

На правах рукописи



**Медведев Роман Евгеньевич**

**Алгоритмы автоматизированного формирования баз знаний  
для систем дистанционного обучения**

Специальность

05.13.11 – Математическое и программное обеспечение  
вычислительных машин, комплексов и компьютерных сетей

**Автореферат**

диссертации на соискание ученой степени  
кандидата технических наук

Рязань 2013

Работа выполнена на кафедре «Вычислительной и прикладной математики» ФГБОУ ВПО «Рязанский государственный радиотехнический университет» (ФГБОУ ВПО «РГРТУ»).

**Научный  
руководитель:**

**Каширин Игорь Юрьевич,**  
доктор технических наук, профессор  
каф. ВПМ ФГБОУ ВПО «РГРТУ»

**Официальные  
оппоненты:**

**Шибанов Александр Петрович**  
доктор технических наук, профессор  
каф. САПР ВС ФГБОУ ВПО «РГРТУ»

**Чепайкин Алексей Олегович**  
кандидат технических наук, начальник  
отдела поддержки инфраструктуры  
ООО ТБ Информ г. Рязань

**Ведущая  
организация:**

**ФГБОУ ВПО «Московский  
государственный университет печати  
имени Ивана Федорова», г. Москва**

Защита состоится **6 ноября 2013 года в 12.00 часов** на заседании диссертационного совета Д 212.211.01 в ФГБОУ ВПО «Рязанский государственный радиотехнический университет» по адресу:

**390005, г. Рязань, ул. Гагарина, д.59/1.**

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВПО «Рязанский государственный радиотехнический университет».

Автореферат разослан «\_\_» октября 2013 года.

Ученый секретарь  
диссертационного совета

канд. техн. наук, доцент



**В. Н. Пржегорлинский**

## Общая характеристика работы

**Актуальность темы.** В настоящее время во всем мире наблюдается непрерывный рост интеллектуализации технологий во всех отраслях человеческой деятельности, порождающий спрос на высококвалифицированные кадры, которые в кратчайшие сроки могут повысить свою квалификацию или переквалифицироваться (пройти специализированные курсы или переподготовку). В связи с этим, всё большую ценность стал приобретать специалист, способный при необходимости получить дополнительные знания без отрыва от производства, с минимальными временными и материальными затратами, тем самым повысить свои профессиональные навыки. Растет число людей, которые в силу таких причин, как инвалидность, нехватка времени, удаленность учебного заведения, высокая стоимость образования и т.д., не могут реализовать свои потребности в образовании. Перечисленные факторы заставляют прибегнуть к поиску новых путей получения образования и обратить внимание на дистанционное обучение (ДО), основанное на широком применении компьютерных и сетевых технологий.

При организации ДО важную роль занимает проблема выбора обучающей среды, так как в случае ошибочного выбора снижается эффективность обучения. Это происходит за счет неправильного понимания и не усвоения учебного материала, отсутствия интереса к обучению, неправильной оценки знаний и не достижению заявленных целей обучения. За время развития ДО были созданы и стали доступны тысячи учебных приложений и программ учебного назначения, наиболее распространенными из которых являются системы дистанционного обучения (СДО). Под СДО понимается сложная многокомпонентная система, представляющая собой совокупность организационных, телекоммуникационных, педагогических и научных ресурсов, а также технических средств, вовлеченных в создание и практическое осуществление программ с использованием дистанционных технологий обучения.

Проведенный анализ современных СДО, результатов научных исследований и публикаций показал, что на настоящий момент в СДО недостаточно автоматизирован процесс накопления знаний, учебных материалов и тестовых заданий, хранящихся на сторонних ресурсах. В системах отсутствует модель, позволяющая адекватно описывать знания о той или иной предметной области и использовать их при дальнейшем ДО. В системах недостаточно автоматизирован процесс формирования учебно-методического материала и контрольных тестовых заданий, в ходе которого учитывались бы знания об индивидуальных характеристиках обучаемых, накопленные за время работы обучаемого с системой. Структура учебного курса задается самим преподавателем, что снижает адаптивность ДО и вызывает неоднородность трактовки содержания учебного материала и поставленных целей обучения, если над курсом работают несколько преподавателей. В результате этого возникает необходимость в использовании новых методов и программных средств, позволяющих обеспечить

индивидуальный подход и квалифицированную помощь при автоматизированном подборе учебного материала.

**Цель диссертационной работы** состоит в разработке алгоритмов формирования баз знаний для СДО с целью осуществления автоматизированного накопления, подбора и организации последовательности единиц учебного материала, учитывающей индивидуальные характеристики обучаемого.

