и выявить взаимосвязи между этими процессами.

Для создания данной диаграммы проанализированы внешние объекты, влияющие на процесс управления системой оценки результатов образования: обучающийся (студент), преподаватель» [43, стр. 73].

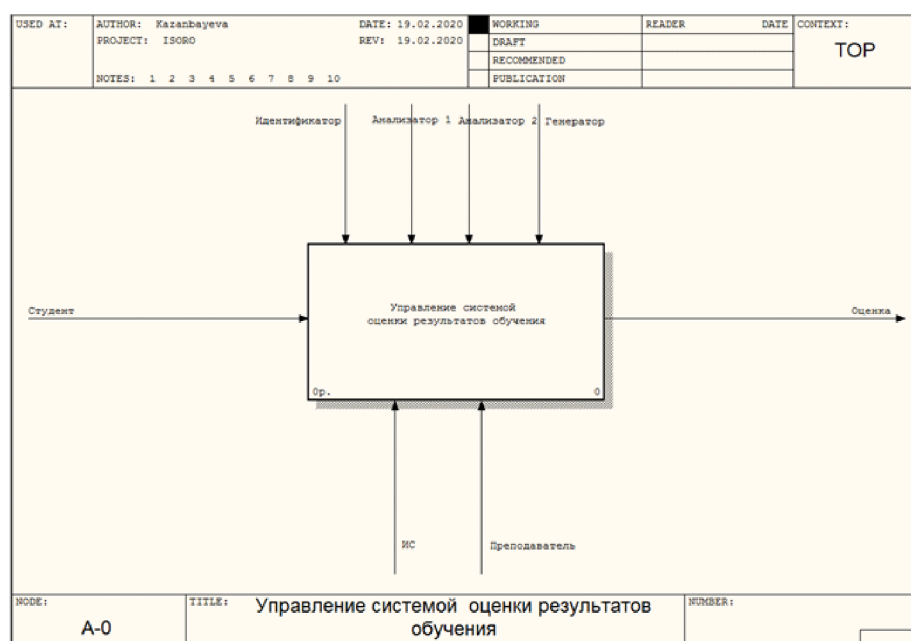


Рисунок 4.1 – Информационная поддержка управления СОРО

Входной информацией являются данные, которые предварительно внесены при оформлении и зачислении обучающегося, а выходной – рейтинг, успеваемость, статус обучающегося.

«Проектирование и организация системы соответствуют требованиям основных нормативных документов: правил приема в университет, закона об образовании, государственного общеобразовательного стандарт, устава университета, учебного плана и рабочих программ учебных дисциплин»[43, стр.74].

В целом система разбивается на фрагменты, показанные на рисунке 4.2.

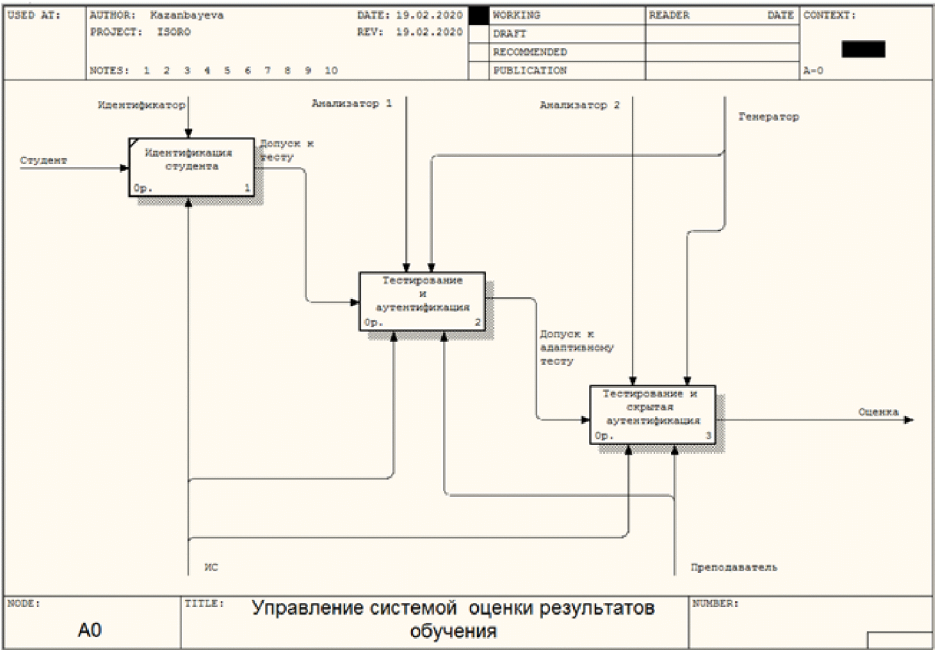


Рисунок 4.2 – Декомпозиция процесса «Управление СОРО»

Кроме того, схема работы, показанная на рис.4.3, отражает тот факт, что работа пользователя обеспечивается только после определения его статуса.

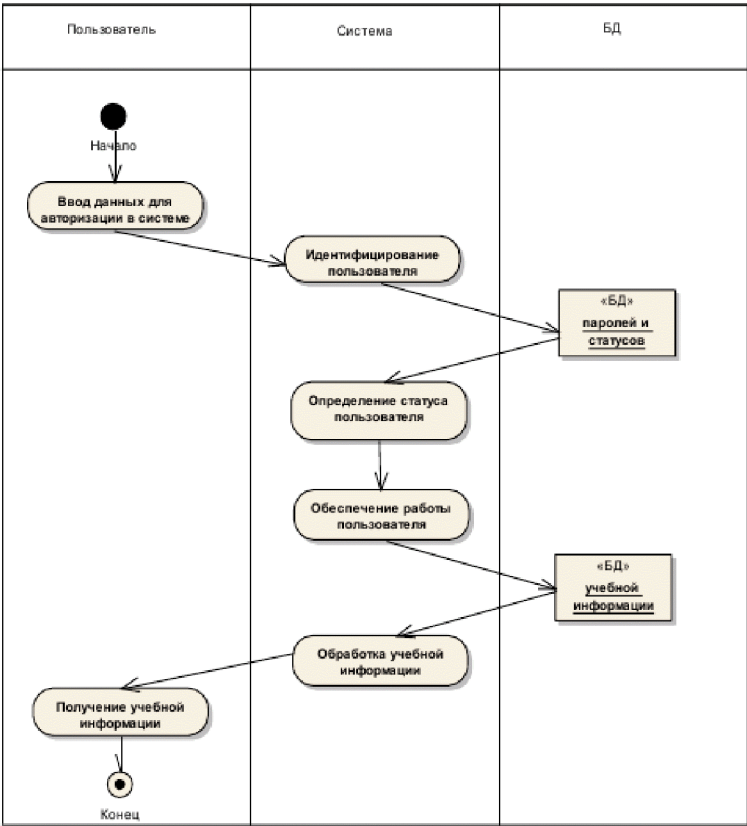


Рисунок 4.3 – Диаграмма указывающая статус пользователя

«Возможность работы с системой пользователю предоставляется только после определения его статуса. При доступе пользователей к системе требуются данные в виде имен пользователей и паролей. Входные данные проверяются в базе данных, и если они верны, то система предоставляет пользователям права, доступ к данным и действия, соответствующие статусу. Если идентификационные данные введены неправильно, пользователю будет отказано в доступе к системе. Во время работы пользователей обеспечивается доступ к базе данных системы, где хранятся все данные об учебной информации» [43, стр. 74].

Процесс декомпозиции тестирования и аутентификации допуска к адаптивному тестированию двумя процессами (рис. 4.4):

- 1) Тестирование 1 (производится методом генератора).
- 2) Аутентификация.

Доступ к первому процессу определяется обычной идентификацией (логин и пароль выдаются администратором).

Далее вам будет предложен ряд тестовых заданий (открытого типа) для прохождения и определения знаний.

Второй процесс включает в себя аутентификацию путем ввода индивидуального идентификационного номера и получения кода сообщением (sms) на связанный номер телефона.

После подтверждения данных и ввода кода открывается доступ к адаптивному тесту, сгенерированному ранее пройденной проверкой знаний.

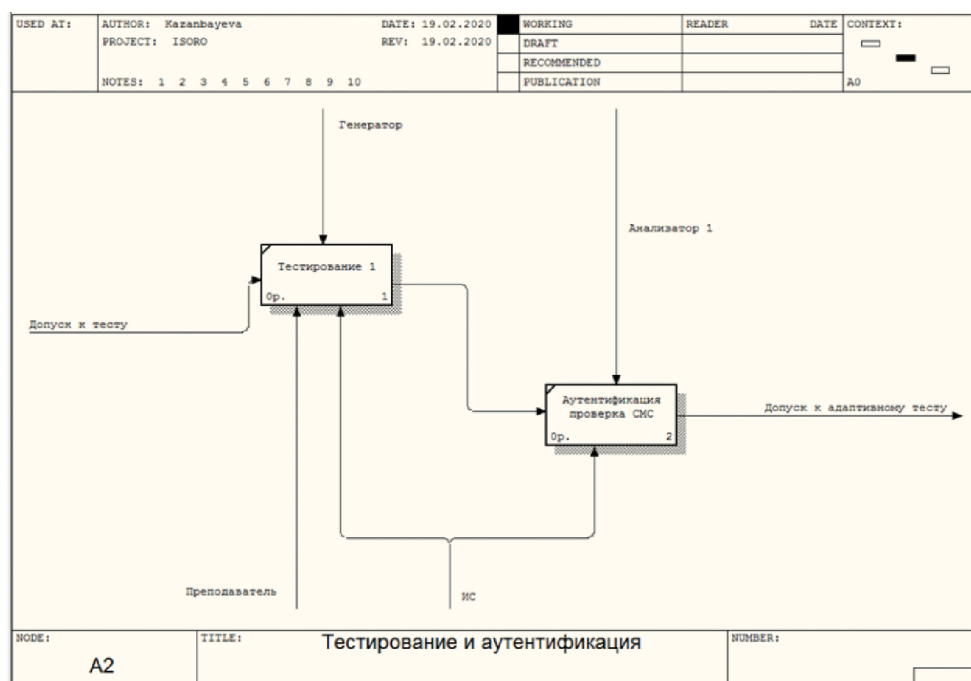


Рисунок 4.4 – Декомпозиция процесса «Тестирование и аутентификация»

Таким образом, процесс выхода, предоставленный на рисунке 4.4 допуска к адаптивному тестированию позволит студентам пройти комплексную проверку знаний, включающую в себя тестовые задания закрытого типа,

соответствие, определение последовательности, а также вопрос, позволяющий им ответить устно с использованием модуля СОРО (рис.4.5).

В конце процедуры проверки выводится итоговая оценка.

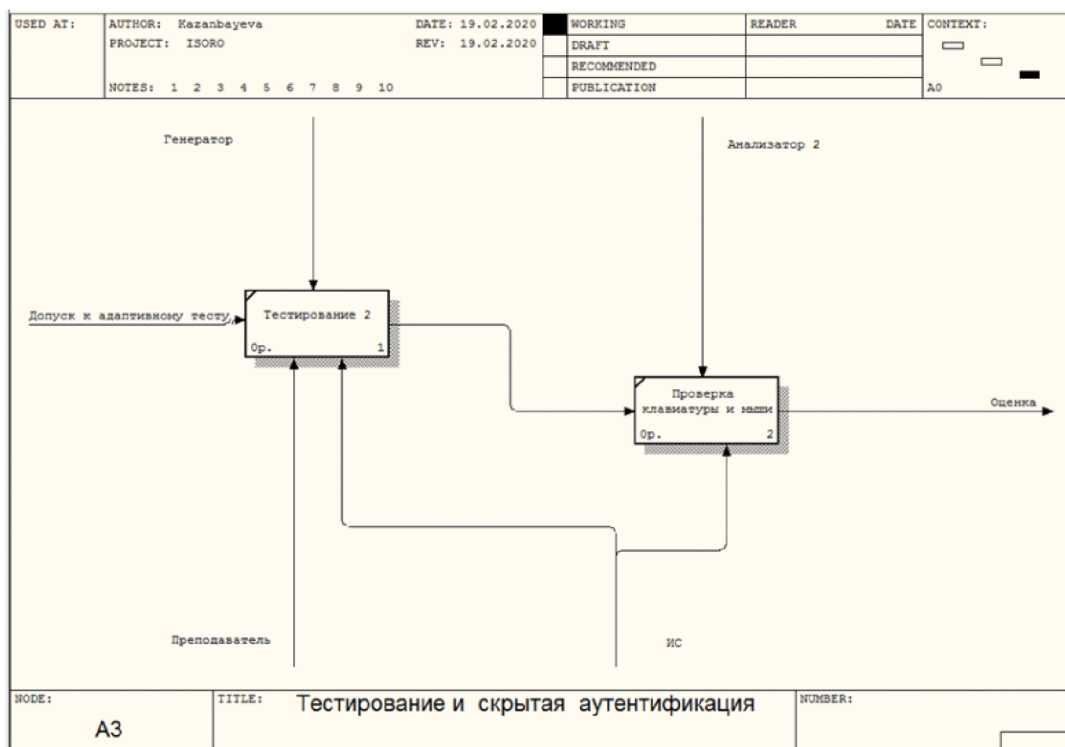


Рисунок 4.5 – Декомпозиция процесса «Тестирование и скрытая аутентификация»

В данном разделе разработаны модели, отражающие основные особенности системы управления организации системы оценки результатов обучения при подготовке обучающихся.

