

На правах рукописи

Проскурнин Алексей Александрович

**ПРЕДСТАВЛЕНИЕ, ОБРАБОТКА И КОНТРОЛЬ ЗНАНИЙ
НА ОСНОВЕ ЭЛЕМЕНТАРНЫХ ВЕРБАЛЬНЫХ ЕДИНИЦ**

05.13.17 – Теоретические основы информатики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Москва – 2010

Работа выполнена на кафедре «Медиасистемы и технологии»
ГОУ ВПО «Московский государственный университет печати»

Научный руководитель: кандидат технических наук, доцент
Филиппович Юрий Николаевич

Официальные оппоненты: доктор технических наук, профессор
Попов Дмитрий Иванович
кандидат технических наук, доцент
Сазонов Борис Алексеевич

Ведущая организация: Институт системного анализа РАН

Защита диссертации состоится 23 декабря 2010 г. в 12 часов 30 минут на заседании диссертационного совета Д 212.147.03 при Московском государственном университете печати (127550 Москва, ул. Прянишникова, 2А).

С диссертацией можно ознакомиться в библиотеке Московского государственного университета печати.

Автореферат разослан «_____» ноября 2010 г.

Учёный секретарь
диссертационного совета
д. т. н., профессор

Агеев В.Н.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования.

Современный этап развития науки, образования и народного хозяйства характеризуется разработкой и использованием информационных технологий, основанных на знаниях, на базе средств вычислительной техники, применением методов искусственного интеллекта при создании информационных систем. Возник и развивается широкий класс информационных систем — автоматизированные системы контроля знаний (АСКЗ).

В работе к классу АСКЗ отнесены системы, основная функция которых состоит в формировании модели персональных знаний (МПЗ) пользователя в процессе взаимодействия человека и компьютера — автоматизированного контроля знаний (АКЗ).

В большом количестве работ в области построения систем контроля знаний и значительном разнообразии предлагаемых подходов, можно выделить две базовые парадигмы их создания: парадигму оценки (измерения) знаний и парадигму выявления и анализа знаний.

В рамках первой парадигмы МПЗ представляет собой скаляр (так называемая «скалярная модель обучаемого»). В этом случае основной целью контроля знаний является их оценка, формирование некоторой меры (рейтинга, ранга), характеризующей знания человека, их носителя. Системы, построенные на основе этой парадигмы, наиболее часто относят к классу тестирующих, или тестового контроля знаний. Они дают человеку формальную оценку его знаний, т.е. отвечают на вопрос «на сколько?» он знает тот или иной предмет.

Вторая парадигма подразумевает, что МПЗ — это модель данных, для которых характерна внутренняя интерпретируемость, структурированность, связность, иерархичность информационных единиц. Такие системы дают содержательное описание знаний человека, отвечают на вопрос «что он знает или не знает?» о предмете.

Большой вклад в развитие методов формирования оценки знаний, относящихся, прежде всего, к теории педагогических измерений, внесли В.С. Аванесов, А. Бирнбаум, Ф. Лорд, А.Н. Майоров, В.И. Михеев, Ю.М. Нейман, М. Новик, Г. Раш, А.О. Татур, М.Б. Чельшкова, А.О. Шмелев и другие ученые.

Методы создания интеллектуальных систем обучения и контроля знаний исследовали многие отечественные и зарубежные ученые: С.В. Астанин, П. Брусиловский, А.И. Башмаков, И.А. Башмаков, В.И. Васильев, Т.А. Гаврилова, Д.О. Жуков, Дж. Карбонелл, В.М. Курейчик, В.А. Петрушин, Д.И. Попов, Л.А. Растринин, И.Д. Рудинский, Г.В. Рыбина, В.Л. Стефанюк, Т.А. Таран.

В разработку методов создания систем, основанных на знаниях, большой вклад внесли В.Н. Вагин, А.П. Еремеев, А.С. Нариньяни, Г.С. Осипов, Э.В. Попов, Г.С. Поспелов, Д.А. Поспелов, В.К. Финн, В.Ф. Хорошевский и другие.

Подходы к созданию АСКЗ в парадигме выявления и анализа знаний, на сегодняшний день, исследованы в гораздо меньшей степени, т.к. в большинстве работ наивысший приоритет отдается решению задачи формирования интегральной оценки знаний.

Предлагаемый в диссертации подход к решению задачи автоматизированного контроля знаний ориентирован на содержательное описание и анализ персональных знаний обучаемого о предметной области.

Актуальность этого подхода обуславливается необходимостью внедрения инновационных образовательных технологий, основанных на активных и интерактивных формах проведения занятий и контроля знаний обучающихся, использования в процессе обучения разнообразных источников информации. Реализация подхода ориентирована также и на устранение следующих характерных недостатков в уже существующих АСКЗ, предназначенных для формирования содержательного описания знаний обучаемого:

1. Использование методов АКЗ, максимально учитывающих специфику конкретной предметной области, и неприменимых для других областей.

2. Моделирование в АСКЗ учебного курса (предмета), а не предметной области как таковой.

3. Поверхностность, низкий уровень детализации описания знаний обучаемого в АСКЗ.

4. Необходимость наличия достаточно больших репрезентативных выборок обучаемых.

5. С одной стороны, закрытая форма заданий обладает существенными недостатками; с другой стороны, для открытой формы заданий, подразумевающей анализ свободного естественно-языкового ответа обучаемого в виде предложения, крайне трудно гарантировать высокую адекватность такого анализа.

6. Применение в АСКЗ таких методов формирования МПЗ, при которых принципиально невозможно пояснить пользователю ход рассуждений, с помощью которых была получена модель его знаний.

7. Отсутствие в АСКЗ инструментов анализа знаний обучаемого, выявленных системой.

8. Ориентированность АСКЗ на один сеанс взаимодействия с обучаемым, после которого ему/ей выдаются результаты, а не на многие сеансы, подразумевающие «диалоги» с системой в течение продолжительного периода времени, когда с каждым сеансом взаимодействия пополняется (уточняется) МПЗ обучаемого.

Цель и задачи исследования.