Для достижения поставленной цели в диссертационной работе решаются следующие **задачи**:

- анализ принципов и технологий создания СДО;
- разработка модели знаний, позволяющей адекватно описывать знания о предметных областях, предназначенные для последующего дистанционного обучения;
- исследование способов отображения структурных элементов учебно-методического материала в соответствующие структурные элементы тестовых задач;
- разработка алгоритма идентификации психологического типа личности обучаемого, на основе его опыта работы с СДО;
- создание процедуры формирования последовательности учебных материалов и тестовых заданий с использованием разработанной модели знаний, учитывающей психологический тип личности обучаемого;
- разработка программного инструментария для апробации предложенных моделей и алгоритмов.

**Методы исследования.** Для решения поставленных задач использовались методы общей алгебры, теории алгоритмов, теории множеств, объектно-ориентированного проектирования, теории унификации, теории графов и дескриптивная логика.

**На защиту выносятся следующие научные результаты.**

1. Оригинальная онтологическая модель знаний, позволяющая адекватно описывать знания о предметных областях, предназначенные для последующего дистанционного обучения.
2. Алгоритм унификации DL-выражений на основе онтологической модели знаний предметной области СДО для выбора тестового и учебного материала.
3. Общие принципы подбора учебного материала, достаточного для эффективного усвоения понятий предметной области обучающимся.
4. Алгоритм идентификации психологического типа личности, позволяющий классифицировать типы личностей обучающихся.
5. Процедура формирования учебных материалов и тестовых задач, использующая онтологию как модель знаний о предметной области, и индивидуальную траекторию обучения, учитывающую особенности личности обучаемого.

**Научная новизна.**

1. Разработана оригинальная онтологическая модель знаний, адекватно описывающая знания о предметных областях СДО.

2. Введено понятие изоморфизма подразделов онтологических описаний, а также приведено конструктивное доказательство изоморфизма задачной и содержательной частей предложенной прикладной онтологии.

3. Показано, что изоморфизм содержательной и задачной составляющих онтологии является необходимым требованием для использования такой онтологии в обучении.

4. Впервые создана семантическая метрика для контроля процесса обучения на основе формального анализа контрольных точек индивидуальных трасс обучения.

5. Показана взаимосвязь индивидуальных трасс обучения с онтологической системой представления знаний о предметной области, дающая возможность разделить процессы формирования кластеров учебного материала и индивидуальные траектории обучения.

**Практическая ценность.** На основе методов, алгоритмов и подходов, предложенных в диссертационной работе, разработан программный комплекс поддержки индивидуализации дистанционного обучения SSSIDL v.1.1, выполненный на входных языках программирования PHP и MySQL. Разработанный программный комплекс позволяет структурировать учебный материал и на его основе строить индивидуальные трассы обучения. Комплекс может использоваться в СДО с целью повышения эффективности ДО, за счет: возможности экспорта готовых единиц учебного материала из описаний на языках разметки XML и OWL источников, автоматизации поиска учебного материала, учета психологических характеристик и уровня знаний обучаемого при формировании индивидуальной трассы обучения.

**Реализация результатов работы.** Результаты диссертационной работы используются:

- в учебном процессе Рязанского государственного радиотехнического университета при обучении студентов по специальностям 080801 – «Прикладная информатика (в экономике)» и 230106 – «Применение и эксплуатация автоматизированных систем специального назначения»; материалы диссертации, модели, методы и алгоритмы формирования баз знаний, полученные в ходе диссертационного исследования применяются при подготовке студентов по следующим учебным курсам: «Информатика и программирование», «Объектно-ориентированное программирование», «Предметно-ориентированные информационные системы»,

- в учебнике для ВУЗов «Предметно-ориентированные экономические информационные системы», рекомендованным УМО по образованию в области прикладной информатики в качестве учебника для студентов ВУЗов по направлению и специальности «Прикладная информатика»,

- в программном обеспечении в рамках образовательного портала администрации г. Рязани,

- в рамках научно-исследовательской работы с администрацией г. Рязани по теме «Создание математического обеспечения и программного инструмен-

тария информационной поддержки административных регламентов и единой базы данных системы общего образования» (муниципальный контракт № 1419).

**Апробация работы.** Основные положения диссертационной работы отражены в докладах 17-ой всероссийской научно-технической конференции молодых ученых и специалистов «Новые информационные технологии в научных исследованиях» г. Рязань 2012 г., в научно-практической конференции «Наука и образование в 21 веке» г. Тамбов 2012 г., в 9-ой международной научно-практической конференции «Новости передовой науки» г. София 2013 г.

**Публикации.** Основные результаты диссертации опубликованы в 18 работах. В том числе: 3 статьи в изданиях, рекомендуемых ВАК, 9 статей в межвузовских сборниках научных трудов, 6 тезисов докладов на международных и всероссийских конференциях.