4.2 Применение устного/письменного экзамена в интеллектуальной системе

«В условиях рыночной экономики, как уже было изложено ранее, меняются потребности и уровень требований потребителей образовательных продуктов, что диктует необходимость постоянного совершенствования образовательного процесса как для успешной работы на рынке образовательных услуг, так и для потребностей рынка труда. Одним из путей решения этой проблемы является внедрение передовых коммуникационных информационно-образовательных технологий, позволяющих образовательным учреждениям идти в ногу со временем и предоставлять студентам новые возможности и повышенную гибкость в процессе получения качественного образования» [43, стр. 88].

Модуль интеллектуальной системы оценки результатов обучения (рис. 4.6) позволяет на определенном уровне прохождения тестирования, а именно

в разделе включающие в себя вопрос для устного/письменного ответа провести контроль устного/письменного экзамена с помощью наличия встроенного или стационарного микрофона и самое главное наличие подключения к сети Интернет.

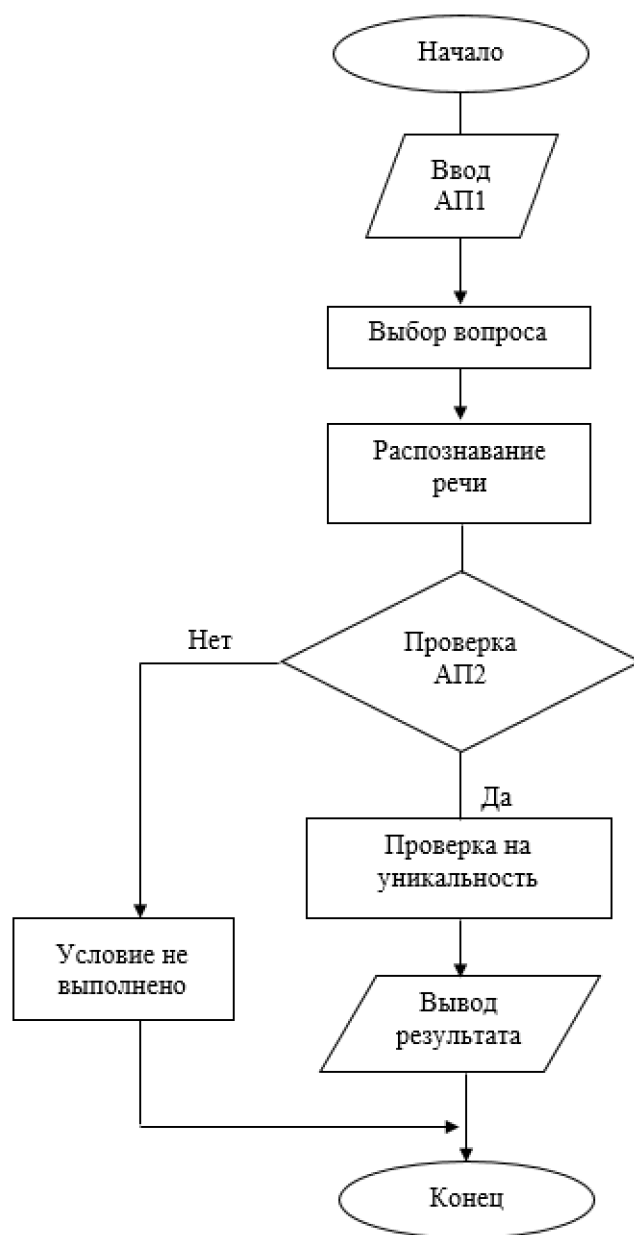


Рисунок 4.6. – Блок алгоритма модуля интеллектуальной системы оценки результатов обучения

Прохождение данного контроля включает вступительное тестирование, при котором на каждой ступени «сложности учебного материала необходимо ответить на одинаковое количество заданных вопросов, различной сложности. По результатам прохождения система самостоятельно или с помощью эксперта определяет следующий уровень сложности материала в данном разделе и генерирует его. Например, если обучающийся правильно ответил менее чем на треть вопросов, то он отправляется на следующий уровень

сложности; если процент правильных ответов составил от 1/3 до 2/3 и от 2/3 до 1, то обучающийся отправляется на аутентификацию и на следующий уровень сложности»[43, стр. 90]. То есть переход к вопросу, требующему устного / письменного экзамена. Однако может возникнуть ситуация, когда испытуемый не прошел проверку и будет автоматически удален из системы, в Приложение В, представлены скриншоты сообщений о возникших проблемах.

Модуль интеллектуальной системы оценки результатов обучения предназначен для автоматизации и реализации оценки результатов обучения на заданные вопросы с помощью микрофона, то есть устного ответа.

Модуль системы реализован средствами веб технологий, который позволяет производить устный контроль, а также автоматически распознавать речь, преобразовывая его в текст проверить его на уникальность, тем самым выдавая на экране процент уникальности.

На рисунке 4.7 представлен интерфейс на начальной входной аутентификации тестируемого.

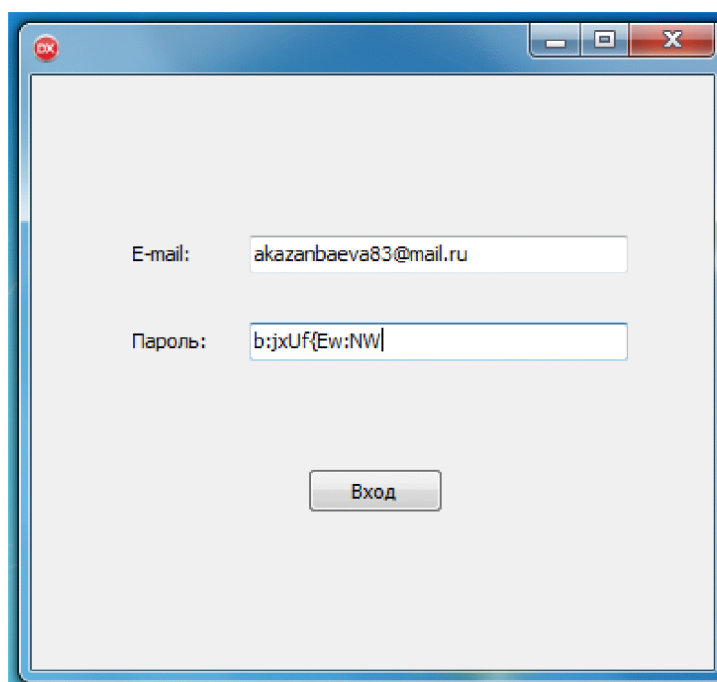


Рисунок 4.7 – Идентификация пользователя

Интерфейс позволяет вводить вопросы, предназначенные для контроля, в настройках с помощью кнопки добавить, имеется возможность удалить с помощью кнопки очистить.

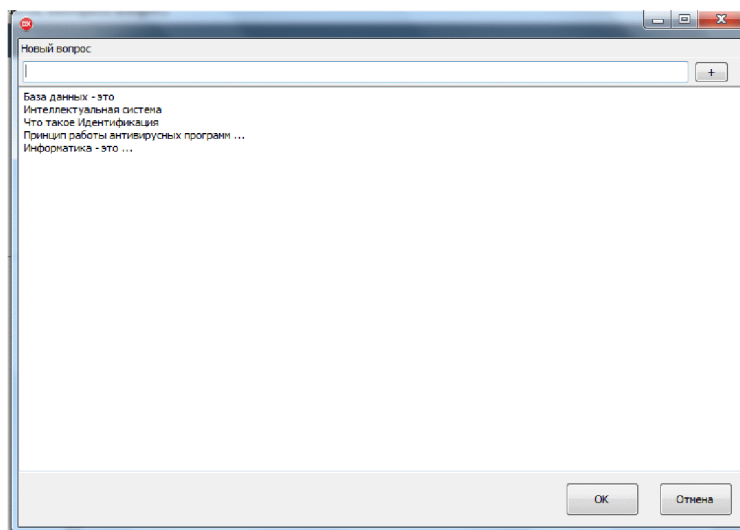


Рисунок 4.8 – Форма ввода заданий/вопросов для контроля

На рисунке 4.9 продемонстрирован интерфейс тестируемого, который, в свою очередь, сопровождается с выбором вопроса и возможностью ответить.

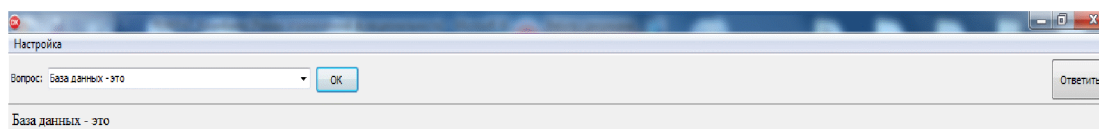


Рисунок 4.9 – Форма выбора вопроса

При нажатии кнопки ответить, открывается следующее окно, для ввода устного ответа с помощью микрофона, распознавание речи (Рис. 4.10).

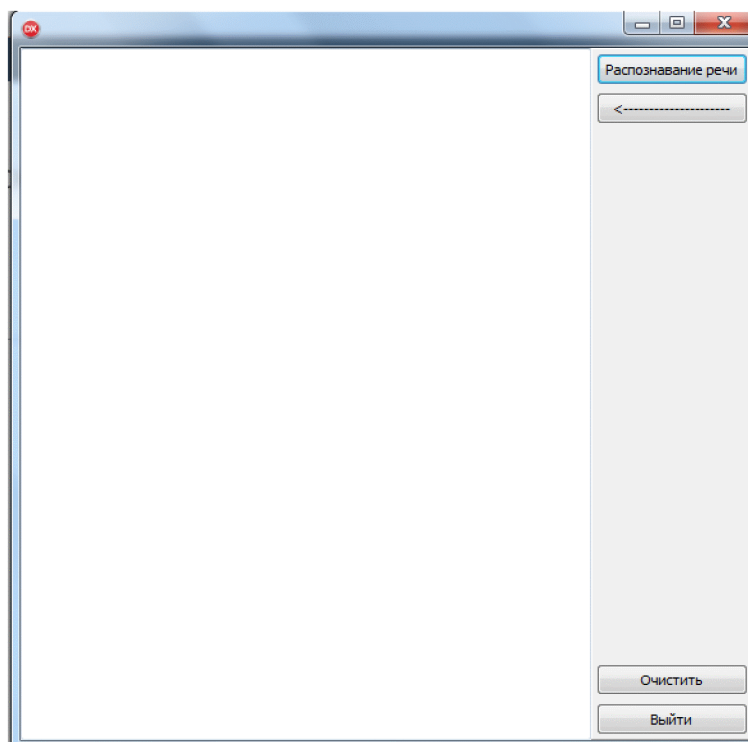


Рисунок 4.10 – Интерфейс распознавание речи

При переходе на кнопку «Распознавание речи» открывается окно, записывающее голосовой ответ, то есть ввод устного развернутого варианта ответа (Рис. 4.11).

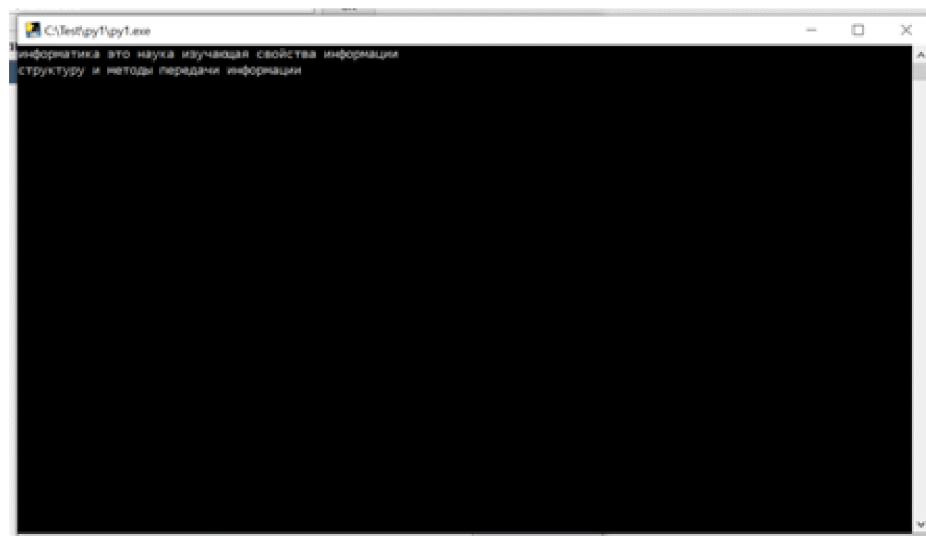


Рисунок 4.11 – Запись звука

С помощью данного модуля устный ответ будет автоматически преобразован в текстовый формат и перенаправлен на проверку плагиат, то есть на уникальность ответа тестируемого (рисунок 4.12).