Целью диссертационного исследования является разработка и практическая реализация в виде программной системы новой методики АКЗ, ориентированной на решение задачи формирования содержательного, структурного описания персональных знаний обучаемого, с учетом устранения описанных выше недостатков в существующих подходах к решению этой задачи.

Для достижения этой цели в диссертационной работе решены следующие основные задачи:

1. Проведены анализ и классификация существующих методов и моделей АКЗ, а также моделей представления знаний и моделей рассуждений, используемых для решения задачи АКЗ.

2. Разработаны математические модели описания экспертных знаний и знаний обучаемого о предметной области для решения задачи АКЗ.

3. Разработаны алгоритмы АКЗ, основанные на использовании различных способов вывода на знаниях.

4. На основе предложенных моделей и алгоритмов разработана методика АКЗ.

5. Разработана и апробирована программная система, реализующая предлагаемую методику АКЗ.

6. Проведены экспериментальные исследования методики АКЗ.

Объектом исследования в работе являются модели представления знаний и модели рассуждений.

Предметом исследования являются способы применения моделей представления знаний и моделей рассуждений в АСКЗ, которые ориентированы на решение задачи формирования содержательного, структурного описания персональных знаний обучаемого, исходя из взаимодействия с ним.

Научная новизна исследования.

Научно новыми являются следующие результаты, полученные в диссертационном исследовании:

1. Разработана новая теоретико-множественная модель описания экспертных знаний о предметной области для решения задачи АКЗ, отличающаяся от известных использованием элементарных единиц вербального языкового знания.

2. Разработана новая теоретико-множественная модель описания персональных знаний обучаемого о предметной области для решения задачи АКЗ, отличающаяся от известных использованием целевых единиц знания, а также выделением трех уровней описания, соответствующих трем различным типам анализа персональных знаний обучаемого – структурному анализу, логическому анализу, и анализу соответствия эталонным образам знания.

3. Разработаны новые алгоритмы АКЗ, позволяющие формировать описание персональных знаний обучаемого, исходя из взаимодействия с ним, и основанные на использовании различных способов вывода. Отличительной особенностью этих алгоритмов является обработка знаний, формальное описание которых дано в соответствии с предлагаемыми теоретико-множественными моделями, а также применение четырехзначной логики аргументации.

4. На основе предлагаемых моделей и алгоритмов разработана новая методика АКЗ.

Методы исследования.

В работе использованы теория множеств и отношений, методы искусственного интеллекта, аппарат нечетких множеств и нечеткого логического вывода, методы исследования операций (решение задач многокритериальной оптимизации). При разработке программного продукта использовались объектно-ориентированный и реляционный подходы к построению программных систем.

Основные положения, выносимые на защиту.

1. Теоретико-множественные модели описания экспертных знаний и персональных знаний обучаемого о предметной области для решения задачи АКЗ.

2. Алгоритмы АКЗ, основанные на использовании различных способов вывода на знаниях.

3. Комплекс дополнительных моделей и алгоритмов, которые совместно с другими составляют основу для реализации методики АКЗ.

4. Методика АКЗ, основанная на предлагаемых моделях и алгоритмах.

Обоснованность и достоверность научных положений, рекомендаций и выводов.

Обоснованность и достоверность научных положений, рекомендаций и выводов определяется корректным применением соответствующих методов исследования, а также подтверждается результатами экспериментальных исследований предлагаемой методики АКЗ. Предложенные определения понятий и классификации апробированы на конференциях и в научных публикациях.

Практическая ценность работы.

Разработанный программный продукт, который является оболочкой системы, основанной на знаниях, позволяет создавать АСКЗ в различных предметных областях путем «наполнения» базы знаний системы. Разработана база знаний предметной области «Основы информатики», и АСКЗ на основе этой базы знаний. Материалы проведенного исследования были использованы в учебном процессе кафедры «Автоматизированные системы обработки информации и управления» МГТУ им. Баумана, в рамках дисциплин «Семиотика информационных технологий» и «Компьютерные технологии в науке и образовании».

Апробация работы.

Основные результаты диссертационной работы докладывались и были одобрены на заседаниях комиссий по аттестации аспирантов кафедры «Автоматизированные системы обработки информации и управления» МГТУ им. Н.Э. Баумана в 2005-2008 гг. Также материалы работы были представлены на следующих научных конференциях и семинарах: Общероссийской научно-технической конференции «Новые технологии в азотной промышленности» (г. Невинномысск, 2003); I Международной научно-практической конференции «Молодежь и наука: реальность и будущее» (г. Невинномысск, 2008); Научно-методическом семинаре НОК CLAIM, на кафедре «Автоматизированные системы обработки информации и управления» МГТУ им. Н.Э. Баумана (г. Москва, 2007, 2009); Научной школе для молодых ученых «Компьютерная графика и математическое моделирование (Visual Computing)» (г. Москва, 2009); III школе-семинаре «Задачи системного анализа, управления и обработки информации» (г. Москва, 2009); I Всероссийской дистанционной научно-практической конференции «Инновации в образовательных системах» (г. Челябинск, 2009); Научной межвузовской конференции преподавателей, аспирантов, молодых ученых и специалистов «Печатные средства информации в современном обществе» (г. Москва, 2010).

Публикации по теме диссертации.

Основные результаты по теме диссертации опубликованы в 11-ти печатных работах, в том числе 1 – в журнале, включенном в Перечень ВАК РФ.

Объем и структура работы.

Диссертация состоит из списка основных сокращений, введения, четырех глав, заключения, списка использованных источников из 294 наименований и 9 приложений. Основной текст изложен на 197 страницах, включая рисунки и таблицы. Приложения выполнены на 46 страницах.

СОДЕРЖАНИЕ РАБОТЫ

Во введении к диссертации обосновывается актуальность работы, определена цель и поставлены задачи исследования, сформулирован перечень основных положений, выносимых на защиту, научная новизна проведенного исследования, практическая ценность результатов работы.

Первая глава диссертации посвящена обзору существующих методов и моделей АКЗ. Дана общая постановка задачи АКЗ, описаны классификации методов решения этой задачи по различным критериям, рассматриваются модели представления знаний и модели рассуждений в задачах АКЗ.