**Структура и объем диссертации.** Диссертация состоит из введения (7 с.), 4-х глав (128 с. – 24 рисунка и 15 таблиц). Библиографический список включает 87 наименований (9 с.).

### **Краткое содержание работы**

**Во введении** дается обоснование актуальности темы диссертационной работы, формулируются цели и задачи работы, указывается научная новизна и практическая ценность, определяются основные положения, выносимые на защиту, кратко излагается содержание диссертации.

**Первая глава** носит обзорный характер. В главе рассмотрены достоинства ДО, различные средства его поддержки, предложена их классификация. Рассмотрен результат процесса внедрения интеллектуальных технологий в дистанционное образование, нашедший свое отражение в появлении интеллектуальных систем дистанционного обучения (ИСДО). Определены основные принципы реализации ИСДО, приведены их отличительные особенности, преимущества, структура и методы взаимодействия с обучаемым. В главе рассмотрена архитектура ИСДО и выделены основные компоненты системы: модель обучаемого, модель предметной области, модель процесса обучения. Приведены методы построения каждой из моделей, рассмотрены их преимущества и недостатки. Рассмотрены достоинства и недостатки различных моделей представления знаний: логических моделей, фреймов, семантических сетей, продукционных моделей, онтологий. Выявлены и проанализированы методы адаптации ИСДО под нужды конкретного обучаемого, рассмотрены достоинства и недостатки каждого из методов. Выделена проблема представления знаний в ИСДО и определены задачи для ее решения. Представлены результаты проведенного анализа существующих ИСДО, определены и сформулированы их недостатки.

**Вторая глава** посвящена разработке оригинальной технологии накопления знаний в системах дистанционного обучения на основе онтологической модели представления знаний, с применением инструментария OWL-S и WSDL для повышения эффективности использования Internet-ресурсов.

В главе спроектирована прикладная онтология, позволяющая адекватно описывать знания о предметных областях СДО. В рамках предложенной при-

кладной онтологии (рис. 1) целесообразно сформулировать ключевую фразу задачного подхода: «для решения задачи студент должен усвоить все понятия, которые необходимы при ее решении».

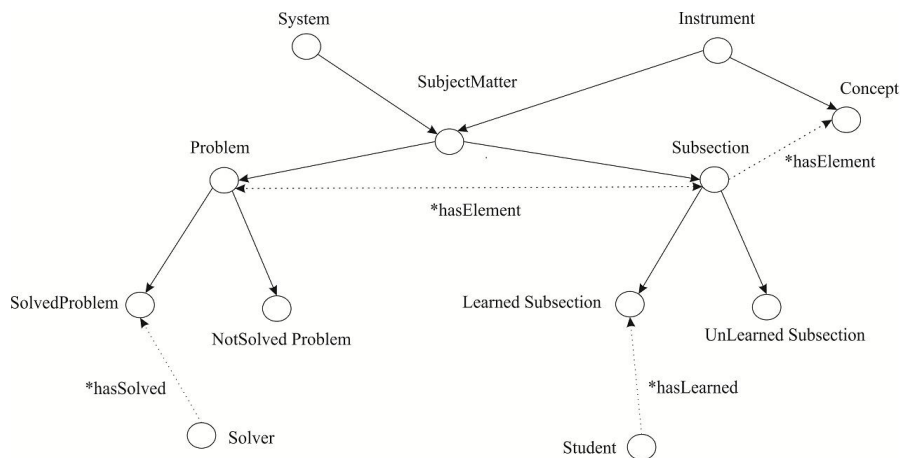


Рисунок 1 - Семантическая сеть фрагмента онтологии «Problem, Subsection»

Для конструктивной работы с формулами дескриптивной логики определены два раздела выражений: TBox и ABox, описывающие абстрактную и конкретную составляющие прикладной онтологии. Нумерация формул прикладной онтологии обучения представлена в табл. 1.