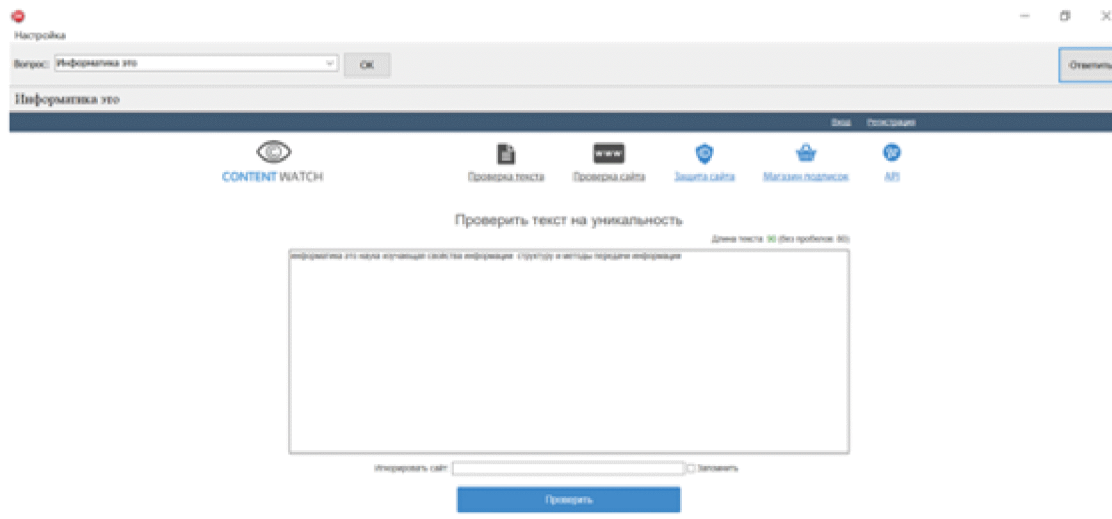


Рисунок 4.12 – Интерфейс подтверждения на проверку текста на уникальность

Б) Отсеву подвергаются не только субъекты обучения, но и документы, использующиеся в процессах, если те или другие не обеспечили необходимого уровня в рамках соответствующей подсистемы.

В) Пул документов обладает ровно такой избыточностью, что, так или иначе применяется (неактуальные документы применяются с вероятностью 0, т.е. фактически отсутствуют).

Итак, всего документов $N(t)$, одновременно или во временной серии (сессии) используется часть документов объема $n_{ij}(t)$.

Варианты обозначений:

- первый индекс маркирует субъекта, второй – серию, в которой соответствующий объем документов применен;

- первый маркирует временной срез, а второй – субъекта, к которому на этом этапе (срезе) были применены соответствующие объемы документов.

Обоснованное предположение об эргодичности дает возможность отождествлять усреднение по времени и по ансамблю. Более того, стационарная эргодичность позволяет отождествлять, а) усредненные характеристики, определенные на основании наблюдения над многими подобными системами в один и тот же момент времени, и б) средние по времени, определенные на основании наблюдения над одной из этих систем для достаточно большого числа последующих моментов времени. Соответственно, для систем, ведущих себя стационарно-эргодически, не имеет значения трактовка индексов n_{ij} , так как средние по ансамблю наблюдений или по времени совпадут.

Определяя интенсивности $p_{ij}(t) = \frac{n_{ij}(t)}{N(t)}$, $\sum_{ij}^N p_{ij} = 1$, формулируем *идеальное поведение системы как эволюцию энтропии S применения документов $S \rightarrow \text{extr}$* , при соблюдении всех правил функционирования подсистемы.

Системы контроля сопровождаются формированием показателей, отражающих содержательную сторону контроля, выражаемых в количественной форме. То есть, всегда существует набор показателей/индикаторов в объеме R , заданных в виде функций от каких-то качественных характеристик протекающих процессов, $f_r(x_{ij})$, $r=\overline{1, R}$, где x_{ij} - содержательные характеристики взаимодействия субъекта i с потоком документов j описываемой подсистемы. Характеристики x_{ij} в сводной форме важны системе для целей принятия решений по совершенствованию подсистемы. Обычно известны (или назначены) средние характеристики E_r . Тогда формулировка оптимизационной задачи может иметь вид:

$S \rightarrow \max$

$$\sum_{ij}^N p_{ij} f_r(x_{ij}) = E_r \text{ для любого } r, r=\overline{1, R}$$

$$\sum_{ij}^N p_{ij} = 1 \quad (16)$$

Решение задачи (в смысле оценки энтропии документооборота в системе) определит план интенсивностей применения документов, опирающийся на достижение именно запланированных показателей.

Типовое решение описанной выше задачи предполагает введение функции Лагранжа с неопределенными множителями (учтем все R требований). Функция Лагранжа с неопределенными множителями Лагранжа α, β_r :

$$L(p_{ij}, \alpha, \beta_r) = S - \alpha(\sum_{ij}^N p_{ij} - 1) - \sum_r \beta_r (\sum_{ij}^N p_{ij} f_r(x_{ij}) - E_r) \quad (17)$$

Распределение p_{ij} по критерию экстремума энтропии есть решение набора дифференциальных уравнений:

$$\frac{\partial L}{\partial p_{ij}} = 0 \quad (18)$$

Энтропия S может быть явно выписана в смысле Шеннона, Тсаллиса или Реньи [69].

Задаваясь энтропией Шеннона $S = -\sum_{ij}^N p_{ij} \log p_{ij}$, и единственным индикатором $r=1$, получаем для каждого i и j :

$$\frac{\partial L}{\partial p_{ij}} = -\log p_{ij} - 1 - \beta f(x_{ij}) - \alpha \quad (19)$$

то есть

$$\begin{aligned} -\log p_{ij} - 1 - \beta f(x_{ij}) - \alpha &= 0 \\ p_{ij} &= e^{-1-\alpha-\beta f(x_{ij})} \end{aligned} \quad (20)$$

Откуда, после суммирования и с учетом $\sum_{ij}^N p_{ij} = 1$, имеем:

$$\sum_{ij}^N e^{-1-\alpha-\beta f(x_{ij})} = e^{-1-\alpha} \sum_{ij}^N e^{-\beta f(x_{ij})} = 1 \quad (21)$$

или

$$e^{\alpha} = \sum_{ij}^N e^{-\beta f(x_{ij})-1} \quad (22)$$

Сложность исследования этой связи множителей Лагранжа определяется нетривиальностью вида зависимости $f(\cdot)$, с помощью которой введен единственный индикатор E . Получив пару рассчитанных, например, приближенными методами, множителей Лагранжа, можно однозначно установить все этапные объемы документов в стабильно работающей системе контроля в каждый момент времени (или раскладку объемов на субъекты контроля), в том числе – предсказать, как показатель E повлияет на статус какого-либо субъекта (документа) в подсистеме.

Как правило, тестология и экзамен «по билетам» опираются на равновероятное распределение экзаменационных вопросов/билетов. Реально же существуют тестовые задания/билеты «практически невостребованные», статистика по ним отсутствует, и, соответственно, оценить их качество статистически невозможно со всеми вытекающими рисками ошибок отсева и контроля.

Применяя формулу энтропии Тсаллиса $S_T = \frac{1}{q-1} (1 - \sum_{ij}^N p_{ij}^q)$ вместо уже использованной прежде энтропии Шеннона, не прибегая к предположению о единственном индикаторе, имеем:

$$\frac{\partial L_R}{\partial p_{ij}} = \frac{q}{1-q} p_{ij}^{q-1} - \alpha - \sum_r \beta_r f_r(x_{ij}) = 0 \quad (23)$$

Получаем параметрическое уравнение (смотреть трактовку в случае энтропии Шеннона):

$$\sum_{ij}^N \left(\frac{1-q}{q} (\alpha + \sum_r \beta_r f_r(x_{ij})) \right)^{\frac{1}{q-1}} = 1 \quad (24)$$

Применяя энтропию Реньи $S_R = \frac{1}{1-q} \log \sum_{ij}^N p_{ij}^q$, при $r=1$, имеем:

$$\frac{\partial L_R}{\partial p_{ij}} = \frac{1}{1-q} \frac{p_{ij}^{q-1}}{\sum_{ij}^N p_{ij}^q} - \alpha - \beta f(x_{ij}) = 0 \quad (25)$$

Умножая на p_{ij} и просуммировав по i и j , имеем:

$$\sum_{ij}^N \left(\frac{1}{1-q} \frac{p_{ij} p_{ij}^{q-1}}{\sum_{ij}^N p_{ij}^q} - \alpha p_{ij} - \beta p_{ij} f(x_{ij}) \right) = 0 \quad (26)$$

$$\frac{1}{1-q} - \alpha \sum_{ij}^N p_{ij} - \beta \sum_{ij}^N p_{ij} f(x_{ij}) = 0 \quad (27)$$

А так как $\sum_{ij}^N p_{ij} = 1$, то, обозначив $E = \sum_{ij}^N p_{ij} f(x_{ij})$, получим окончательно формулу связи (q, E, α, β):

$$\frac{1}{1-q} = \alpha + \beta E \quad (28)$$

Содержательная трактовка:

E – индикатор;

q – параметр Реньи;

α – множитель Лагранжа для учета интенсивностей;

β – множитель Лагранжа для учета индикатора;

формула связи (q, E, α, β) буквально означает: единственный индикатор качества «в среднем» может быть успешно заменен на параметр энтропии Реньи.

Более того, допуская подбор α, β при $q=0$, энтропию можно признать «структурной» и фактически выразить формулой $S = \log N$. Это была бы формула Хартли, если бы мы приняли равновероятное распределение p_{ij} (как в случае формулы Шеннона и, как упоминалось выше, вследствие обычной практики при экзаменационном или тестовом контроле). Однако, здесь это лишь совпадение вследствие «удачного» разбиения структуры документов подсистемы и равную интенсивность вовсе не полагает.

Кроме того:

- декларируя энтропию Реньи при $q = 0$;
- «измельчая объемы» документов (пропорционально раздувая их число) подсистемы под масштабный параметр ε , такой, что весь пул можно по-прежнему ассоциировать с индексами, указывающими на связь поддокументов разного объема с вновь расщепленными процедурами оценивания;
- допуская при этом существование числового предела, указывающего на обобщенную фрактальную размерность Реньи

$$d_0 = \lim_{\varepsilon \rightarrow 0} \frac{s}{\ln\left(\frac{1}{\varepsilon}\right)} = \lim_{\varepsilon \rightarrow 0} \frac{\ln\left(\sum_{i=1}^{N(\varepsilon)} p_{ij}^0\right)}{\ln\left(\frac{1}{\varepsilon}\right)}, \quad (29)$$

получаем размерность Хаусдорфа, характеризующую фрактальность структуры подсистемы документов оценивания [70]. Такая фрактальность пула документов однозначно полезна для эффективного масштабирования материалов для оценивания с ростом числа субъектов или количества серий.

В ходе исследования предложено для любого учреждения образования Республики Казахстан (статистически не подтверждено, выведено из личного опыта авторов), модельные допущения конкретизируются следующим образом:

А) Как правило, реализованы все три подсистемы контроля: входного (отсева), текущего (регулярного), выходного (итогового). Отметим: «выгрузка» данных протокола процессов для обработки, как правило, представляет отдельную проблему.

Б) Как правило, «отсев балластных» документов либо не производится, либо не связан причинно-следственным образом с обеспечением декларируемого качества функционирования подсистем.

В) Как правило, пул документов пассивно достигает некоторого уровня избыточности способом «само получилось».

Рассмотрим подсистему Внешней Оценки Учебных Достижений (ВОУД):

- 4 набора заданий по 25 вопросов;
- каждый вопрос соотнесен с группами от 5 до 8 потенциальных ответов;
- каждая группа содержит от 1 до 3 ответов, маркированных как «правильные»;
- результатом ВОУД для аттестуемых будет начисление баллов за выбор маркированных вариантов в диапазоне от 0 до 200.

- единственным критерием $E = \sum_{ij}^N p_{ij} f(x_{ij})$ выступает условие: доля Z испытуемых должна набрать баллы не менее W , что может трактоваться и как «средний» по аттестуемым набор баллов не ниже порогового: $E \geq W$.

Будем считать, что с ростом объема испытуемых значения Z и W не меняются (согласуется с применительной практикой ВОУД), и функция $f(x_{ij})$ такова:

- аргумент x_{ij} играет роль события «встречи» партии аттестуемых с выборкой тестовых заданий;

- содержательно $f(x_{ij})$ представляет на всем тестовом пуле N итог для i -ой партии аттестуемых (партия может состоять из единственного аттестуемого) на j -ой выборке тестовых документов, совершаемой по правилам внешней оценки учебных достижений;

- от каждого x_{ij} функция $f(x_{ij})$ неотрицательна.

Тогда $f(x_{ij})$ допустимо принять константной, то есть равной E ($E \geq W$), что следует из $E = \sum_{ij}^N p_{ij} f(x_{ij})$ с учетом нормирующего условия $\sum_{ij}^N p_{ij} = 1$.

Например, возможны такие интерпретации «практически невероятных» процессов аттестаций ВОУД:

И) *ситуация формируется неестественным путем, за счет искусственного создания партии аттестуемых.* В частности, все неспособные (нежелательные) аттестоваться «получают» больничный лист вместо теста ВОУД. Имеем ситуацию равных $f(x_{ij})$ (если способные сдать ВОУД отвечают только до порогового уровня E). Это соответствует ситуации равных p_{ij} для функции энтропии Шеннона-Хартли $S = \log N$ (трактуемой как максимально неопределенная), но приводит к гарантированному $E \geq W$. Иными словами, «равновероятный» тестологический подход дает, тем не менее, гарантированный результат.

П) *«подставная» аттестация ВОУД.* Существует только одно значение p_{ij} , отличное от нуля (и равное единице): процесс имеет абсолютно регулярную основу, так как даже *весь пул* тестов, предъявляемый одной-единственной группе с характеристикой x_{ij} «экспертов по тестам» (несанкционированно помогающим аттестуемым) всегда дает индикатор, равным экспертному $E \geq W$.