Проведенный обзор работ позволяет выделить основные направления интеллектуализации АСКЗ: семантический анализ развернутых естественно-языковых ответов обучаемого на вопросы системы; «динамическая» генерация заданий для контроля знаний, в том числе вопросов на естественном языке; реализация различных механизмов адаптивного управления процессом контроля знаний на основе построения модели обучаемого, и ряд других.

Спектр используемых в АСКЗ способов представления знаний очень широк: применяются как логические, так и эвристические модели, включая фреймы, продукции, семантические сети; ряд работ использует конкретные специфические формализмы для моделирования предметной области, например, сети Петри, метод анализа формальных понятий (Formal Concept Analysis), теорию пространств знаний (Knowledge Space Theory) и другие; находят применение тезаурусный подход, в частности, тезаурус семантических полей, и онтологии предметной области, в том числе онтологии в Internet, в рамках концепции Semantic Web. Вывод на знаниях в существующих интеллектуальных АСКЗ реализуется на основе классических продукционных систем, байесовского подхода, нечеткой логики, аргументации, рассуждений по аналогии, рассуждений на основе прецедентов и других подходов.

Несмотря на значительное разнообразие используемых в АСКЗ способов представления и обработки знаний, в основном они применяются в рамках парадигмы оценки (измерения) знаний.

Во второй главе рассматриваются модели представления знаний и алгоритмы их обработки, которые предлагаются в работе для решения задачи АКЗ.

В качестве основы представления знаний о предметной области выбрана модель, предложенная Ю.Н. Карауловым, компонентами которой являются: фигуры знания — элементарные вербальные единицы; концепты и концептосферы — обобщенные понятия, принятые в модели для обозначения предметных областей и их частей.

Фигура знания является пятикомпонентной структурой. Ее интенционал, т.е. сущностное содержание, образуют три следующих компонента: знак — это слово или словосочетание, обозначающее некоторый предмет, формула смысла — естественно-языковое выражение (например, предложение), задающее смысл знака, и способ задания смысла — речевой оборот (стилистическая фигура), определяемый формулой смысла и знаком. Помимо указанных трех интенциональных компонентов, в фигуру знания входят еще два компонента, опреде-

ляющих ее экстенционал – это когнитивная (референтная) область, к которой относится фигура знания, а также функция, определяющая ценность (вес) данной элементарной единицы знания.

В диссертации, в описании фигур знания определенной предметной области, для каждой фигуры выделяются основные бинарные связи между ее знаком и компонентами формулы смысла — вербальными единицами (основными словами и/или словосочетаниями).

В таблице 1 даны примеры фигур знания в предметной области «Основы информатики» с выделенными для каждой фигуры знания связями.

Предлагаемая в диссертационной работе модель представления знаний о предметной области основывается на идее сочетания знаний о мире и знаний о языке. Несколько упрощая и огрубляя модель, можно сказать, что компьютер «знает» о том, какие понятия и имена существуют в предметной области, «знает» определенные отношения, которые существуют между этими понятиями и именами, а также «знает», каким образом определенные аспекты смысла тех или иных понятий или имен, задаваемые некоторой совокупностью их связей с другими понятиями и именами, могут быть выражены в виде кратких естественно-языковых формулировок с использованием определенных способов задания смысла.

Таблица 1. Примеры фигур знания

Знак	Способ задания смысла	Формула смысла	Когнитивная Область	Функция знания	Связи
Информация	Дескрипция	Она передается от источника к приемнику в виде сигналов.	Основные понятия	Рецепт	Информация--- Источник Информация--- Приемник Информация---Сигнал
Процессор	Метафора	Мозг компьютера	Устройство компьютера	Рецепт	Процессор--- Компьютер
Бит	Дескрипция	Единица измерения информации	Единицы измерения	Рецепт	Бит--- Информация

Для проверки знаний человека в АСКЗ ему задаются вопросы. В формулировке каждого из вопросов содержится формула смысла некоторой фигуры знания. Человек должен понять формулировку, т.е. определить, какое понятие или имя она задает, и ответить, набрав на клавиатуре это понятие или имя. Ответ на вопрос оценивается в дихотомической шкале («верно», «неверно»).

В диссертационной работе предлагается следующая идея контроля знаний. Если человек отвечает на вопрос правильно, т.е. определяет знак фигуры

знания по ее формуле смысла, можно предположить, что в его персональное знание о предметной области входят определенные бинарные связи между знаком, который является правильным ответом на вопрос, и определенными вербальными единицами, входящими в формулировку вопроса. Аналогично, если человек не смог определить правильно знак фигуры знания по ее формуле смысла, можно предположить отсутствие в его знаниях о предметной области определенных бинарных связей.

Например, пусть человеку задается вопрос, формулировка которого содержит формулу смысла «Популярный язык программирования, разработанный Никлаусом Виртом», включающую вербальные единицы «язык программирования» и «Никлаус Вирт». Если человек дает правильный ответ «Паскаль», фиксируется аргумент в пользу знания связей «язык программирования - Паскаль» и «Никлаус Вирт - Паскаль». Если же человек не дает никакого ответа или отвечает неправильно, фиксируется аргумент против знания связи «Никлаус Вирт - Паскаль», так как именно эта связь является ключевой с точки зрения определения правильного ответа на вопрос.

Кроме фигур знания, для моделирования предметной области используются две другие единицы знания – концепты и так называемые целевые единицы знания (ЦЕЗ). Концептом является имя или понятие, которое может как принадлежать, так и не принадлежать предметной области. ЦЕЗ в общем случае описывает определенный аспект смысла связи между двумя концептами, при этом один из двух концептов обязательно должен принадлежать предметной области. ЦЕЗ является той бинарной связью, присутствие или отсутствие которой в персональном знании человека можно предполагать исходя из его ответов на вопросы АСКЗ, с поправкой на то, что ЦЕЗ в общем случае может выражать не только смысл этой связи как таковой, а определенный аспект смысла связи.

Над множеством ЦЕЗ задается любое число классификаций (декомпозиций) предметной области, каждая из которых представляет собой дерево, т.е. устанавливаются связи между элементами множества листьев каждого из деревьев и элементами множества ЦЕЗ.