Таблица 1 – Нумерация формул прикладной онтологии обучения

№ п/п	Понятие / Концепт	Формула
1	System	$= \forall \text{hasElement. Element}$
2	Causa	$= \exists \text{hasConsequence. Consequence}$
3	Instrument	$= \exists \text{hasActor. Person} \cap \exists \text{isApplied} \bar{.} \text{Problem}$
4	Solver	$= \exists \text{isSolved. Problem} \cap \text{Student}$
5	SolvedProblem	$= \exists \text{isSolved} \bar{.} \text{Problem} \cup \exists \text{isSolved. Solver}$
6	NotSolvedProblem	$= \exists \text{hasSolver.} \perp$
7	Problem	$= \exists \text{isSolved} \bar{.} \text{Problem} \cup \exists \text{isSolved. Solver}$
8	SubjectMatter	$\subseteq \text{System} \cap \text{Instrument}$
9	SubjectMatter	$= \forall \text{hasElement. Subsection}$
10	Subsection	$\subseteq \text{SubjectMatter}$
11	Problem	$\subseteq \text{SubjectMatter}$
12	Subsection	$= \forall \text{hasElement. Problem} \cap \forall \text{hasElement. Concept}$
13	Learned Subsection	$= \text{Subsection} \cap ( \exists \text{hasLearned} \bar{.} \text{Subsection} \cup \exists \text{hasLearned. Student} )$
14	Un Learned Subsection	$= \text{Subsection} \cap \exists \text{hasLearned.} \perp$
15	Concept	$\subseteq \text{Instrument}$
16	Student	$\subseteq \text{Person}$
17	Process	$= \text{Dynamic}$
18	Learning	$\subseteq \text{Process} \cap \exists \text{hasConsequence. SolvedProblem}$
19	Learn_the_Concept	$= \exists \text{hasLearn. Concept}$
20	ProblemConcepts	$= \text{Concept} \cap \forall \text{isApplied} \bar{.} \text{SolvedProblem}$
21	LearnProblemConcepts	$= \text{Learning} \cap \forall \text{hasLearn. ProblemConcepts}$

В главе выделена формальная составляющая процесса обучения, представляющая собой множество средств обучения: учебно-методические материалы, пособия, лабораторные работы, стенды, тестовые задачи. Без существенной потери качества эти средства можно свести к двум основным подмножествам: учебные материалы и тестовые задачи. От точности соответствия друг другу элементов этих множеств зависит качество процесса обучения. Оба множества имеют схожую структуру и опираются на базовые множества лежащих в их основе понятий. С точки зрения математики имеет смысл рассматривать различные отображения структурных элементов учебного материала в соответствующие структурные элементы тестовых задач. Наиболее качественным из отображений предлагается признать изоморфизм, предполагающий взаимнооднозначное соответствие элементов базовых множеств. Для математически строгого описания изоморфизма онтологий рассматривается множество правильно построенных формул (ППФ) этих онтологий как множество упорядоченных пар  $(C, F)$ , где  $C$  – множество концептов,  $F$  – множество формул из раздела ТВох. Таким образом, с помощью универсума  $C \times F$  задается отношение дескрипции  $R_d \subseteq C \times F$ .

Далее можно рассмотреть только правые части  $F$  этого отношения, как множество, представляющее собой проекцию на второй аргумент отношения. При этом выражения ППФ в части  $F$  можно представить как результат конечно-го множества операций композиции элементов  $\{c_1, c_2, \dots, c_k\}$  из множества  $C$ . Эти выражения получаются из имен концептов  $c_j$  с помощью специальных знаков дескриптивной логики:  $\{\subseteq, =, \neg, \forall, \cup, \cap, ;, -\}$ . Приведем семантику таких операций композиции  $\Omega$ , учитывая замкнутость этих операций на множестве  $F$  (табл. 2).

Таблица 2 – Операции композиции ППФ

№ п/п	Обозначение операции из $\Omega$	Семантика (результат операции)
1	$\omega_{\cap} (F_j, F_i)$	$F_j \cap F_i$
2	$\omega_{\cup} (F_j, F_i)$	$F_j \cup F_i$
3	$\omega_{\exists R_j} (c_i)$	$\exists R_j . c_i$
4	$\omega_{\forall R_j} (c_i)$	$\forall R_j . c_i$
5	$\omega_{\forall R_j^-} (c_i)$	$\forall R_j^- . c_i$
6	$\omega_{\exists R_j^-} (c_i)$	$\exists R_j^- . c_i$
7	$\omega : (c_i)$	$: c_i$

В ППФ можно выделить два отношения  $\{\subseteq, =\}$ , которые связывают левую и правую части онтологических определений. В сделанных обозначениях на множестве ППФ онтологии можно задать алгебраическую систему  $A$ :

$$A = \langle F, \Omega, \{\subseteq, =\} \rangle.$$

В алгебраической системе  $A$  можно выделять подалгебры, замкнутые на своих множествах-носителях. Например, определим две подалгебры  $G$  и  $D$  для алгебры  $\langle F, \Omega \rangle$ :  $G = (F_g, \Omega_g)$  и  $D = (F_d, \Omega_d)$ . В них справедливы соотно-



шения  $F_g \subseteq F, \Omega_g \subseteq \Omega, F_d \subseteq F, \Omega_d \subseteq \Omega$ . Предположим также, что эти подалгебры имеют один и тот же тип.

Алгебры  $G$  и  $D$  являются алгебрами одинакового типа. Алгебры  $G$  и  $D$  являются *гомоморфными*, так как существует отображение  $\Gamma: M_g \rightarrow M_d$ , удовлетворяющее условию:

$$\Gamma(\omega_g(f_{g1}, f_{g2})) = \omega_d(\Gamma(f_{g1}), \Gamma(f_{g2})), \quad (1)$$

где  $f_{g1}, f_{g2} \subseteq F_g$ .