Однако, исключение ситуаций И) и П) (т.е. возможности «подчищать справками» списки аттестуемых и/или фактически аттестоваться «экспертам») приведет к изменению в значениях $f(x_{ij})$ и в оценке энтропии. Действительно, какое-либо распределение p_{ij} , отличающееся от равномерного, провоцирует не константную $f(\cdot)$, поскольку ненулевое E дает коррелированность p и $f(\cdot)$ в трактовке p и $f(\cdot)$ как случайных величин. Содержательные требования ВОУД – это требование попадания двумерной случайной величины $E(\cdot, \cdot) = \sum_{ij}^N (p_{ij} f(x_{ij}) \stackrel{\text{def}}{=} E_{ij})$ в эллипс рассеивания [71], параметризованный величинами Z и W .

Разобранные выше случаи И) и П) процессов аттестаций ВОУД в предложенной трактовке – это искусственное снижение дисперсии разброса результатов со сдвигом среднего в сторону максимума. Следовательно, И) и П) можно распознать, сравнивая значения энтропии, рассчитанной и по результатам сессий, и по результатам ВОУД. «Искусственность» ситуации И) эквивалентна «срежисированности» ситуации П) (и обе отличны от сессионной ситуации, когда, так или иначе, но аттестуются *все и самостоятельно*), на роль энтропии в первую очередь должна претендовать

энтропия Реньи с q-параметром, пересчитанным из уровня показателя E, основанного на W ($E = \sum_{ij}^N p_{ij} f(x_{ij})$, $E \geq W$).

Рассмотрим модельный пример.

Допустим, малочисленная группа из 3 человек сдает сессию:

- каждый аттестуемый отвечает на 100 из общей базы объемом в 300 вопросов;

- при этом, 34, 33 и 33 вопроса имеют 5-8 вариантов ответов с, соответственно, 1-м, 2-м и 3-м «правильно»-помеченными ответами;

- документы тестового протокола атомарны (атом равен одному вопросу), причем:

- $\sqrt{p_{ij}} \in \{0, 1/300, 2/300, 3/300\}$, $af(x_{ij}) \in \{0, 1/3, 2/3, 1, 4/3, 5/3, 2\}$ для сессии;

- $\sqrt{p_{ij}} \in \{0, 1/200, 2/200\}$, $a f(x_{ij}) \in \{0, 1/2, 1, 3/2, 2\}$ для ВОУД (троечник «спрятан»);

- «троечник» всегда ставит только одну пометку, с трудом переваливает за суммарные 100 баллов и сессию сдает, в ВОУДе не участвует (имеет оправдательный документ);

- «хорошист» никогда не выбирает три пометки, сдает сессию и ВОУД;

- «отличник» не боится потерять баллы, выбирая «лишнее» (неверный ответ), в надежде набрать максимально возможный балл, успешно сдает сессию и ВОУД.

Суммируя $\sum_{ij}^{300} p_{ij} f(x_{ij})$, рассчитаем:

ЭШ) Энтропия Шеннона $S = - \sum_{ij}^{300} p_{ij} \log p_{ij}$.

Значение энтропии Шеннона:

- для сессии (расчеты для трех участников) находится в интервале от $\log(300) \approx 8.23$ (предельный случай, когда все отвечали на разные вопросы), до $3\log(33, (3)) \approx 15.18$ (предельный случай, когда все отвечали на один набор вопросов) (данный расчет годится для сессии с адаптивным тестовым набором заданий, от полностью совпадающим, до полностью измененным по составу вопросов для одного и того же испытуемого);

- для ВОУД (аналогичные расчеты для двоих участников) от $\log(200) \approx 7.65$ до $2\log(50) \approx 11.29$.

Итак, интервалы $[8.23; 15.18]$ и $[7.65; 11.29]$ дают перекрытие, то есть имеем:

- неразличимую ситуацию сессии от ВОУД, если распределения p_{ij} совершенно независимы;

- если же p_{ij} «однотипны», например, вопросы не совпадали ни у кого ни с кем, то случай «сессии» даст значение энтропии строго большего энтропии «ВОУД».

ЭР) Энтропия Реньи $S_R = \frac{1}{1-q} \log \sum_{ij}^{300} p_{ij}^q$.

Аналогичные расчеты дают интервалы [3.47; 8.23] и [4.64; 7.65], соответственно, для сессии и для ВОУД, и неразличимую ситуацию, за исключением «однотипных» (как и для энтропии Шеннона).

Обратим внимание: верхняя Реньи-оценка интервалов от q не зависит, и совпадает с нижней Шеннон-оценкой, которая получена для минимально возможного набора документов (тестовых заданий) – когда все тестируемые получают одни и те же 100 вопросов (такие оценки дают результаты пробного ВОУД-тестирования по одному единственному варианту заданий).

$$\text{ЭТ) Энтропия Тсаллиса } S_T = \frac{1}{q-1} (1 - \sum_{ij}^N p_{ij}^q).$$

Энтропия Тсаллиса, как известно [72], достигает максимума $S = \frac{N^{1-q}-1}{1-q}$ для $q>1$ в случае равновероятных распределений, то есть при $p_{ij}=1/N$ для модельного примера. Значения энтропии:

- для сессии $N=300, S \approx 8.23$ при $q \approx 1$;
- для ВОУД $N=200, S \approx 7.65$ при $q \approx 1$.

Если вопросы у тестируемых хотя бы частично совпали, энтропия Тсаллиса заметно падает, вплоть до $S \approx 4,61$ и для сессии, и для ВОУД.

Ситуация и в данном случае выглядит подобной разобранной выше, так как при ВОУД падает разнообразие комбинаций заданий (часть книжек тестов останется не выданной по упомянутым ранее причинам). Возможно, стоит разбивать тестовое задание на документальное множество пар «вопрос-ответ» с явным учетом того, что часть пар потребует содержательной «подсказки» из-за внутренней зависимости вариантов ответов, ранее объединявшихся вопросом. Как раз для таких вариантов учета и привлекается энтропия Тсаллиса с показателем $q \neq 1$.

Предлагается «энтропийная» схема замены концепции адаптивный тестовый набор (или другой внешней проверки остаточных знаний) перерасчет оценки за экзамен с учетом предыдущих тестирований итогов сессии.

1. Выбирается значение S_R^* энтропии Реньи из внутренних точек перекрывающихся интервалов, например, на основании соответствия значению индикатора Е. Для модельного примера это не ниже 4.64, а возможный выбор равенства верхней оценке 7.65 позволяет вариации параметра q в широких пределах, без принципиального влияния на результат.

2. Из использованной на сессии базы тестовых заданий собирается пул документов, статистика по которым дает именно выбранное значение энтропии S_R^* .

3. Исследуются результаты аттестуемых на построенном пуле тестовых заданий.

В сессионных базах содержится весь контингент аттестуемых (нет эффекта замены теста на справку), и отсутствует специально подобранная «повышающая» результаты группа «экспертов» (в сессию тестирование идет под видеонаблюдением; более того, возможно эффективное протоколирование процесса по побочным источникам информации [73]).

4. Результаты на выборочном пуле осредняются в соответствии с требованиями адаптивного тестового набора.

5. Повторное тестирование в форме «адаптивной схемы» (с выбором нескольких верных ответов) на выпускном курсе по сессионной базе и учет, в том числе, по предложенной схеме (1–4.) новых (повторных) результатов, даст оценки устойчивости итогов по критерию наличия «остаточных» знаний.

На рисунке 4.14 представлена схема применения энтропийной оценки для адаптивного теста.

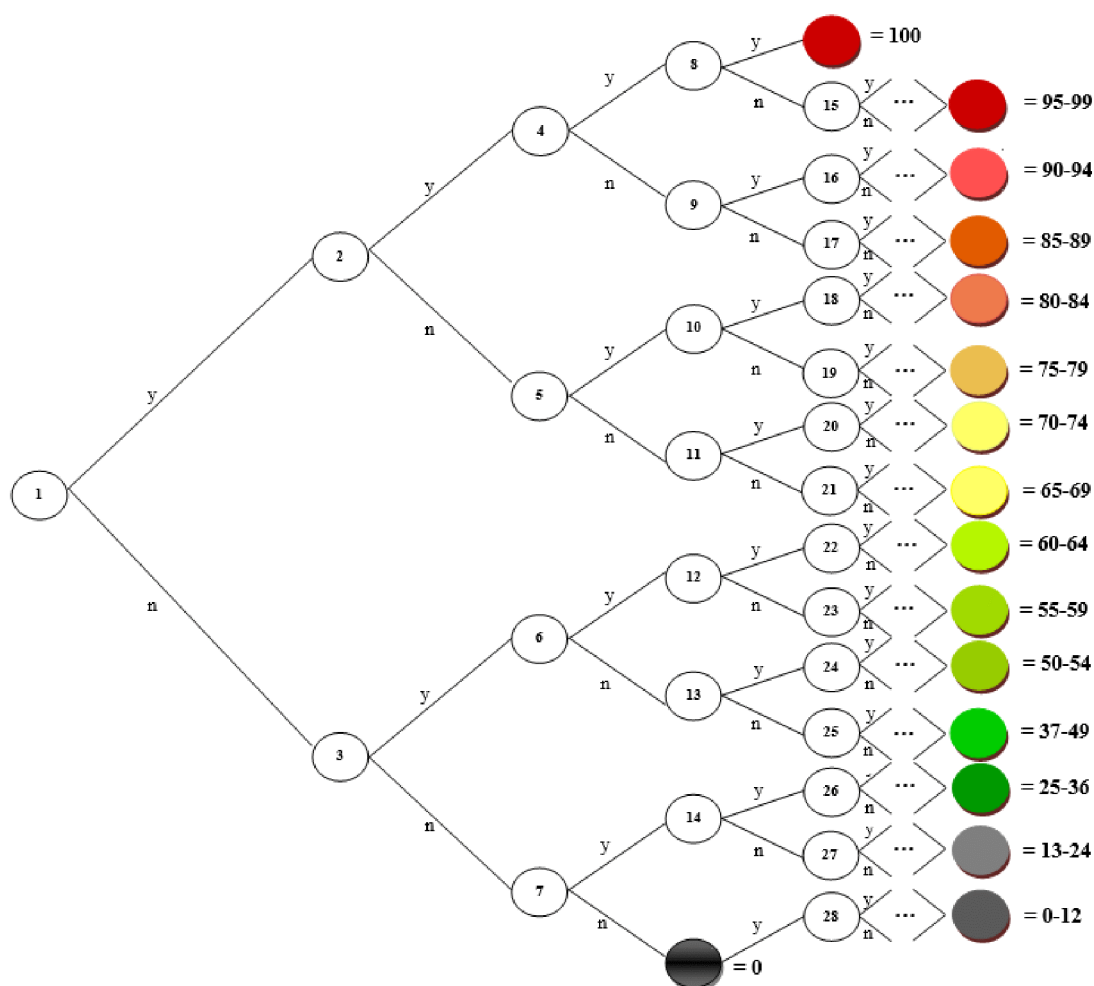


Рисунок 4.14 - Схема применения энтропийной оценки для адаптивного теста

На данной схеме отражены и помечены номерным знаком вопросы адаптивного характера с применением энтропийной оценки при переходе на каждый последующий вопрос. При положительном ответе (y) осуществляется переход ко второму вопросу более сложного характера (подача вопросов поступает от наименьшего к большему). Согласно схеме видно, что при положительном ответе (y) на все последующие сложные вопросы, а также при соблюдении всех правил прохождения адаптивного тестирования в системе ИСОРО обучающийся автоматически может получить «100» (сто) баллов, в

схеме данный уровень помечен бордовым цветом. При отрицательном ответе (n) осуществляется переход ко второму вопросу того же уровня что и первый вопрос. После трех неудачных ответов (n) система автоматически выставляет «0» (ноль) баллов, в схеме данный уровень помечен черным цветом. Уровни от темно-серого, до красного помечены и отражены в виде процентного значения согласно бально-рейтинговой системе.

Обратим внимание (статистически не доказано): поскольку, де-факто, база тестов с несколькими выборами ответа состоит из сессионных материалов головных вузов Республики Казахстан, для их контингента, аттестуемых сценарий (1.–4.) и происходит (в отсутствие энтропийных оценок нет гарантии, что не реализуется схема самой неудачной версии выборки/пула тестовых заданий).

Подчеркнем, что модельный расчет приведен только для простоты проверки вычислений, все выводы и предложения не меняются и для любой более объемной и реалистичной ситуации [72, стр. 98].

В системах контроля знаний формируются показатели, отражающих содержательную сторону контроля, в усредненной количественной форме. Обычно они задаются в виде функций от характеристик протекающих процессов и явно, или неявно учитывают, или стремятся учесть «типичные» проявления, предполагая наличие статистически значимых и достаточных для анализа данных с устойчивыми характеристиками [72, стр. 99].

4.4 Моделирование интерфейса и разработка оценочных заданий для работы в системе ИСОРО

«Система оценки результатов обучения ИСОРО предназначена в первую очередь для выполнения процедур компьютерного тестирования и предусматривает следующие виды работ» [74]:

- «создание и выполнение ТЗ с различными типами ответов: альтернативный, выбор одного ответа из нескольких многовариантных ответов, из разных графических ответов, свободное оформление ответов, комплексный ответ с возможностью ответа на слова семантического движения (семантические области);

- редактирование ранее созданных тестовых заданий (текст и ответы)» [74, стр. 87].;

- верные ответы преподавателя и правильная постановка задачи;

- тест производительности сети;

- выведение на экран протокола для преподавателя (ведомость экзаменационной сессии, подготовка и представление статистических данных о выводах экзаменационных процедур).