Необходимо отметить, что для большинства бинарных связей между концептами выделять несколько аспектов понимания их смысла не нужно. Например, связи «массив-структура данных» или «системный блок-компьютер» одноаспектные и их смысл определяется соответствующими типами связей «is a» и «part of». Тем не менее, другие связи более сложные, поэтому возможно отдельно учитывать различные аспекты смысла таких сложных связей. Например, для связи «процессор-компьютер» могут быть выделены следующие аспекты: «процессор – часть компьютера» и «процессор – главное управляющее устройство компьютера», т.к. возможно понимание человеком первого из указанных аспектов и при этом непонимание второго.

Множество ЦЕЗ можно представить в виде графа, вершинам которого соответствуют концепты, а ребрам – ЦЕЗ. Такой граф, в общем случае, допускает кратные ребра, т.к. возможно для одной и той же пары концептов выделение различных аспектов смысла бинарной связи между этими концептами. Граф не обязательно должен быть связным; число компонент связности в нем зависит от

степени семантической близости различных подобластей внутри моделируемой предметной области.

При описании знаний о предметной области выделяется 3 уровня этого описания: онтологический уровень, вербальный уровень, и таксономический уровень, при этом задаются связи между первым из них и двумя другими:

$$M_{\text{ПР.ОБЛ.}} = \langle M_{\text{ОНТ.}}, M_{\text{ВЕРБ.}}, M_{\text{ТАКС.}}, ML_{\text{ОНТ.-ВЕРБ.}}, ML_{\text{ОНТ.-ТАКС.}} \rangle.$$

Входящие в $M_{\text{ПР.ОБЛ.}}$ компоненты задаются следующим образом:

$$M_{\text{ОНТ.}} = \langle T, E, A, L, f_{ET}, f_{LEEA}, H_L \rangle \quad M_{\text{ВЕРБ.}} = \langle Q, P, V, Z, W, f_{QP}, R_{QZW} \rangle$$

$$M_{\text{ТАКС.}} = \langle D, C, f_{CD}, R_{CC}, (\tilde{R}_d^{3H.})_{d \in D}, (\tilde{R}_d^{HE3H.})_{d \in D}, (H_{Cd})_{d \in D}, I, G \rangle$$

$$ML_{\text{ОНТ.-ВЕРБ.}} = \langle f^{3H.}, f^{HE3H.}, f_{QE} \rangle \quad ML_{\text{ОНТ.-ТАКС.}} = \langle (R_{CLd})_{d \in D}, (f_{EGi})_{i \in I} \rangle$$

Описание компонентов представлено ниже.

Онтологический уровень:

T – множество типов концептов; E – множество концептов; A – множество аспектов; L – множество ЦЕЗ; $f_{ET} : E \rightarrow T$ – отображение, задающее каждому концепту его тип; $f_{LEEA} : L \rightarrow E \times E \times A$ – инъективное отображение, задающее смысл ЦЕЗ; H_L – множество логических выражений (формул), в которых высказывания вида «человек знает ЦЕЗ $l \in L$ » могут объединяться с помощью логических операций И, ИЛИ, а также скобок, задающих приоритет выполнения этих операций.

Вербальный уровень:

Q – множество фигур знания (множество вопросов); P – множество свойств (атрибутов) фигур знания; V – множество значений атрибутов фигур знания; Z – множество знаков; W – множество словоформ знаков; $f_{QP} : (Q \times P) \rightarrow V$ – отображение, определяющее значения атрибутов фигур знания; $R_{QZW} \subseteq Q \times Z \times W$ – отношение, задающее для каждой фигуры знания множество знаков, и для каждого из этих знаков – множество словоформ.

Таксономический уровень:

D – множество классификаций (декомпозиций) предметной области; C – множество классификационных единиц (КЕ); $f_{CD} : C \rightarrow D$ – отображение, задающее для каждой КЕ определенную классификацию; $R_{CC} \subseteq C \times C$ – отношение «родительская КЕ \leftrightarrow дочерняя КЕ», определяющее иерархическую структуру каждой классификации. При этом, если $(c_i, c_j) \in R_{CC}$, то $f_{CD}(c_i) = f_{CD}(c_j)$; также для $\forall c \in C : (c, c) \in R_{CC}$.

$(\tilde{R}_d^{3H.})_{d \in D}$ представляет собой семейство бинарных нечетких отношений: для каждой классификации $d \in D$ определяется бинарное нечеткое отношение $\tilde{R}_d^{3H.}$, которому соответствует функция принадлежности, задаваемая выражением $\mu_d^{3H.} : (C_d^{\text{нижн.}} \times C_d^{\text{нижн.}}) \rightarrow [0, 1]$, где $C_d^{\text{нижн.}}$ – множество КЕ самого «нижнего» уровня классификации $d \in D$:

$$C_d^{\text{нижн.}} = \{c \in C : f_{CD}(c) = d \wedge (\neg \exists c' \in C : (c, c') \in R_{CC})\}.$$

Таким образом, с помощью отношения \tilde{R}_d^{3H} для КЕ нижнего уровня классификации $d \in D$ задаются нечеткие правила вида «ЕСЛИ человек знает КЕ А, ТО он также знает КЕ В». Семейство нечетких бинарных отношений $(\tilde{R}_d^{HE3H})_{d \in D}$ имеет аналогичный смысл, но при этом задает нечеткие правила вида «ЕСЛИ человек НЕ знает КЕ А, ТО он также НЕ знает КЕ В».

$(H_{Cd})_{d \in D}$ – это семейство множеств: для каждой классификации $d \in D$ определяется множество логических выражений (формул) H_{Cd} , в которых высказывания вида «человек знает КЕ $c \in C(f_{CD}(c) = d)$ » могут объединяться с помощью логических операций И, ИЛИ, а также скобок, задающих приоритет выполнения этих операций.

I – множество эталонных образов знания; каждый из этих образов задается как совокупность требований к знанию ЦЕЗ, и может отражать как определенный уровень, так и характер, направленность знания; G – множество возможных значений уровня требований к знанию определенной ЦЕЗ; $G = \{\text{«знание строго обязательно»}, \text{«знание критично»}, \text{«знание желательно»}, \text{«знание опционально»}, \text{«знание не нужно»}\}$.