Независимо от того, выполнена ли сначала операция композиции в  $G$  и затем произведено отображение  $\Gamma$ , либо сначала произведено отображение  $\Gamma$ , а затем в  $D$  выполнена соответствующая операция композиции  $\omega_d$ , результат будет одинаков.

Алгебры  $G$  и  $D$  являются *изоморфными* или *однозначно гомоморфными*, так как существует обратное отображение  $\Gamma^{-1}: F_d \rightarrow F_g$ , удовлетворяющее условию:

$$\omega_g(\Gamma^{-1}(f_{d1}), \Gamma^{-1}(f_{d2})) = \Gamma^{-1}(\omega_d(f_{d1}, f_{d2})), \quad (2)$$

где  $f_{d1}, f_{d2} \subseteq F_d$ .

Кроме того, потребуем для всех элементов множеств-носителей из  $G$  и  $D$  соответствие в алгебраических системах всех элементов множества носителя каждого определяемого концепта отношениям  $\{\subseteq, =\}$ . То есть, если концепт из  $G$  соединен в итоговое описание отношением  $\subseteq$ , то и соответствующий ему концепт из  $D$  также должен быть описан с помощью отношения  $\subseteq$ . Если же при описании концепта из  $G$  было задействовано отношение  $=$ , то соответствующий концепт из  $D$  также должен быть описан с помощью равенства.

Равенство (2) приводится к равенству (1) следующим образом.

1. Заменяем в условии (1) левые части на правые

$$\omega_d(\Gamma(f_{g1}), \Gamma(f_{g2})) = \Gamma(\omega_g(f_{g1}, f_{g2})).$$

2. Применим  $\Gamma^{-1}$  к обеим частям получившегося равенства

$$\Gamma^{-1}(\omega_d(\Gamma(f_{g1}), \Gamma(f_{g2}))) = \Gamma^{-1}(\Gamma(\omega_g(f_{g1}, f_{g2}))) = \omega_g(f_{g1}, f_{g2}).$$

3. Так как  $\Gamma(f_{gi}) = f_{di}$ ,  $f_{di} \in F_d$ , тогда  $f_{gi} = \Gamma^{-1}(f_{di})$ , следовательно,

$$\omega_g(\Gamma^{-1}(f_{d1}), \Gamma^{-1}(f_{d2})) = \Gamma^{-1}(\omega_d(f_{d1}, f_{d2})).$$

Мощности основных множеств изоморфных алгебр  $G$  и  $D$  равны, т.е. изоморфизм алгебр  $G$  и  $D$  называется *изоморфизмом на себя* или *автоморфизмом*.

Конструктивное доказательство изоморфизма двух локальных фрагментов онтологий может рассматриваться как полное структурное совпадение соответствующих ППФ с точностью до определения соответствия концептов. Изоморфизм прикладных онтологий позволяет применять пары («Материал», «Тест») в обучающих системах, и дает возможность использовать аналогичное доказательство при автоматическом подборе материала в раздел онтологии АВох.

Для автоматического доказательства изоморфизма разделов онтологии АBox, формируемых из информационных хранилищ, необходимо воспользоваться двумя новыми механизмами, раскрываемыми посредством понятий «тезаурус» и «унификация».

Под тезаурусом понимается концептуальный словарь, содержащий для каждой из составляющих его лексем одно родовидовое отношение, отношение синонимии, возможно, отношение антонимии и другие отношения, которые могут задаваться администратором тезауруса. При этом родовидовое отношение описывает связь лексемы с более общим понятием в родовидовой таксономии. Структура лексемы предлагаемого тезауруса выглядит следующим образом:

Наименование лексемы + [Нормальная форма] + [Корневая основа] + [Лексема предок] + {[Антоним]} + {[Синоним]}

Унификация онтологических описаний предполагает осуществление попытки привести два описания к единственному эквивалентному им описанию при выполнении реорганизации исходных описаний. Под терминами понимаются ППФ дескриптивной логики, описывающие предметную область. Для унификации в онтологиях находятся классы эквивалентности и подобия всех составляющих. Алгоритм унификации (рис. 2) имеет целью поиск всех отношений, в которых находятся два онтологических описания. При этом прикладная часть описаний может быть унифицирована на множестве прикладных формул, в то время как для задачной онтологии для раздела АBox должен применяться