Система ИСОРО обеспечивает следующие функции и содержит функционал:

Для эксперта/преподавателя (педагога):

- создание тестового материала и «ведение различных работы с базами данных ТЗ по ведущим дисциплинам и темам;
 - создание и управление логическими базами данных, способность проведения оценочных занятий (рубежных) по любым темам одной или разных дисциплин в комплексе;
 - использование и применение различных моделей для создания ТЗ и ответов;
 - имеется возможность представить верный ответ в различных формах: введение свободного текст, формула, набор вариантов ответа, по ключевым словам, введение устного ответа;
 - проведение различных методов контроля, мониторинг и выведение анализа результатов;
 - настраивание определенного допуска администрирования, как для каждого так и отдельного обучающегося или создание предварительно общей настройки;
 - установка фильтра и определение даты, времени представления и сдачи, количества проделанных попыток и места представления контроля» [74, стр. 81];
 - оформление и выведение детальной статистики системы и возможность детального учета проделанной работы/протокол;
 - представление результатов проверки в различных образцах;
 - выведение и представление обработанных и собранных статистических результатов/данных о пройденной оценке знаний/контроль знаний в графическом виде;
 - картографирование: каждым обучающим прохождением тестового материала производится индивидуально; обработка и ранжирование результатов по группам; выведение формы дидактических единиц развития факторов контролируемой дисциплины; учебный план по результатам обучения в различных образовательных учреждениях;
 - проведение сбора статистических данных для проведения анализа качества, выведение из базы данных тестовых материалов и установление критериев;
- Обучающийся/тестируемый в системе ИСОРО:
- при прохождении контроля имеется возможность паузы/стоп;
 - возможность заявки к «порядку апелляции/обжалования результатов проверки согласно требованиям с восстановлением ответа и возможностью оценки ответа в интерактивном режиме, с правом корректировки» [74, стр. 81];
- администратор
- разделение прав на управление и введения учетных записей;
 - создание и введение демо-пользователей с правами по организации тестирования;
 - настраиваемая и регулируемая политика безопасности информации и т.д.

На рис. 4.15 представлен интерфейс системы оценки результатов

обучения, который ничем не отличается для всех пользователей системы.



Рисунок 4.15 – Интерфейс интеллектуальной системы оценки результатов обучения

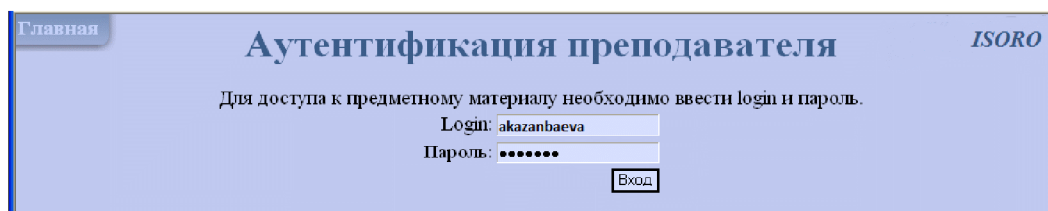


Рисунок 4.16 – Окно аутентификации преподавателя при входе в систему

После ввода данных преподаватель переходит в окно главного меню системы предоставленная на рисунке 4.17.

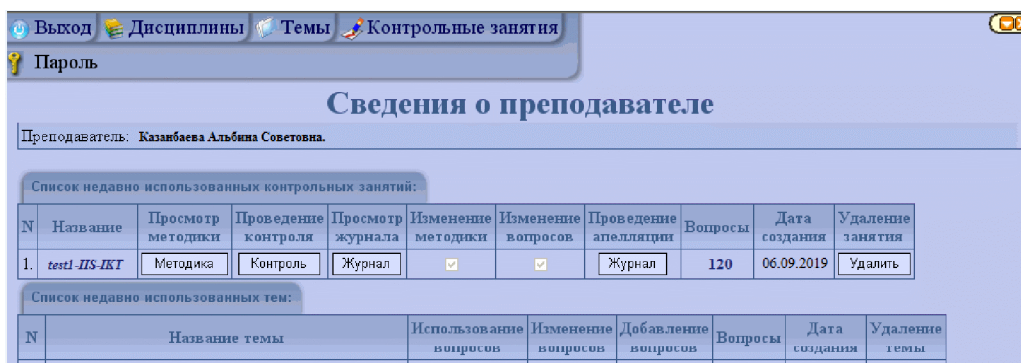


Рисунок 4.17 – Меню с методической работой

Окно содержит информацию о преподавателе, который вошел в систему. В главном пункте меню есть кнопка выхода, которая, соответственно, возвращает нас на титульную страницу системы. Пароль также можно

изменить с помощью вставки «Пароль», эта процедура позволяет повысить безопасность предметного материала, созданного ранее для контрольных работ (рис.4.18).

The screenshot shows a web application interface with a top navigation bar containing icons and labels for 'Выход' (Exit), 'Преподаватель' (Teacher), 'Дисциплины' (Disciplines), 'Темы' (Topics), and 'Контрольные занятия' (Control sessions). The main title is 'Смена пароля преподавателя' (Change teacher password). Below the title, the user is identified as 'Преподаватель: Казанбаева Альбина Советовна.' (Teacher: Kazanbaeva Albina Sovetovna). There are three input fields: 'Текущий пароль:' (Current password), 'Новый пароль:' (New password), and 'Подтверждение нового пароля:' (Confirmation of new password). A button labeled 'Изменить пароль' (Change password) is located at the bottom right of the form.

Рисунок 4.18 – Смена пароля преподавателя

Дальнейшая работа с подготовкой материала должна начинаться с вкладки «Дисциплины», которая закрепляется за преподавателем, а также для выбора новой дисциплины осуществляется либо непосредственно из всего списка, либо через фильтр. На рисунке 4.19 представлено окно выбора дисциплин из списка.

The screenshot shows a web application interface with a top navigation bar containing icons and labels for 'Выход' (Exit), 'Преподаватель' (Teacher), 'Дисциплины' (Disciplines), 'Темы' (Topics), and 'Контрольные занятия' (Control sessions). The main title is 'Список дисциплин для выбора' (List of disciplines for selection). Below the title, the user is identified as 'Преподаватель: Казанбаева Альбина Советовна.' (Teacher: Kazanbaeva Albina Sovetovna). There is a tab labeled 'Список дисциплин:' (List of disciplines:). Below the tab is a list of disciplines with checkboxes. The first four are checked: 'Информационная безопасность' (Information security), 'Компьютерные сети' (Computer networks), 'Операционные системы и системное программирование' (Operating systems and system programming), and 'Управление IT-проектами' (IT project management). The remaining disciplines are unchecked: 'вопрососборник' (Questionnaire), 'Internet-Intranet', 'NET-технологии' (NET technologies), 'NP-полные задачи' (NP-complete problems), 'PR в межкультурной коммуникации' (PR in intercultural communication), 'WEB-дизайн' (Web design), 'Автоматизация аудиторской деятельности' (Automation of audit activities), and 'Автоматизация бухгалтерской деятельности' (Automation of accounting activities).

Рисунок 4.19 – Выбор дисциплин

В данном окне необходимые дисциплины выделяем галочками, после чего они будут отображаться в отдельном окне как активные.

На рис. 4.20 представлены дисциплины, по которым был подготовлен тестовый материал. Видны заголовки полей таблицы: созданные темы (созданные автором / созданные только в дисциплине); количество вопросов по теме (созданные автором / созданные всего); проведенные контрольные работы (автор/только для дисциплины); используемые в контрольных занятиях (автор всех для всех предметов дисциплины). Эта подробная информация необходима для ведения статистики по использованию созданного материала во всех действующих дисциплинах.

N	Название	Количество тем	Количество вопросов в темах	Количество контрольных занятий	Количество вопросов в занятиях
1.	Информационная безопасность	16/117	226/4783	1/97	59/8915
2.	Компьютерные сети	15/16	403/402	1/1	52/52
3.	Операционные системы и системное программирование	1/1	117/117	0/0	0/0
4.	Управление IT-проектами	4/4	130/130	0/0	0/0

Рисунок 4.20 – Активные дисциплины преподавателя

Каждая тема относится к определенной дисциплине. Создание темы дисциплины начнется с выбора вкладки «Темы» в главном меню, а затем появится панель «Управление темами» (рис.4.21).

Тема сопровождается определенными данными от автора, названием, количеством созданных и заполненных вопросов; правами на этап; датой создания; может ли тема быть удалена, если тема не используется (или тема, для которой не был создан контрольный урок). Если преподаватель ранее создал темы для дисциплины, то список тем в таблице будет отображаться в столбце.

N	Название темы	Использование вопросов	Изменение вопросов	Добавление вопросов	Вопросы	Дата создания	Удаление темы
1.	Идентификация компьютера в сети	☒	☒	☒	60	07.10.2019	
2.	Базовая модель организации взаимодействия открытых систем (модель OSI).	☒	☒	☒	37	07.10.2019	
3.	Информационные процессы	☒	☒	☒	25	03.12.2019	

Рисунок 4.21 – Управление темами

Чтобы создать новую тему, необходимо перейти на вкладку «Создать тему», выбрать дисциплину, которую вы хотите отредактировать, введите имя темы, а затем нажмите кнопку «Создать» и перейдите в окно «Управление вопросами темы».

При разработке и выполнении тестового задания преподаватель имеет возможность предложить ответы на следующие виды тестовых заданий: выбор одного ответа из нескольких (закрыты вид); выбор нескольких из многих; свободная раскладка/построение; графический ответ; сконструированный ответ.

Чтобы ввести новое тестовое задание, преподаватель в режиме Управление темами вопроса нажимает кнопку «Добавить вопрос», а затем открывается редактор «Создать вопрос» представленный на рисунке 4.22.

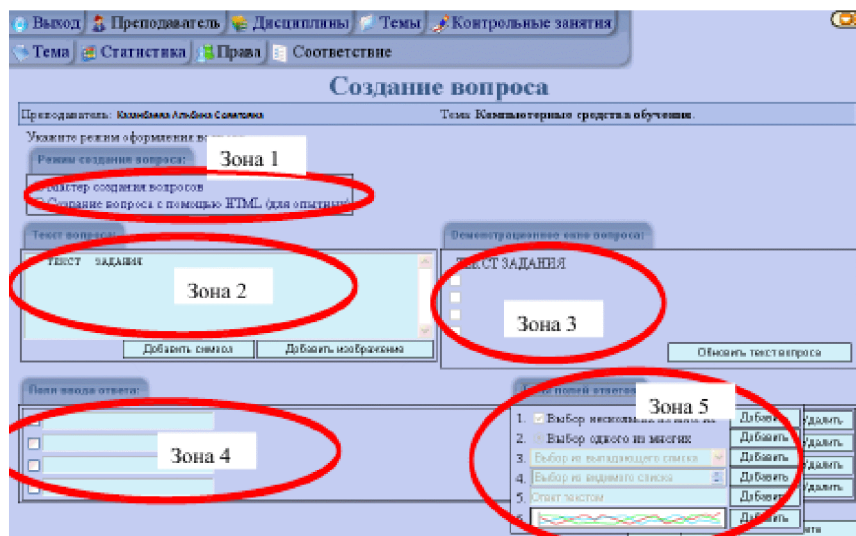


Рисунок 4.22 – Окно Создание вопроса

Данное окно состоит из нескольких зон: создание вопросов, текст вопроса; демонстрационное окно вопроса; поля ввода ответов; типы полей ответов. Который в свою очередь позволяет создавать вопросы, вводить формулировку, выводить вопрос для просмотра, вводить варианты ответов.

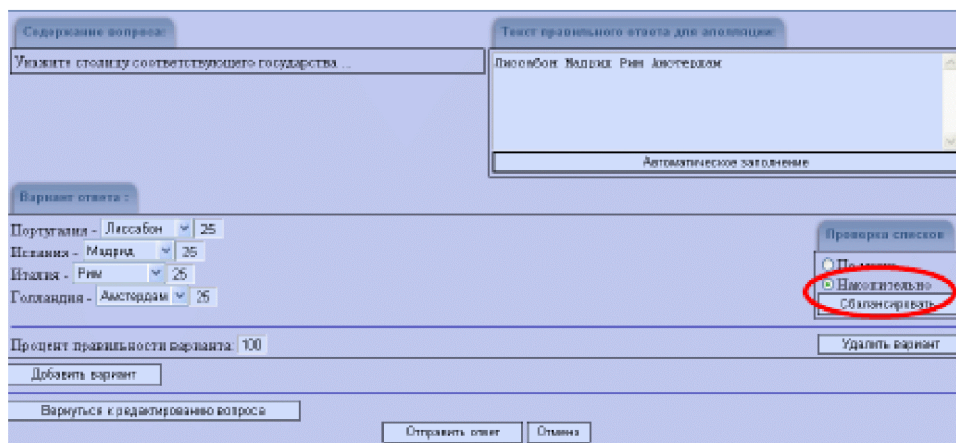


Рисунок 4.23 – Оформление правильного ответа на соответствие

Интуитивно понятно, что конструкция тестовых заданий закрытого типа с одним вариантом ответа, с выбором нескольких вариантов и вопросы открытого типа понятны.

На рис. 4.23 представлена версия вопроса на соответствие. Необходимо выбрать модель оценки ответов, также должны в поле «Проверка списков» выбрать модель оценки ответа «Накопительно» и нажать кнопку «Сбалансировать», поскольку все ответы равны проценту правильности каждого совпадения.

Данная система позволяет провести подготовку контрольного занятия как для отдельной темы, так и для нескольких вариантов тем. Редактирование во время процесса, то есть дополнительно добавить, отложить или удалить некоторые вопросы. Назначение сеанса контроля, согласно расписанию, либо дополнительно внеурочное время для самоконтроля обучающегося.