Связь онтологического и вербального уровня:

Функции f^{3H} и f^{HE3H} задают продукции следующего вида: если задан некоторый вопрос, и получен правильный ответ, зафиксировать знание для определенных ЦЕЗ; аналогично, если задан вопрос, и получен неверный ответ, зафиксировать незнание определенных ЦЕЗ:

$$f^{3H} : Q \rightarrow 2^L, f^{HE3H} : Q \rightarrow 2^L, \forall q \in Q : f^{HE3H}(q) \subseteq f^{3H}(q).$$

$f_{QE} : Q \rightarrow E$ – функция, которая для каждой фигуры знания задает концепт, соответствующий ее знаку (множеству знаков-синонимов).

Связь онтологического и таксономического уровня:

Семейство бинарных отношений $(R_{CLd})_{d \in D}$ для каждой классификации $d \in D$ задает бинарное отношение $R_{CLd} \subseteq C_d^{\text{нижн}} \times L$, где $C_d^{\text{нижн}}$ – множество КЕ самого «нижнего» уровня этой классификации. Т.е. для каждой КЕ «нижнего» уровня задается множество входящих в нее ЦЕЗ, при этом одна и та же ЦЕЗ может относиться к нескольким КЕ.

Семейство функций $(f_{EGi})_{i \in I}$ для каждого эталонного образа знаний $i \in I$ задает совокупность требований этого образа к знанию ЦЕЗ, а именно, значение уровня требований $g \in G$ для каждой из ЦЕЗ: $f_{EGi} : L \rightarrow G$.

Модель персональных знаний (МПЗ) обучаемого задается следующим выражением:

$$M_{ПЗ} = \langle M_{ПР.ОБЛ.}, S_Q, S_L, S_C, f_Q, f_L, f_C, f_{HL}, f_{HC}, f_I, f_{DC}, f_{DHC}, f_{DI} \rangle$$

Здесь: $S_Q = \{\text{«верный ответ»}, \text{«неверный ответ»}, \text{«не задан»}\}$ – множество значений (статусов), одно из которых ставится в соответствие каждому из вопросов; значения определяются с помощью анализа ответов испытуемого на вопросы; $S_L = \{\text{«знание»}, \text{«незнание»}, \text{«неизвестно»}, \text{«противоречие»}\}$ – мно-

жество статусов ЦЕЗ; $S_C = \{\text{«абсолютное незнание»}, \text{«определенно, незнание»}, \text{«скорее всего, незнание»}, \text{«неопределенность»}, \text{«скорее всего, знание»}, \text{«определенно, знание»}, \text{«абсолютное знание»}\}$ – множество статусов КЕ; $f_Q : Q \rightarrow S_Q$ – функция, которая задает текущие статусы вопросов; $f_L : L \rightarrow S_L$ – функция, которая задает текущие статусы ЦЕЗ; $f_C : C \rightarrow S_C$ – функция, которая задает текущие статусы КЕ; $f_{HL} : H_L \rightarrow S_L$ – функция, определяющая значения логических выражений (формул) из H_L ; значения вычисляются на основе выполнения операций И, ИЛИ в рамках четырехзначной логики; $f_{HC} : \bigcup_{d \in D} (H_{Cd}) \rightarrow S_C$ – функция, определяющая значения логических выражений (формул) из множеств H_{Cd} ($d \in D$); значения вычисляются на основе выполнения операций И, ИЛИ в рамках семизначной логики; $f_I : I \rightarrow [0, 1]$ – функция, задающая степень соответствия знаний испытуемого каждому из эталонных образов знания $i \in I$; $f_{DC} : C \rightarrow [0, 1]$ – функция, задающая степени достоверности текущих определенных статусов КЕ; $f_{DHC} : \bigcup_{d \in D} (H_{Cd}) \rightarrow [0, 1]$ – функция, задающая степени достоверности текущих определенных значений логических выражений (формул) из множеств H_{Cd} ($d \in D$); $f_{DI} : I \rightarrow [0, 1]$ – функция, задающая степени достоверности текущих определенных значений степеней соответствия знаний испытуемого каждому из эталонных образов знания $i \in I$.

Выражение для $M_{ПЗ}$ может быть переписано следующим образом:

$$M_{ПЗ} = \langle M_{ПР.ОБЛ.}, M_Q, M_L, M_C, M_H, M_I \rangle, \text{ где}$$

$M_Q = \langle S_Q, f_Q \rangle$ – МПЗ на уровне вопросов; $M_L = \langle S_L, f_L \rangle$ – МПЗ на уровне ЦЕЗ; $M_C = \langle S_C, f_C, f_{DC} \rangle$ – МПЗ на уровне КЕ; $M_H = \langle f_{HL}, f_{HC}, f_{DHC} \rangle$ – МПЗ на уровне логических критериев знания; $M_I = \langle f_I, f_{DI} \rangle$ – МПЗ на уровне эталонных образов знания. Уровни представления МПЗ и преобразования между ними схематично изображены на рис. 1.

Таким образом, всего предлагается 5 уровней представления персональных знаний. Уровни 1 и 2 являются основными, базовыми – это само выявленное знание. Уровни 3, 4 и 5 интерпретируются как инструменты анализа выявленных знаний.

Исходя из преобразований уровней 2→3, 2→4, и 2→5, можно выделить 3 типа анализа персональных знаний: 1) Структурный анализ основан на идее аналогии с интеллектуальным многомерным анализом данных (технологией OLAP); 2) Логический анализ основан на вычислении логических критериев знания с помощью использования логических связей И, ИЛИ для многозначной логики; 3) Анализ соответствия эталонным образам знания позволяет вычислять степень соответствия текущих выявленных знаний пользователя АСКЗ некоторым предварительно заданным образам знания, которые могут отражать как определенный уровень знаний, так и их характер (направленность).