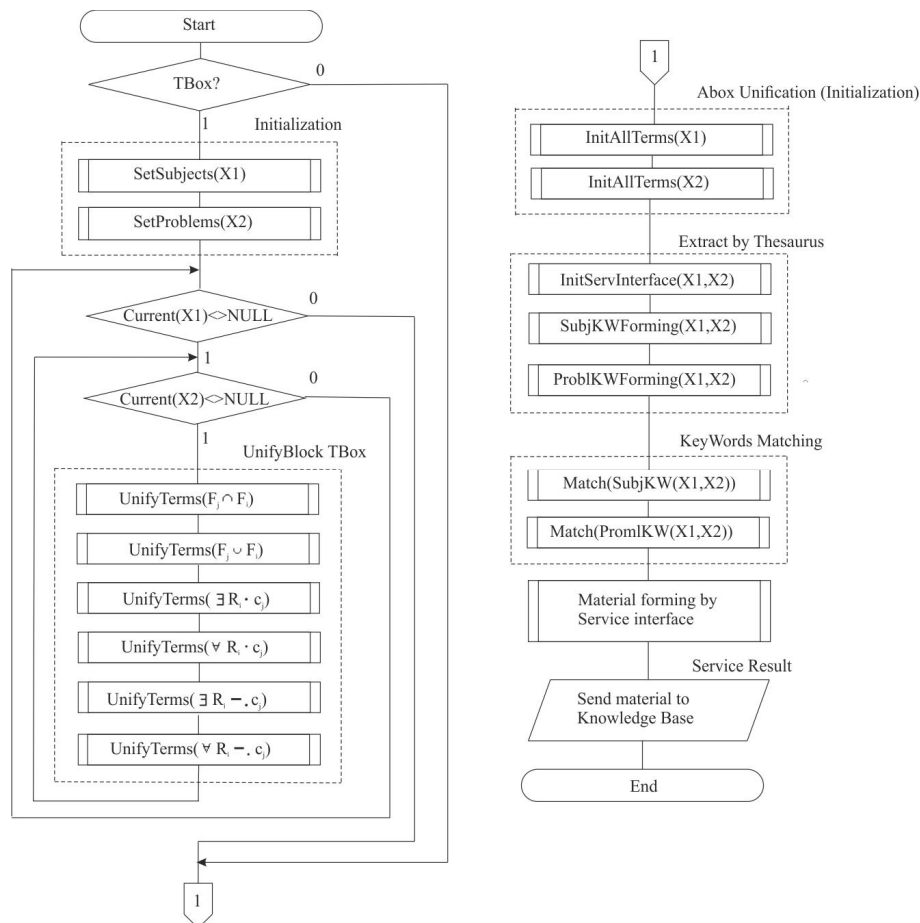


Рисунок 2 – Алгоритм унификации термов разделов онтологии

алгоритм, использующий соотношения концептов с точностью до классов эквивалентности в лексемах тезауруса.

В начале алгоритма осуществляется проверка изоморфизма соответствующих разделов ( $X1$ ,  $X2$ ) из  $TBox$ . Для этого проверяются на симметричный гомоморфизм соответствующие описания, выполненные на основе операций объединения и пересечения термов, а также существования у концептов взаимоотображаемых свойств. В проверке используются вложенные циклы, реализующие перебор всех термов ППФ из  $X1$  и  $X2$ . После проверки на унифицируемость подразделов  $TBox$  реализуется унификация соответствующих ( $X1$  и  $X2$ ) подразделов  $ABox$ , как доказательство изоморфизма ППФ конструкций.

Завершается алгоритм сопоставлением ключевых слов на их наличие в агрегации тех синтаксических связей, которые задаются соответствующими описаниями прикладной онтологии. Предложенный алгоритм унификации является достаточно эффективным и обладает свойствами адаптивности, поскольку дает возможность управлять точностью выбора содержательной части лексем, корректируя классы эквивалентности тезауруса.

**В третьей главе** рассматриваются алгоритмы проектирования индивидуальных трасс обучения на основе психологического типа личности обучаемого.

Цель теоретических исследований, проведенных в данной главе, может быть сформулирована следующим образом: «необходимо получить математический формализм и инструментальные программные средства, достаточные для адекватной классификации обучающихся по широкому спектру психологических свойств их личностей для выявления наиболее эффективных учебных ресурсов, пригодных обучаемым с конкретными психологическими особенностями». При этом в главе не ставится цели исследовать множество существующих психологических типов личностей.

Перед тем как предложить обучающемуся тестовую задачу, нужно обеспечить все основные элементы ситуации, в которой эта задача может быть решена, а именно:

- обучающемуся должен быть предоставлен учебно-методический материал, достаточный по своему содержанию для решения тестовой задачи,
- учебно-методический материал должен соответствовать уровню подготовленности обучающегося, для того, чтобы материал мог быть правильно понят и усвоен,
- материал должен соответствовать по форме изложения, сложности и подробности психологическому типу личности обучающегося,
- тестовая задача на конкретный момент времени должна соответствовать уровню подготовленности с учетом психологического типа личности,
- последовательность предлагаемого материала и тестовых заданий должна соответствовать психологическому типу обучающегося.