Основная деятельность интеллектуальной системы оценки результатов обучения является – непосредственное прохождение контроля знаний студента (обучающегося).

Для этого сеанса обучающемуся необходимо пройти на главную страницу, представленную на рисунке 4.15 (Интерфейс интеллектуальной системы оценки результатов обучения) и нажать кнопку «Тестирование».

Поэтому после нажатия на эту кнопку будет предложено завершить идентификацию, введя логин и пароль, выданные администратором при регистрации в учебную группу.

4.5 Иллюстрация результатов функционирования ИСОРО

Для апробации модуля системы оценки результатов обучения были использованы 114 итогов по семестру/результатов линейного тестирования/ответов адаптивного опроса (в идеологии ИСОРО), обучающихся на 3-4 курсе по специальности «Вычислительная техника и программное обеспечение» и «Строительство и эксплуатация зданий» Петропавловского строительного-экономического колледжа.

Адаптивный опрос обучающихся проводился во время промежуточной и итоговой аттестаций, то есть выдавались задания, ответ на который предоставлялся в устной/письменной форме при открытой форме задания или выбором, в зависимости от типа тестового задания. Оценивание результатов проводилось разработанным модулем системы оценки результатов обучения.

Процедура ответа, обучающегося на заданный вопрос представлен и описан в пункте 4.2, письменный/устный ответ, преобразованный в текст, направляется экспертам и/или экзаменатору/преподавателю для оценивания без машинным способом, весь алгоритм поступающих заданий формировался по реакции экзаменатора на полученные ответы в интерактивном режиме (если тестовая среда оказывалась не способна сама определить содержательность ответа).

Для наглядного представления полученных данных была применена сортировка результатов без учета принадлежности предмету и испытуемому, изображения динамики показателей (рис.4.24).

1	50	40	25	39	78	85	75	77	90	95	85
2	50	43	35	40	79	85	75	78	90	95	85
3	50	50	35	41	79	85	75	79	91	95	90
4	50	55	35	42	79	85	75	80	91	95	90
5	50	55	45	43	79	85	75	81	91	95	90
6	51	65	55	44	79	86	75	82	91	95	90
7	51	68	55	45	80	88	75	83	92	95	90
8	51	68	55	46	81	88	80	84	92	95	90
9	52	68	55	47	81	88	80	85	93	95	90
10	53	68	55	48	81	88	80	86	93	95	90
11	56	70	55	49	81	88	80	87	93	95	90
12	56	70	65	50	82	88	80	88	93	95	90
13	57	70	65	51	83	88	80	89	94	95	90
14	62	70	65	52	83	88	80	90	94	95	90
15	62	73	65	53	84	90	80	91	94	95	90
16	66	73	65	54	84	90	80	92	95	95	90
17	70	73	65	55	85	90	80	93	95	95	90
18	70	73	65	56	85	90	80	94	95	95	95
19	70	73	65	57	85	90	80	95	95	95	95
20	71	73	65	58	85	90	80	96	95	98	95
21	71	75	65	59	86	90	80	97	95	98	95
22	72	75	65	60	86	90	80	98	95	98	95
23	72	75	65	61	86	90	80	99	95	98	95
24	72	75	65	62	87	93	80	100	96	98	95
25	72	75	65	63	87	93	80	101	96	98	95
26	72	78	65	64	87	93	85	102	97	98	95
27	73	78	65	65	88	93	85	103	97	98	95
28	73	78	65	66	88	93	85	104	97	98	95
29	73	80	75	67	88	93	85	105	97	98	95
30	74	80	75	68	88	93	85	106	97	98	95
31	75	83	75	69	89	93	85	107	98	98	95
32	76	83	75	70	89	93	85	108	98	98	95
33	77	83	75	71	89	93	85	109	98	98	95
34	77	83	75	72	89	93	85	110	99	98	95
35	77	83	75	73	89	93	85	111	99	98	95
36	77	83	75	74	89	93	85	112	100	98	95
37	77	85	75	75	90	93	85	113	100	98	95
38	78	85	75	76	90	95	85	114	100	100	100

Рисунок 4.24 – Сортировка результатов без учета принадлежности предмету и испытуемому

Первая колонка показателей – семестровые промежуточные итоги результатов обучения, вторая – результаты линейного теста (экзамена), третья – результаты проверки заданий в ИСОРО. Первую колонку можно трактовать как результат экспертизы группой экспертов, последнюю – как адаптируемую схему экспертизы с опорой на преимущественно одного эксперта.

Оценивание линейной схемой тестирования выглядит завышенным, по сравнению с оцениванием группой экспертов, оценивание по схеме ИСОРО – заниженным по сравнению с оценкой той же группой экспертов, в последнем случае, возможно, сказалась меньшее количество градаций оценки (с шагом не менее пяти).

Те же самые итоги в привязке к испытуемому, будучи отсортированными по возрастанию итогов за семестр (первая колонка таблицы, трактуемых как оценивание группой экспертов) построены в excel (рис.4.25) и демонстрируют существенное количество инверсий оценки полученной в итоге линейного

теста (второй колонки), как с завышением, по сравнению с итогами семестра, так и с занижением.

1	78	65	39	98	80	77	93	85
2	68	55	40	83	80	78	93	85
3	65	55	41	88	75	79	93	85
4	50	35	42	75	75	80	93	85
5	43	25	43	88	75	81	98	90
6	83	75	44	93	80	82	93	85
7	68	55	45	95	80	83	95	95
8	55	35	46	95	80	84	95	90
9	90	80	47	88	80	85	95	95
10	70	45	48	88	80	86	95	90
11	70	55	49	68	65	87	93	90
12	70	65	50	85	80	88	95	90
13	73	65	51	90	75	89	95	90
14	75	65	52	85	80	90	73	75
15	73	65	53	90	90	91	88	90
16	73	65	54	98	80	92	95	95
17	95	65	55	80	80	93	90	90
18	85	65	56	90	80	94	98	90
19	78	65	57	90	80	95	98	95
20	75	65	58	88	75	96	98	90
21	40	55	59	93	80	97	95	95
22	75	75	60	93	85	98	98	95
23	83	75	61	78	75	99	95	95
24	95	75	62	85	85	100	95	95
25	83	65	63	93	80	101	98	95
26	88	65	64	90	85	102	98	95
27	95	80	65	93	90	103	98	95
28	55	35	66	85	85	104	95	95
29	68	65	67	95	90	105	98	95
30	83	65	68	80	80	106	98	90
31	83	75	69	93	95	107	98	100
32	85	75	70	95	85	108	98	95
33	70	65	71	95	85	109	98	95
34	85	75	72	90	85	110	93	85
35	73	65	73	90	85	111	98	95
36	73	55	74	75	75	112	98	95
37	93	75	75	88	85	113	100	95
38	86	75	76	95	90	114	98	95

Рисунок 4.25 – Итоги в привязке к испытуемому

Сама инвертированность отражает показатель системой линейных тестов, не удовлетворяющий как преподавателей, так и студентов (отражено в обработке анкет в первой главе работы на рисунке 1.4. Третья колонка, полученная по схеме ИСОРО, так же содержит инверсии, по сравнению с оценкой за семестр (имеются уникалы, как провалившие все виды тестирования, так и наоборот, превзошедшие ожидания по итогам семестра), но, очевидно, в гораздо меньшей степени. Особенно выделяется средняя часть «спектра» оценок, схема ИСОРО, в среднем более занижающая (как следует

из предыдущей иллюстрации (рис. 4.24) оценку допускает инверсии, во-первых, в меньших масштабах, а во-вторых, в основном, в «соседние» области оценок, за исключением нескольких уже упомянутых «уникумов».

Данная иллюстрация результатов позволяет предположить, что адаптация по схеме ИСОРО не сильно отклоняется от результатов работы в течении семестра, как у экспертов, так и у испытуемых, и сопровождается более закономерным результатом.

Более подробно можно увидеть на следующем рисунке 4.26, на котором дополнительно собраны итоги по предмету по текущим правилам итогового оценивания предмета, а именно - 60% вклада оценки, накопленной за семестр и 40% оценки за итоговое тестирование (экзамен). Левее (внутри центральной белой зоны) учтены по принятой схеме итоги линейного теста, правее (крайняя правая одиночная колонка) - тестирование по схеме ИСОРО.

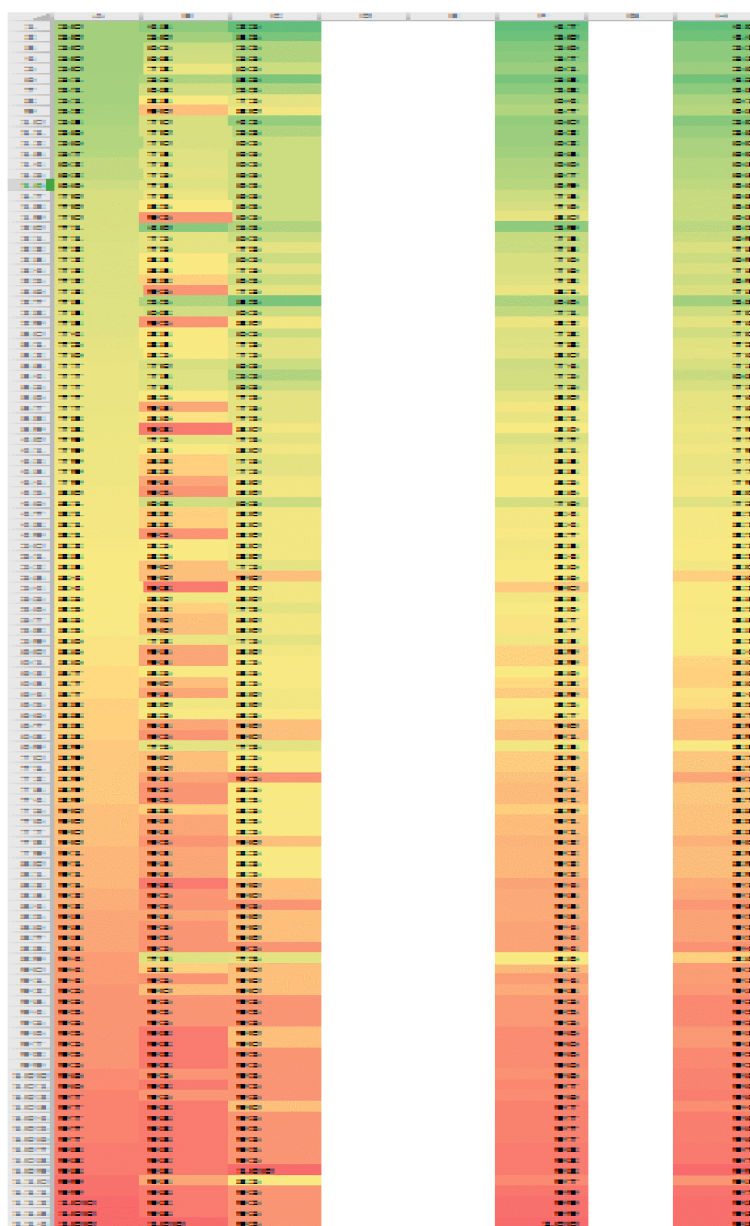


Рисунок 4.26 – Итоги по предмету

На данном рисунке видно, что основная часть «полосатости» из-за инверсий оценок тестированием примерно перекликается (уникумы «сидят» ровно в одних и тех же позициях), это позволяет предполагать, что юридическая замена сессионных итогов с линейной системы тестирования на систему ИСОРО не повлияет пагубно уже сложившуюся годами схему подведения итогов обучения студентов в рамках сессии.

Поскольку «балльная» система оценивания не может быть, безусловно признана количественной, более того, идея суммирования «с процентами» вообще выглядит не понятной с математической точки зрения, хотя и «освященной» законодательно и общепринятой, решено было рассмотреть итоговую таблицу с результатами в ее номинальном виде. На рисунке 4.27 все данные отображаются на номинально построенном дереве (учитывающем всю таблицу рисунка 4.25), ветвящемся, в итоге, на 114 кластеров. Таким образом, успешно можно рассматривать меньшее количество кластеров, например, 8 или 9, чтобы ассоциировать разбиение с принятой нынче буквенной нотацией от А до F.

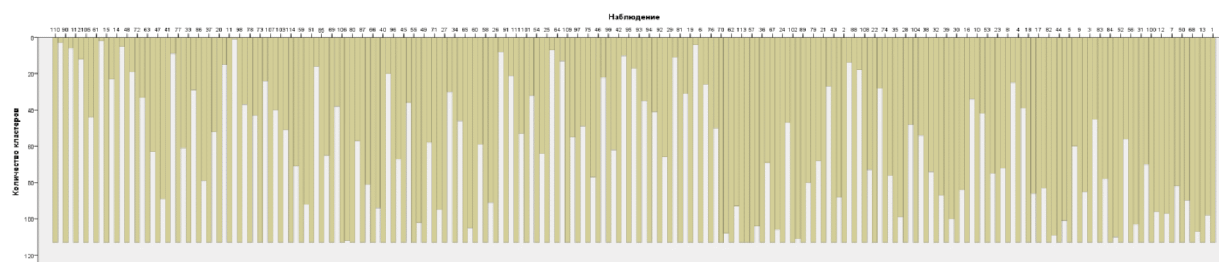


Рисунок 4.27 – Номинальное дерево на 114 кластеров

Это отражено на рисунке 4.28 (то же самое отражено на рисунке 4.27, но немного мешает масштабирование на десятки кластеров), из которого следует, что попытка учесть влияние всех добытых «оценок» (в номинальном смысле) приводит к хорошо представленным (А, А-, В+, В, В-, С+, следует читать список по зонам от правого края иллюстрации) и «скомканным», хотя и сопоставимым между собой (С, С-, D+, D здесь крайне левая область).