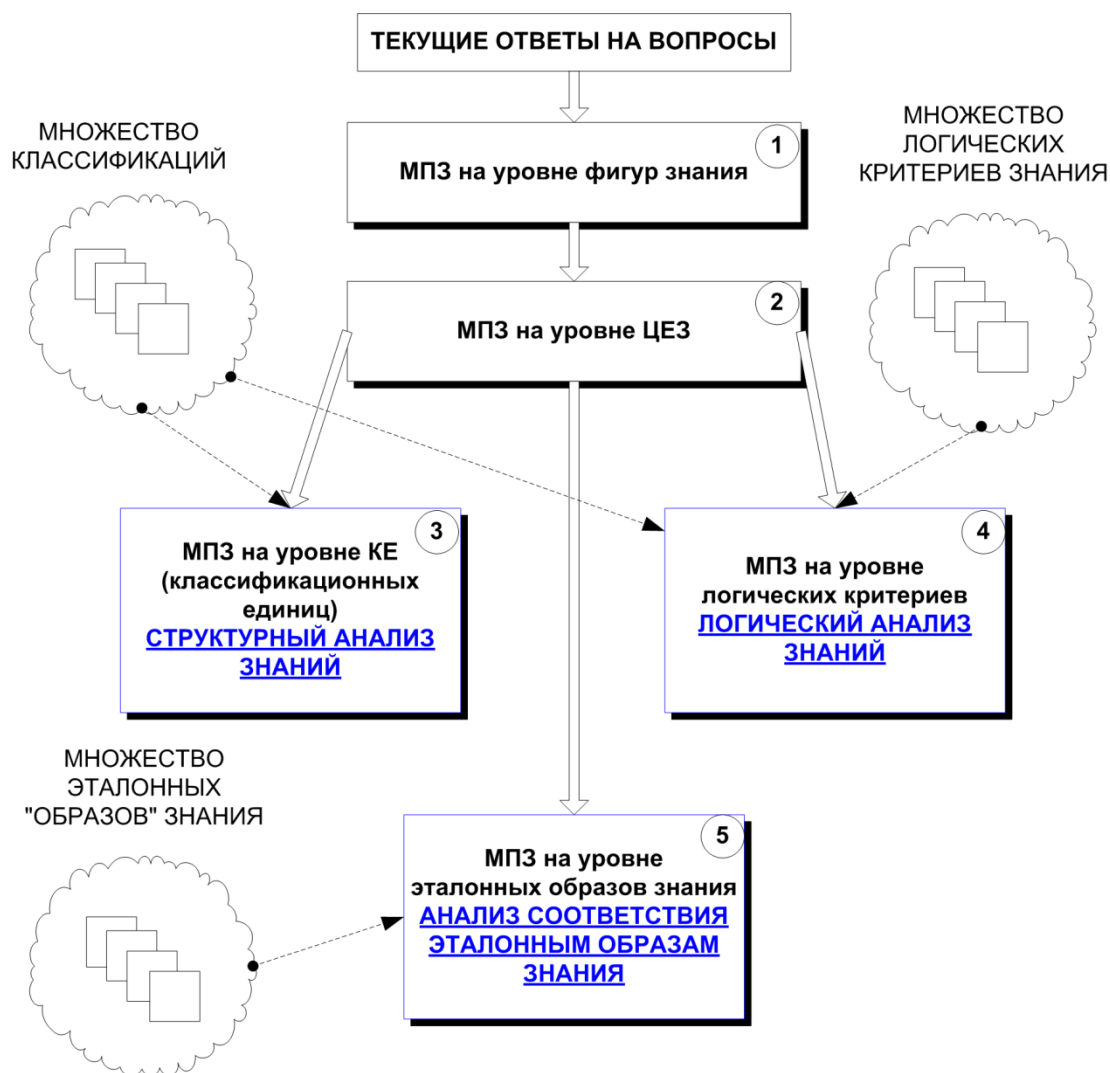


Рис. 1. Уровни представления МПЗ и преобразования между ними

Алгоритмы АКЗ, формирующие МПЗ обучаемого, основаны на моделировании следующих рассуждений человека-эксперта, осуществляющего контроль знаний: 1) Рассуждения, связанные с формированием суждений о присутствии в сознании человека определенных когнитивных единиц на основании информации, полученной от человека в процессе взаимодействия; 2) Рассуждения, связанные с формированием суждений о степени знания одних когнитивных единиц, исходя из степени знания других единиц, когда и те, и другие единицы находятся на одном уровне иерархии в структурной декомпозиции предметной области; 3) Рассуждения, связанные с формированием суждений о степени знания некоторого множества элементов, исходя из знания/незнания этих элементов (операция «свертки»); 4) Рассуждения, связанные с выбором следующего вопроса, который будет задан человеку в процессе адаптивного взаимодействия, исходя из текущей МПЗ человека и выбранной стратегии взаимодействия.

В основе моделирования рассуждений 1-го типа лежит применение семантики логики аргументации A_4 , предложенной В.К. Финном. Продукции, задаваемые функциями $f^{ЗН}$ и $f^{НЕЗН}$, интерпретируются как инструменты размещения соответствующих ЦЕЗ в два мультимножества: мультимножество

знания $m_{3H.} : L \rightarrow N$ (аргументы) и мультимножество незнания $m_{HE3H.} : L \rightarrow N$ (контраргументы), где N – множество целых неотрицательных чисел. Пусть $l \in L$, $m_{3H.}(l) = N_1$, $m_{HE3H.}(l) = N_2$. Тогда: если $N_1=0$ и $N_2=0$, то $f_L(l) =$ «неизвестно»; если $N_1>0$ и $N_2=0$, то $f_L(l) =$ «знание»; если $N_1=0$ и $N_2>0$, то $f_L(l) =$ «незнание»; если $N_1>0$ и $N_2>0$, то $f_L(l) =$ «противоречие».

Рассуждения 2-го типа моделируются как классический нечеткий логический вывод на основе применения композиционного правила вывода Заде: нечеткое отношение между входной и выходной переменными в данном случае – это либо $\tilde{R}_d^{3H.}$, либо $\tilde{R}_d^{HE3H.}$.

Рассуждения 3-го типа реализуются с помощью определенного варианта выполнения операции «свертки». В диссертационной работе предлагается 3 варианта реализации операции «свертки»: 1) на основе использования разности долей элементов со статусами «знание» и «незнание»; 2) на основе применения системы нечеткого логического вывода (СНЛВ) с трапециевидными функциями принадлежности; 3) на основе связки СНЛВ с «грубыми» функциями принадлежности и настраивающего эту систему генетического алгоритма (СНЛВ + ГА). При этом в качестве основы для реализации СНЛВ и связки СНЛВ+ГА был использован подход, предложенный А.П. Ротштейном.