Для подбора материала и формирования индивидуальной трассы обучения используются следующие функции:

-  $MCom(m_i): m_i \rightarrow [0,1]$  и функция определения сложности учебного материала  $m_i$ :

$$MCom(m_i) = |M'(C_{req})|/|C_{chapt}|,$$

где  $|M'(C_{req})|$  - мощность множества требуемых понятий единицы учебно-методического материала;  $|C_{chapt}|$  - мощность множества всех понятий курса,

-  $PCom(p_i): p_i \rightarrow [0,1]$ , функция определения сложности задачи  $p_i$ :

$$PCom(p_i) = |P'(C_{req})|/|C_{chapt}|,$$

где  $|P'(C_{req})|$  - мощность множества требуемых понятий тестовой задачи,

-  $MType(m_j, tm_k): m_j \times tm_k \rightarrow [0,1]$ , функция, ставящая в соответствие каждому элементу из множества единиц учебного материала  $m_j$ , подмножество психологических типов, которые наиболее просто смогут усвоить этот материал:

$$MType(m_j, tm_k) = \sum_{i=1}^{|Tm|} MAcc(m_j, tm_{k,i})/|Tm|,$$

где  $MAcc$  - функция определения соответствия учебного материала  $m_j$  характеристике типа личности  $tm_{k,i}$ ;  $|Tm|$  - мощность множества характеристик единицы учебного материала, используемых при определении психологического типа личности обучаемого,

-  $PType(p_j, tp_k): p_j \times tp_k \rightarrow [0,1]$ , функция, ставящая в соответствие каждому элементу из множества тестовых задач  $p_j$ , подмножество психологических типов  $tp_k$ , которые наиболее просто смогут их решить:

$$PType(p_j, tp_k) = \sum_{i=1}^{|Tp|} PAcc(p_j, tp_{k,i})/|Tp|,$$

где  $PAcc$  - функция определения соответствия тестовой задачи  $p_j$  характеристике типа личности  $tp_{k,i}$ ;  $|Tp|$  - мощность множества характеристик задачи, используемых при определении психологического типа личности обучаемого,

-  $MSem(c_i, m_j): c_i \times m_j \rightarrow [0,1]$ , функция соответствия каждого понятия  $c_i$  учебно-методическому материалу  $m_j$ :

$$MSem(c_i, m_j) = 1/|Mc_{view}|, \quad c_i \in Mc_{view},$$

где  $|Mc_{view}|$  - мощность множества рассмотренных понятий единицы учебно-методического материала,

-  $PSem(c_i, p_j): c_i \times p_j \rightarrow [0,1]$ , функция соответствия каждого понятия  $c_i$  выбираемой тестовой задаче  $p_j$ :

$$PSem(c_i, p_j) = 1/(|Pc_{view}| + |Pc_{serv}|), \quad c_i \in Pc_{view} \vee c_i \in Pc_{serv},$$

где  $|Pc_{view}|$  - мощность множества рассмотренных понятий тестовой задачи;  
 $|Pc_{serv}|$  - мощность множества вспомогательных понятий тестовой задачи,

-  $MVal(m_i): m_i \rightarrow [0,1]$ , функция, определяющая меру важности единицы учебно-методического материала  $m_i$ :

$$MVal(m_i) = |Mc_{req} = Mc'_{view}| / |M|,$$

где  $|Mc_{req} = Mc'_{view}|$  - мощность множества единиц учебного материала, для которых рассмотренные понятия единицы учебного материала  $m_i$  являются требуемыми;  $|M|$  - мощность множества всех учебных материалов курса,

-  $PVal(p_i): p_i \rightarrow [0,1]$ , функция, определяющая меру важности задачи  $p_i$ :

$$PVal(p_i) = |Mc_{req} = Mc''_{view}| / |M|,$$

где  $|Mc_{req} = Mc''_{view}|$  - мощность множества единиц учебного материала, для которых рассмотренное понятие текущей тестовой задачи являются требуемым понятием,

-  $CSt(c_i, s_j): c_i \times s_j \rightarrow [0,1]$ , выражение, формализующие степень усвоения понятий предметной области обучающимся  $s_j$ :

$$CSt(c_i, s_j) = |Pc_{view} = c_i| / |Pc_{view}' = c_i|,$$

где  $|Pc_{view} = c_i|$  - мощность множества задач решенных  $j$ -ым обучаемым, в которых текущее понятие считается рассмотренным;  $|Pc_{view}' = c_i|$  - мощность множества задач как решенных  $j$ -ым обучаемым, так и не решенных, в которых текущее понятие считается рассмотренным,

-  $TSt(t_i, s_j): t_i \times s_j \rightarrow [0,1]$ , выражение, формализующие степень принадлежности обучающегося  $s_j$  определенному типу личности:

$$TSt(t_i, s_j) = \frac{|P_{stud,j}|}{\sum_{k=1}^{|P_{stud,j}|} PType(p_k, tp_i)} / |P_{stud,j}|,$$

где  $|P_{stud,j}|$  - мощность множества всех решенных пользователем задач.