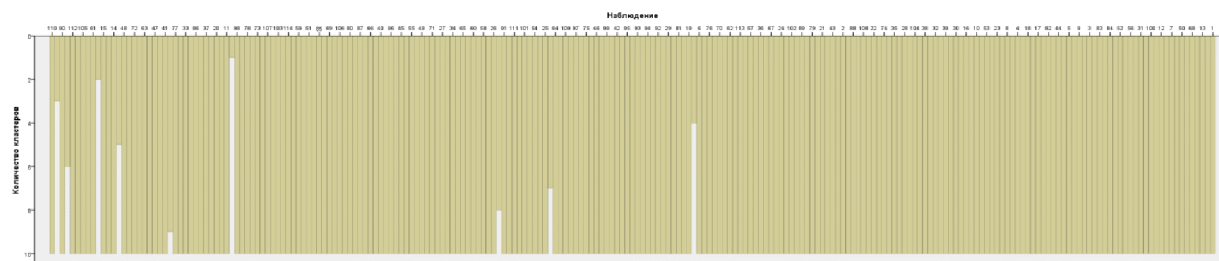


Рисунок 4.28 – Номинальное дерево на 10 кластеров

Введение порядкового прочтения содержимого таблицы отражено на рисунке 4.29, с учетом автоматически рассчитываемых средних рангов картину несколько перестраивает, оставляя, в целом, тоже самое ощущение «скомканности» левой области, что может быть естественным образом

объяснено, поскольку учитывается очень инвертированная «картина мира» вносимая итогами линейного теста.

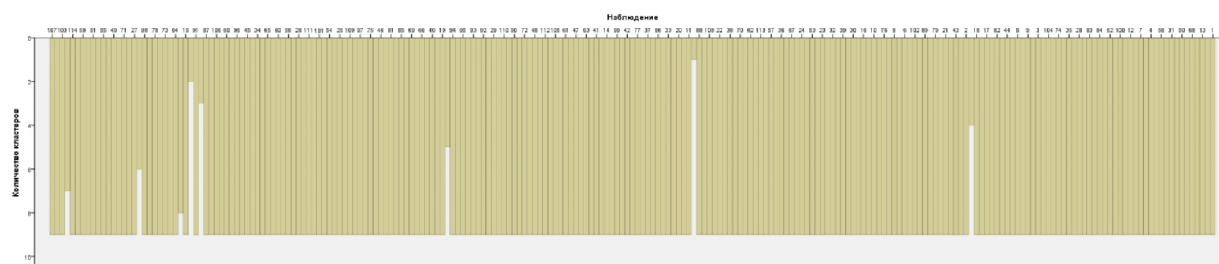


Рисунок 4.29 – Итоги линейного теста

«Чистый» результат иллюстрации, но без перехода в количества и данные с процентами, как на рисунке 4.26, отражен на рисунке 4.30 (семестровые итоги и итоги линейного теста кластеризованы в порядковой шкале с автоматически рассчитанными средними рангами).



Рисунок 4.30 – Итоги семестровые и линейного теста

Также на рисунке 4.31 (семестровые итоги и итоги теста ИСОРО кластеризованы в порядковой шкале с автоматически рассчитанными средними рангами).



Рисунок 4.31 – Итоги семестровые и теста ИСОРО

Кластеризация на рисунке 4.30 указывает на некоторую «несамостоятельность» оценок В⁺ и В⁻, как и в случае с рисунком 4.31, где, «сплошной» ощущается и оценка С.

В целом, вывод остается тем же, работоспособность схемы ИСОРО никак не хуже тех возможностей, что законодательно доступны классической схеме с линейным тестом.

4.5 Выводы по четвертой главе

1. Сформулированы и предложены рекомендации по управлению информационной поддержки системы оценки результатов обучения. Рекомендации представлены с помощью методологии IDEF0, обозначены основные фрагменты деятельности и критерии управления системой оценки учебных достижений.
2. Разработан модуль интеллектуальной системы оценки результатов обучения, позволяющий прохождение тестирования, провести контроль устного/письменного экзамена.
3. Также в работе показано наличие и возможности иных методов анализа, опирающихся, прежде всего на энтропийные характеристики, менее требовательные к объемам и качеству данных, в то же время, активно использующие, например, уже имеющиеся в системе материалы и общепринятый документооборот. Новые энтропийные подходы, возможно, использовать в системе контроля для целей принятия решений по ее совершенствованию.
4. Предложена краткая инструкция работы по созданию и редактированию тестовых заданий, сформированных для прохождения адаптивного тестирования (для преподавателя).
5. Проведена апробация и анализ эффективности модуля системы оценки результата обучения. Согласно выводам, работоспособность схемы ИСОРО никак не хуже тех возможностей, что законодательно доступны классической схемы линейной оценки результатов обучения.

ЗАКЛЮЧЕНИЕ

На сегодняшний день система образования характеризуется интенсивным развитием теории и практики компьютерного тестирования, которое в свою очередь является важной и неотъемлемой частью образовательной информационной среды любой образовательной организации в период пандемии.

Компьютерное тестирование – является методом контроля, представляющий собой стандартизированный процесс применения тестов посредством «компьютера, под управлением специального проекта, обеспечивающей необходимо нужное представление тестовых заданий и обработку результатов тестирования»[75] для достижения целей и задач. При создании содержания теста необходимо основываться на несколько методологические правила, наиболее важными из которых являются значимость, взаимосвязь содержания и формы, однозначность, репрезентативность и вариативность.

Материалы первой главы позволили проанализировать и констатировать необходимость и возможность использования технологии компьютерного тестирования как средство диагностики результатов обучения.

Переход на новую систему оценки знаний в сочетании с правильно подобранным контролем учебной деятельности позволит перейти на качественно новый уровень обучения в вузе. Такой подход к обучению поможет повысить эффективность обучения.

«В практике современного образования нет универсального метода оценки, который бы полностью решал проблему объективности при оценке профессиональной компетентности. Одни методы оценки позволяют судить о результатах формирования определенных видов профессиональной компетентности, другие, соответственно, о других»[23].

Подводя итоги можно сделать несколько выводов по результатам работы:

1) Представлены результаты теоретического исследования современных направлений развития непрерывного контроля. Изучены основные сферы применения данных систем. Проанализированы эффективные методы оценки результатов обучения, востребованные в современных информационных системах. Прикладным аспектом использования полученного в процессе исследования результата является возможность применения в проектировании и разработке системы оценки учебных достижений.

2) Представлена модель интеллектуальной системы оценки результатов обучения. Проведен анализ результатов тестирования, изложены особенности работы системы. Обосновано применение энтропийного подхода к системе обучения и аттестации.

3) Представлены методы и алгоритмы управления процессами интеллектуальной системы оценки результатов обучения. Также представлен метод генерации тестов, основанный на вычислении достижимого и связанного недо-приближения абстракции системы событий.

4) Предложены инструментальные программные средства, позволяющие проведения онлайн аттестации с применением адаптивной, автоматизированной системы.

5) представлен модуль системы (для применения устного/письменного экзамена в формате онлайн) оценки результатов обучения.

6) Сформулированы и предложены рекомендации по принципу организации системы оценки результатов обучения. Рекомендации представлены с помощью методологии IDEF0, обозначены основные фрагменты деятельности и критерии управления системой оценки учебных достижений.

В связи с широким внедрением онлайн обучения, система контроля знаний приобрела особую актуальность разработки контролирующих систем, основанных на тестировании, разработка методики составления тестов. Под оптимизированной обработкой данных тестирования понимается организация ввода результатов тестирования в базу хранения, сопряжение этой базы с базой тестовых материалов, обработка результатов, тестируемых и получение отчетных данных с минимальными затратами всех видов ресурсов.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Послание Президента Республики Казахстан Н. Назарбаева народу Казахстана. – 2017. – 31 января. https://www.akorda.kz/ru/addresses/addresses_of_president/poslanie-prezidenta-respubliki-kazahstan-nazarbaeva-narodu-kazahstana-31-yanvarya-2017-g/
2. Шевчук Е.В., Икласова К.Е., Казанбаева А.С., Козак Е.А. Алгоритм определения смысловой схожести разноязычных текстов для мобильных систем проверки результатов обучения // IV Международная научно-практическая конференция (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук», Тольяттинский государственный университет. Тольятти. 2018, стр.337-342.
3. А.с. №11321. Модуль интеллектуальной системы оценки результатов обучения/А.С. Казанбаева, В.П. Куликов. kazpatent.kz; опубл. 10.07.2020.
4. <http://ru.wikipedia.org/wiki/Аутентификация>
5. Закон Республики Казахстан от 7 января 2003 года №370 (с изменениями и дополнениями от 13.11.2019 г.) «Об электронном документе и электронной цифровой подписи».
6. Аванесов, В.С. Форма тестовых заданий. – М., 2006.
7. Аванесов В.С. Теория и практика педагогических измерений URL: http://charko.narod.ru/tekst/biblio/Avanesov_Teoriya_i_metod_ped_izmer.pdf
8. Аванесов, В.С. Из глубины веков //Педагогическая диагностика. - 2003. - №1. – С.3-7.
9. <http://ibooks.ru/reading.php?short=1&productid=341649>
10. Конопко Е.А. Использование компьютерного тестирования в процессе профессиональной подготовки бакалавров в вузе <http://dlib.rsl.ru/rsl01003000000/rsl01003174000/rsl01003174317/rsl01003174317.pdf>
11. Казанбаева А.С., Процедура подготовки документального сопровождения компьютерного тестирования на базе Северо-Казахстанского государственного университета. Актуальные вопросы образования. Сибирский государственный университет геосистем и технологий. Россия, г.Новосибирск, (2019).
12. Мусин М.Б. Идентификация пользователя средствами NeuralNetwork по клавиатурному почерку. Вестник Павлодарского педагогического университета имени С. Торайгырова. г. Павлодар. №4. 2018. стр. 51-55. – ISSN 1811-1807.
13. Куликов В.П., Мукашев Р.М. Анализ систем распознавания речи с открытым API. Вестник Павлодарского педагогического университета имени С. Торайгырова. г. Павлодар. №1. 2019. стр. 46-54. – ISSN 1811-1831.
14. Денисюк А.В. Применение акустического криптоанализа ЖК мониторов для контроля самостоятельности сдачи тестирования. Вестник

Павлодарского педагогического университета имени С. Торайгырова. г. Павлодар. №2. 2019. стр. 86-91. – ISSN 1811-1858.

15. Мухатаев А.А., Шевчук Е.В., Казанбаева А.С. Білімді бағалаудың түрлері және тестілеудің рөлі. Innovation and global issues in social sciences, Patara, april 27-29, 2017, p.953-958.

16. Чуб Е.В. Формирование общих и профессиональных компетенций студентов в условиях интеграции профессионального и дополнительного образования. Инновации в образовании. 2014. № 9. С. 83-91. <http://elibrary.ru/item.asp?id=21834663>

17. «Государственный общеобязательный стандарт высшего образования» Утвержденный постановлением Правительства РК от 22 августа 2012 года №1080 (с изменениями от 31.10.2018, №604).

18. Майоров, А.Н. Теория и практика создания тестов для системы образования. – М.: «Интеллект-центр», 2001. – С. 82–101.

19. Васильев С.А, Жанаева А.С. К вопросу о формализации аксиоматического аппарата теории общественного выбора //материалы конференции Дискретный анализ и исследование операций (DAOR'04), Новосибирск, 2004, С.197.

20. <https://socioline.ru/rv.php>

21. <https://fdfgroup.ru/poleznaya-informatsiya/stati/vyborka-tipy-vyborok-raschet-oshibki-vyborki/>

22. Куликов В.П., Казанбаева А.С. Комплект диагностических средств оценки результатов. Вестник Павлодарского педагогического университета имени С. Торайгырова. г. Павлодар. №1. 2019. стр. 208-218. – ISSN 1811-1831.

23. Шевчук Е.В., Казанбаева А.С. Система оценивания образовательных результатов бакалавра образования. VII Международная научно-техническая конференция «Энергетика, информатика, инновации – 2017», Том -3, Россия, г. Смоленск. 2017, стр.293-297. <http://www.sbmpei.ru/files/uplfiles/f5a2956e97c36bTOM-3.pdf#2>

24. Куликов В.П., Куликова В.П. Адаптация принципов АСМ ICPC к дистанционному обучению // Материалы XII Всероссийской научно-практической конференции с международным участием «Дистанционное обучение в высшем профессиональном образовании: опыт, проблемы и перспективы развития», СПб ГУП, 2019.

25. Вятчин Д.А. Нечеткие методы автоматической классификации. - Минск, 2004. – 216 с.

26. Куликов В., Куликова В. К вопросу тестирования // Современные тенденции естественно-математического образования: школа – вуз [Текст]: материалы VIII Международной научно-практической конференции, 12 – 13 апреля 2019 года: в 2 ч. Ч. 1 / Соликамский государственный педагогический институт (филиал) ФГБОУ ВО «ПГНИУ», Т. В. Рихтер, составление. – Соликамск: СГПИ; ООО «Типограф», 2019. – 194 с.

http://solgpi.ru/pdf/math_sbornik_2019_1.pdf#1, <https://nkzu.antiplagiat.ru/report/source/44?v=1&source=36&block=19>

27. Шевчук Е.В., Казанбаева А.С. Непрерывный контроль как необходимое условие качественной оценки учебных достижений. Международная научно-практическая конференция «Непрерывное образование в XXI веке: проблемы, тенденции, перспективы развития», Непрерывный контроль как необходимое условие качественной оценки учебных достижений. Россия, г. Шадринск. 2016, стр. 143-147.