Суть моделирования рассуждений 4-го типа состоит в следующем. Для каждого вопроса, входящего в текущее множество вопросов для выбора, вычисляются 2 критерия: $K_i^{3H.}$ – коэффициент соответствия вопроса текущей области знания, $K_i^{HE3H.}$ – коэффициент соответствия вопроса текущей области незнания. Также при отборе вопроса используется 3-ий критерий: W_i – сложность вопроса (он задается экспертом). Затем, в зависимости от выбранной стратегии взаимодействия, определяется постановка задачи многокритериального выбора вопроса: 1) Если Стратегия = «Максимальная лояльность», то: $K_i^{3H.} \rightarrow \max$, $K_i^{HE3H.} \rightarrow \min$, $W_i \rightarrow \min$; 2) Если Стратегия = «Максимальная строгость», то: $K_i^{3H.} \rightarrow \min$, $K_i^{HE3H.} \rightarrow \max$, $W_i \rightarrow \max$.

Задача многокритериального выбора решается следующим образом: 1) Построение множества парето-оптимальных векторов (множества Эджворта-Парето); 2) Выбор определенного вектора (т.е. вопроса) из множества парето-оптимальных векторов случайным образом, либо с помощью аддитивной свертки критериев с учетом весовых коэффициентов.

В диссертационной работе предлагаются следующие критерии эффективности адаптации для стратегий «Максимальная лояльность» и «Максимальная строгость»:

$$K_{ML} = \frac{N^{АДАПТ+} - N^{СЛ+}}{N}, \quad K_{МС} = \frac{N^{АДАПТ-} - N^{СЛ-}}{N}.$$

Здесь N – общее число вопросов, которое задается при одном запуске алгоритма; $N^{АДАПТ+}$ – среднее число правильных ответов, данных пользователем, когда вопросы задавались в режиме адаптации с использованием стратегии

«Максимальная лояльность»); $N^{СЛ+}$ – среднее число правильных ответов, данных пользователем, когда вопросы отбирались случайно; $N^{АДАПТ-}$ – среднее число неправильных ответов, данных пользователем, когда вопросы задавались в режиме адаптации с использованием стратегии «Максимальная строгость»; $N^{СЛ-}$ – среднее число неправильных ответов, данных пользователем, когда вопросы отбирались случайно; усреднение во всех формулах ведется по числу запусков алгоритма.

В третьей главе рассматриваются дополнительные модели и алгоритмы, предлагаемая методика АКЗ, и проведенные экспериментальные исследования методики.

Для реализации предлагаемого подхода в виде программной системы, были разработаны дополнительные модели и алгоритмы: 1) Алгоритм анализа ответа на вопрос с учетом опечаток – основан на вычислении взвешенного расстояния между строками с помощью алгоритма Вагнера-Фишера; 2) Модели отбора подмножеств вопросов; 3) Модель вычисления различных интегральных числовых характеристик базы знаний о предметной области; 4) Модель уменьшения количества ЦЕЗ, имеющих статус «противоречие», на основе решения задачи многокритериальной порядковой экспертной классификации.

Все вместе предлагаемые модели и алгоритмы составляют оригинальную методику АКЗ, которая на самом «верхнем» уровне состоит из 3 этапов: 1) Выбор предметной области; 2) Разработка АСКЗ для данной предметной области, включающая создание базы знаний и загрузку ее в базу данных программы (основная часть базы знаний описывается в файлах Microsoft Word); 3) Применение АСКЗ.

Экспериментальное исследование методики проходило в 2 этапа.

На 1-ом этапе был проведен эксперимент по накоплению фигур знания в предметной области «Основы информатики», разработанных различными субъектами. Затем, с использованием этих данных и других источников была разработана база знаний этой предметной области, некоторые данные о которой представлены в таблице 2.

Таблица 2. Некоторые параметры БЗ «Основы информатики»

№	Название параметра	Значение
1	Число фигур знания (вопросов)	250
2	Число концептов	431
3	Число ЦЕЗ	494
4	Число связей «Фигура знания – ЦЕЗ»	1146
5	Среднее число вопросов на одну ЦЕЗ	1,295

На 2-ом этапе с помощью разработанной базы знаний были проведены 2 экспериментальных исследования: исследование валидности выявления знаний; исследование эффективности механизма адаптации.

Эксперимент по подтверждению валидности выявления знаний основан на сопоставлении суждений о знаниях испытуемых, сформированных 3-мя экспертами и АСКЗ на основании одних и тех же исходных посылок – ответов испытуемых на вопросы.

Ниже, на рис. 2, представлен график, иллюстрирующий сравнение решений экспертов и АСКЗ для 30 испытуемых (решением является статус определенной КЕ, задаваемый в 7-уровневой «серой» порядковой шкале, соответствующей множеству S_C); решения АСКЗ формировались с помощью 3-го варианта реализации операции «свертки».



Рис 2. Сравнение решений системы (вариант 3) и экспертов

Коэффициент ранговой корреляции Кендалла между решениями АСКЗ и средним решением 3-х экспертов (в качестве среднего значения использовалась медиана) для 3-х вариантов реализации операции «свертки» имеет значения: 1) 0,885; 2) 0,855; 3) 0,900. Для каждого из указанных значений корреляция между двумя ранжировками статистически значима на уровне $\alpha = 0,01$.

Суть экспериментального исследования эффективности механизма адаптации состоит в вычислении значений критериев эффективности адаптации для «виртуальных» испытуемых, каждому из которых соответствует «виртуальная» конфигурация знаний, представляющая собой ответы на 250 вопросов. Для каждой такой конфигурации знаний можно имитировать диалог соответствующего ей «виртуального» субъекта с АСКЗ. Теоретически число всевозможных рассматриваемых конфигураций знаний равно 2^{250} , поэтому была использована специальная стратегия для генерации меньшего количества «виртуальных» конфигураций: 7352 для стратегии «максимальная лояльность» и 7397 для стратегии «максимальная строгость». Общее число запусков алгоритма адаптации, выполненных в ходе проведения эксперимента, составило 884940, а общее число ответов на вопросы, «полученных» от «виртуальных» субъектов, 22123500. Результаты эксперимента подтвердили эффективность адаптации для обеих стратегий адаптивного взаимодействия.