28. Мутанов Г. М., Шевчук Е. В. Экспертная система оценки знаний методом тестирования. – Петропавловск, 2001.

29. Баврина А.П. Современные правила использования методов описательной статистики в медико-биологических исследованиях. Медицинский альманах 2020; 2(63): 95–104. Bavrina A.P. Modern rules for the use of descriptive statistics methods in biomedical research. Medicinskij al`manah 2020; 2(63): 95–104.

30. Fox N., Hunn A., Mathers N. Sampling and sample size calculation. – Yorkshire & the Humber: The NIHR RDS for the East Midlands. 2007: 41.

31. Хили Дж. Статистика: Социологические и маркетинговые исследования. СПб: Питер, 2005.

32. Наркевич А.Н., Виноградов К.А. Методы определения минимально необходимого объема выборки в медицинских исследованиях. Социальные аспекты здоровья населения [сетевое издание] 2019; 65(6): 10. URL: <http://vestnik.mednet.ru/content/view/1123/30/lang.ru/>, <https://doi.org/10.21045/2071-5021-2019>

33. Крамер Г. Математические методы статистики - М.: Мир, 1975. — 648 с.

34. Зверев Е.Б. Атрибутивная малая выборка – применение в аудиторских процедурах // Внутренний контроль в кредитной организации" № 4 (36) \ 2017.

35. Методологические положения по статистике (выпуск 1,2,3,4,5) https://www.gks.ru/bgd/free/B99_10/IssWWW.exe/Stg/d020/i020030r.htm

36. Хей Дж. Введение в методы байесовского статистического вывода / Дж. Хей / Пер. с англ. А. А. Рывкина — М.: Финансы и статистика, 1987. — 355 с.

37. Крылова Е.М., Казанбаева А.С. Тестіленушілердің білім деңгейін бағалау әдістерін әзірлеу мәселесі туралы. «Вестник КазННТУ», г. Алматы. 2019. №3. стр. 366-372. ISSN 1680-9211.

38. Сивуха С.В., Козяк А.А О реформе статистического вывода в психологии (Сомнительная значимость статистической значимости) Психология. Журнал Высшей школы экономики. 2009. Т. 6, № 4. С. 66–86.

39. https://studme.org/1260102029269/psihologiya/problemy_kotorye_kazhdyy_raz_nuzhno_reshat_zanovo.

40. Вероятность и математическая статистика: Энциклопедия / Под ред. Ю.В. Прохорова. — М.: Большая российская энциклопедия, 2003. — 912 с.

41. <https://excel2.ru/articles/oshibka-vtorogo-roda-i-krivye-operativnoy-harakteristiki-v-ms-excel>.

42. Мартынов, В.В. Теоретические и методологические основы создания информационных систем управления подготовкой специалистов по требованиям работодателей / В.В. Мартынов, Е.И. Филосова, Ю.В. Шаронова, О.В. Ширяев // Управление экономикой: методы, модели, технологии: материалы XV Международной научной конференции. В 2 т. Т. 2 / Уфимск. гос. авиац. техн. ун-т. – Уфа: Уфимск. гос. авиац. техн. ун-т., 2015. – С. 220-223.

43. Фандрова, Л.П. Применение тестирования в рамках адаптивных обучающих систем [Электронный ресурс]/ Л.П. Фандрова, О.В. Ширяев // Ежегодная конференция «Использование программных продуктов 1С в учебных заведениях». – Режим доступа: <http://www.1c.ru/rus/partners/training/edu/these s/?y=2010&s=46&t=1085>

44. Bride Hadrien, Julliand Jacques, Masson Pierre-Alain. Tri-modal under-approximation for test generation // Science of computer programming. Vol:132. pp.190-208.

45. Julliand, J., Kouchnarenko, O., Masson, P.A., Voiron, G. Test Generation from Event System Abstraction to Cover Their States and Transition // Programming and computer software. - 2018. Vol: 44. - pp. 1-14.

46. Сумин В. И., Кравченко А. С., Рябинин В. В. Адаптивное тестирование. Логические модели Раша и Бирнбаума // Вестник ВГТУ. 2009. №6. URL:<https://cyberleninka.ru/article/n/adaptivnoe-testirovanie-logicheskie-modeli-rasha-i-birnbauma>.

47. Куликов В.П., Шатилова Е.М., «Автоматизация процесса проведения практических занятий по программированию» Материалы V республиканской студенческой научно-технической конференции, Усть-Каменогорск, 2005, 2 с.

48. Куликов В.П., «Квалификационное оценивание в ВУЗе», Научный журнал. Межвузовский вестник. Петропавловск, 2004, 7 с.

49. Куликов В.П., «О «тонкостях» в компьютерных технологиях оценки знаний методами тестирования», Вестник ВКГТУ им. Д.Серикбаева, №1(35), Усть-Каменогорск, 2007, 5 с.

50. Kulikov V., Kulikov A. Modeling of the group effect during the studying process, Материалы конференции, СКГУ им. М. Козыбаева, Петропавловск, 2008, 8 с.

51. Данилова, С.Д. Управление учебной деятельностью на кафедре, [Текст] / Л.В. Найханова, С.Д. Данилова, С.В. Тумашов, Е.Ю. Ангаева, // Сборник научно-методических статей, - Вып.8, - Улан-Удэ: ВСГТУ, - 2002, - С.114-120.

52. Bride, H., Julliand, J., and Masson, P.-A., Tri-modal under-approximation for test generation, *Sci. Comput. Program.*, 2016, vol. 132, no. P2, pp. 190–208.

53. Godefroid, P. Compositional dynamic test generation (2007) ACM SIGPLAN Notices, 42 (1), pp. 47-54.

54. Bué, P.-C., Julliand, J., Masson, P.-A. Association of under-approximation techniques for generating tests from models (2011) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6706 LNCS, pp. 51-68.
55. Godefroid, P., Klarlund, N., Sen, K. DART: Directed Automated Random Testing (2005) Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI), pp. 213-223.
56. Tillmann, N., De Halleux, J. Pex-white box test generation for. NET (2008) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 4966 LNCS, pp. 134-153.
57. Makarov, S.I., Sevastyanova, S.A. Information modeling of the students' residual knowledge level (2020) Advances in Intelligent Systems and Computing, 908, pp. 502-509.
58. Wise, S.L., Kuhfeld, M.R., Soland, J. The Effects of Effort Monitoring with Proctor Notification on Test-Taking Engagement, Test Performance, and Validity, (2019) Applied Measurement in Education, 32 (2), pp. 183-192.
59. Wise, S.L., Gao, L. A General Approach to Measuring Test-Taking Effort on Computer-Based Tests. (2017) Applied Measurement in Education, 30 (4), pp. 343-354.
60. Bershadskaya, M.D., Serova, A.V., Chepurensky, A.Yu., Zima, E.A. Competence-based approach to learning outcomes assessment: Russian experience in sociological Education. (2019) Vysshee Obrazovanie v Rossii, 28 (2), pp. 38-50.
61. Muchuchuti, S., Narasimhan, L., Sidume, F. Classification model for student performance amelioration. (2020) Lecture Notes in Networks and Systems, 69, pp. 742-755.
62. Wise, S.L., Kingsbury, G.G. Modeling Student Test-Taking Motivation in the Context of an Adaptive Achievement Test. (2016) Journal of Educational Measurement, 53 (1), pp. 86-105.
63. Lendyuk, T., Sachenko, S., Rippa, S., Sapojnyk, G. Fuzzy rules for tests complexity changing for individual learning path construction. (2015) Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2015, 2, № 7341443, pp. 945-948.
64. Maravić Čisar, S., Čisar, P., Pinter, R. Evaluation of knowledge in Object Oriented Programming course with computer adaptive tests. (2016) Computers and Education, 92-93, pp. 142-160.
65. Mantzaris, A.V., Walker, T.G., Taylor, C.E., Ehling, D. Adaptive network diagram constructions for representing big data event streams on monitoring Dashboards. (2019) Journal of Big Data, 6 (1), №24,
66. Özyurt, H., Özyurt, O., Baki, A., Güven, B. Integrating computerized adaptive testing into UZWEBMAT: Implementation of individualized assessment module in an e-learning system. (2012) Expert Systems with Applications, 39 (10), pp. 9837-9847.

67. Kazanbayeva A.S., Kulikov V.P., Iklassova K.I., Development of a method for assessing learning outcomes through automated testing management, Periodico Tche Quimica, ISSN 2179-0302. (2019); vol.16 (n°33), pp. 796-812.
68. Шевчук Е.В., Шпак А.В. Модель повышения надежности компьютерного теста // Математическое и компьютерное моделирование. Сборник материалов VI Международной научной конференции, посвященной памяти Б.А. Рогозина. – Омск: Омский государственный университет им.Ф.М. Достоевского, 2018. – С. 77-79.
69. A.I. Wilson, 1978, Entropy in Urban and Regional Modeling, Moscow, Nauka Publishing.
70. O.V. Chumak, 2010, Entropies and Fractals in Data Analysis, Moscow, Sternberg Astronomical Institute Publishing.
71. https://matlab.ru/articles/Ver_7.pdf(Date of access: December 24, 2018).
72. Kulikov V., Iklassova K., Kazanbayeva A. Entropy based decision making method in managing the development of asocioin formational system. Journal of Theoretical and Applied Information Technology, Vol.98 №1. 92-102 pp.-ISSN 1817-31952020.
73. Daniel Genkin, Mihir Pattani, Roei Schuster, Eran Tromer. Synesthesia: Detecting Screen Content via Remote Acoustic Side Channels – Santa Barbara: CRYPTO 2018 conference.
74. Красильникова В.А. Использование информационных и коммуникационных технологий в образовании. Учебное пособие. Оренбург. 2012. http://window.edu.ru/resource/286/76286/files/пособие_ИИКТ_Красильникова.pdf.
75. Данилова, С.Д. Адаптивная нечеткая модель оценивания результатов автоматизированного тестирования с разделением заданий по уровням усвоения: дисс. ... канд. тех. наук: 05.13.01 / Улан-Удэ: ВСГТУ, 2006.http://dlib.rsl.ru/rsl01002000000/rsl01002976000/rsl01002976113/rsl01002976_113.pdf.
76. Кремер Н.Ш. Теория вероятностей и математическая статистика, М.: Юнити, 2000. – 543 с.
77. Общеобразовательный тест. Наука и жизнь URL: <https://www.nkj.ru/archive/articles/10135/> (дата обращения 12.01.2019).
78. World Finals Rules for 2019 - Updated 4 Oct. 2018 URL: <https://icpc.baylor.edu/worldfinals/rules> (датаобращения 12.01.2019).
79. Ошибки первого и второго выбора. <https://ru.wikipedia.org/wiki>.

ПРИЛОЖЕНИЕ А

Свидетельство о внесении в государственный реестр прав на объекты,
охраняемые авторским правом

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  РЕСПУБЛИКА КАЗАХСТАН

СВИДЕТЕЛЬСТВО
О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР
ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ
№ 11321 от «10» июля 2020 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):
КАЗАНБАЕВА АЛБИНА СОВЕТОВНА, КУЛИКОВ ВЛАДИМИР ПАВЛОВИЧ, БЕРТАГИН АРМАН
САРСЕМБАЕВИЧ

Вид объекта авторского права: программа для ЭВМ

Название объекта: Модуль интеллектуальной системы оценки результатов обучения

Дата создания объекта: 09.07.2020





Копия свидетельства по адресу: <https://www.kazpatent.kz> и сайтам
"Авторское право" и "Патент" по адресу: <https://copyright.kazpatent.kz>
Подлинность документа возможно проверить на сайте [kazpatent.kz](https://www.kazpatent.kz)
в разделе «Авторское право» по адресу: <https://copyright.kazpatent.kz>

Подписано ЭЦП Куантыров Е.С.

ПРИЛОЖЕНИЕ Б

Акт внедрения модуля интеллектуальной системы оценки результатов обучения

УТВЕРЖДАЮ
Директор КТ КП Петропавловского
строительно-экономического
колледжа
_____ А.Б. Кушумбаев
« ____ » _____ 2021 г.

АКТ

о практическом использовании результатов диссертационной работы
соискателя степени PhD, СКУ имени М. Козыбаева
Казанбаевой А.С. на тему: «Разработка интеллектуальной системы оценки
результатов обучения»

Настоящий акт составлен в подтверждении того, что результаты теоретических и экспериментальных исследований, приведенные в диссертационной работе А. Казанбаевой применялись при оценке результатов обучения, обучающихся на 3-4 курсе по специальностям «Вычислительная техника и программное обеспечение» и «Строительство и эксплуатация зданий и сооружений» Петропавловского строительно-экономического колледжа. Апробация данного исследования показала, что модуль ИСОРО предназначен для автоматизации оценивания результатов устного ответа, что позволило применить его отдельно при проверки устных/письменных работ, а также в комплексе, как часть системы.

Работоспособность схемы данного исследования никак не хуже тех возможностей, что законодательно доступны классической схемы линейной оценки результатов обучения.

Методист _____

_____ Е.А. Калганова

ПРИЛОЖЕНИЕ В

Скриншоты сообщений о возникших проблемах в модуле ИСОРО