В четвертой главе описана программная реализация предлагаемой методики АКЗ. Рассмотрены архитектура программного продукта, его функции, модель базы данных, форматы файлов Microsoft Word, из которых происходит загрузка основной части базы знаний в базу данных программы.

В соответствии с наличием трех основных категорий пользователей АСКЗ, и необходимостью разделения доступа к данным и действиям для этих категорий пользователей, система реализована в виде трех автоматизированных рабочих мест: эксперта, обучаемого и администратора, каждое из которых является приложением-клиентом по отношению к серверу базы данных.

Основная идея разработанной модели базы данных (БД) состоит в введении сущности «Модель знаний» и хранении в БД ссылки на эту сущность для всех других сущностей, которым соответствуют компоненты модели знаний. Такой подход позволяет хранить в одной БД любое количество баз знаний (БЗ) по различным предметным областям, вести несколько «версий» одной БЗ, формировать новую БЗ на основе уже существующей. Таким образом, создается гибкая среда управления данными, снижается трудоемкость описания знаний. Всего БД содержит 37 таблиц и является частично денормализованной.

Для создания системы использовались среда разработки Delphi и СУБД Firebird, клиентская часть приложения состоит из 43 программных модулей.

Заключение содержит основные полученные результаты и выводы из выполненного в работе исследования.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Проведены анализ и классификация существующих методов и моделей АКЗ, а также моделей представления знаний и моделей рассуждений, используемых для решения задачи АКЗ.

2. Разработана новая теоретико-множественная модель описания экспертных знаний о предметной области для решения задачи АКЗ.

3. Разработана новая теоретико-множественная модель описания персональных знаний обучаемого о предметной области для решения задачи АКЗ.

4. Разработаны новые алгоритмы АКЗ, основанные на использовании различных способов вывода на знаниях.

5. На основе предлагаемых моделей и алгоритмов разработана новая методика АКЗ.

6. Разработана и апробирована программная система, реализующая предлагаемую методику АКЗ.

7. Проведены экспериментальные исследования предлагаемой методики АКЗ, которые подтвердили валидность выявления знаний и эффективность предложенного механизма адаптации.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

В журналах, входящих в Перечень изданий, рекомендованных ВАК РФ.

1. Проскурнин А.А., Филиппович Ю.Н. Методы представления, автоматизированного выявления и анализа декларативных знаний человека о предметной области в компьютерных системах контроля знаний // Проблемы полиграфии и издательского дела. – 2010. – № 2. – С. 71-78.

В других изданиях.

2. Проскурнин А.А. Компьютеризованное обучение химии и химической технологии // Новые технологии в азотной промышленности: Сб. трудов обще-

российской науч.-техн. конф. / Северо-Кавказский государственный технический университет. Ставрополь, 2003. – С. 69-71.

3. Проскурнин А.А. Реферативный обзор по теме «Интеллектуальные системы тестирования и контроля знаний» // Интеллектуальные технологии и системы. Сборник учебно-методических работ и статей аспирантов и студентов. Выпуск 6. / Сост. и ред. Ю.Н. Филипповича. – М.: Изд-во ООО «Эликс+», 2004. – С. 251-259.

4. Проскурнин А.А. Интеллектуальные системы тестирования и контроля знаний // Вестник информационных технологий в образовании. Сборник учебно-методических и научных работ. Выпуск 1. – М.: УМК по специальности ИТО, 2005. – С. 152-160.

5. Проскурнин А.А. Математические модели оценки знаний // Интеллектуальные технологии и системы. Сборник учебно-методических работ и статей аспирантов и студентов. Выпуск 7. / Сост. и ред. Ю.Н. Филипповича. – М.: Изд-во ООО «Эликс+», 2005. – С. 197-210.

6. Проскурнин А.А. Автоматизированная система контроля знаний // Интеллектуальные технологии и системы. Сборник учебно-методических работ и статей аспирантов и студентов. Выпуск 8. / Сост. и ред. Ю.Н. Филипповича. – М.: НОК «CLAIM», 2006. – С. 200-208.

7. Проскурнин А.А. Автоматизированное извлечение декларативных знаний субъекта на основе когнитивного тезауруса // Интеллектуальные технологии и системы. Сборник учебно-методических работ и статей аспирантов и студентов. Выпуск 9. / Сост. и ред. Ю.Н. Филипповича. – М.: НОК «CLAIM», 2007. – С. 225-238.

8. Проскурнин А.А. Автоматизированное выявление и анализ декларативных знаний человека в интеллектуальных системах поддержки процессов обучения // Молодежь и наука: реальность и будущее: Материалы I Международной научно-практической конференции (Невинномысск, 26 февраля 2008). Сборник в 3 томах. Том 2 - Невинномысск: НИЭУП, 2008. – С. 282-285.

9. Филиппович Ю.Н., Проскурнин А.А. Структурный анализ персональных знаний с использованием системы нечеткого логического вывода // Научная школа для молодых ученых «Компьютерная графика и математическое моделирование (Visual Computing)»: тезисы и доклады. – Москва, 2009. – С. 162-178.

10. Проскурнин А.А. Интеллектуальная система автоматизированного выявления и анализа декларативных знаний обучаемого // Инновации в образовательных системах: Материалы I Всероссийской дистанционной научно-практической конференции (25 декабря 2009 г.) / Под ред. И.В. Резанович. – Челябинск, Издательский центр ЮУрГУ, 2009. – С. 12-16.

11. Проскурнин А.А. Отбор подмножеств заданий в системе контроля знаний на основе решения задачи линейного программирования с булевыми переменными // Материалы научной межвузовской конференции преподавателей, аспирантов, молодых ученых и специалистов «Печатные средства информации в современном обществе (к 80-летию МГУП)». Секция «Электронные средства информации в современном обществе». Сб. тез. докл. – М., 2010. – С. 64-65.