Цель выбора требуемого материала как подмножества  $M^+ \subseteq M$  и тестовых задач как подмножества  $P^+ \subseteq P$  описывается следующим выражением:

$$\{M^+, P^+ : m_i, p_t \mid \forall m_i \in M, \forall p_t \in P, \\ MType(m_i, tm_k) > \min_{j=1}^n (m_j) \& PType(p_v, tp_k) > \min_{v=1}^m (p_v)\},$$

где операция  $\min$  определяет порог принадлежности элементов множеств  $M$  и  $P$  по соответствию определенному типу личности.

Следующим набором выражений описывает принцип, основанный на подготовленности обучаемого и сложности задач как подмножества  $P^* \subseteq P$  и материала как подмножества  $M^* \subseteq M$ :

$$\{P^* : p_t \mid \forall c_i \in C, \forall p_t \in P, \exists s_j TSt(t_r, s_j) > \min_{r=1}^m (t_r) \& CSt(c_i, s_j) > \min_{i=1}^n (c_i) \& PSem(c_i, p_t) > \min_{t=1}^m (p_t)\},$$

$$\{M^* : m_t \mid \forall c_i \in C, \forall m_t \in M, \exists s_j TSt(t_r, s_j) > \min_{r=1}^m (t_r) \& CSt(c_i, s_j) > \min_{i=1}^n (c_i) \& MSem(c_i, m_t) > \min_{t=1}^m (m_t)\}.$$

Для окончательного отбора итоговых множеств материала  $M^0$  и задач  $P^0$ , необходимо вычислить пересечение полученных множеств:

$$M^0 \equiv M^* \cap M^+, P^0 \equiv P^* \cap P^+.$$

Полученные множества теперь содержат учебно-методический материал и тестовые задачи, выбранные специально для обучающегося определенного психологического типа, а также для предметной области, понятия которой должен усвоить этот обучающийся.

Заключительным этапом формирования выборки задач и материала должен стать этап упорядочения элементов множеств  $M^0$ ,  $P^0$  для формирования индивидуальной трассы обучения:

$$CortM(m_1, m_2, \dots, m_i, \dots, m_k) : \forall m_i, m_{i+1} MCom(m_i) \leq MCom(m_{i+1}),$$

$$CortP(p_1, p_2, \dots, p_j, \dots, p_l) : \forall p_j, p_{j+1} PCom(p_j) \leq PCom(p_{j+1}),$$

где  $CortM$  и  $CortP$  соответственно кортежи последовательности выбора материалов и последовательности предложения тестовых задач.

В главе рассмотрен геометрический смысл принципа формирования индивидуальных трасс обучения. Для графического представления динамики обучения по осям  $n$ -мерного пространства отложены метрические шкалы, соответствующие степени усвоения каждого из понятий некоторой предметной области. Такое представление даст возможность использовать семантические пространства  $\mathcal{C}$ . Осгуда, точками которого можно характеризовать как задачи, предлагаемые на тестирование, так и материалы, необходимые для изучения (рис. 3, рис. 4).

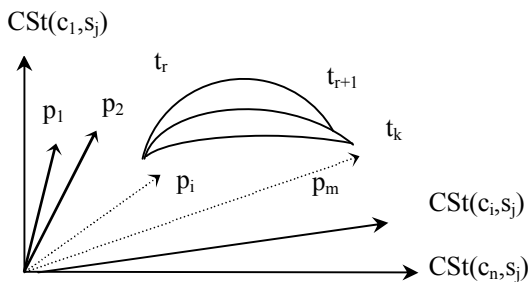


Рисунок 3— Семантическая метрика планирования обучения (задачи)

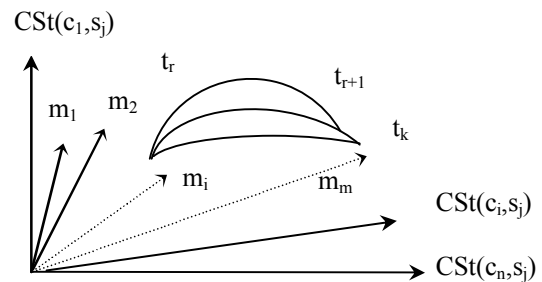
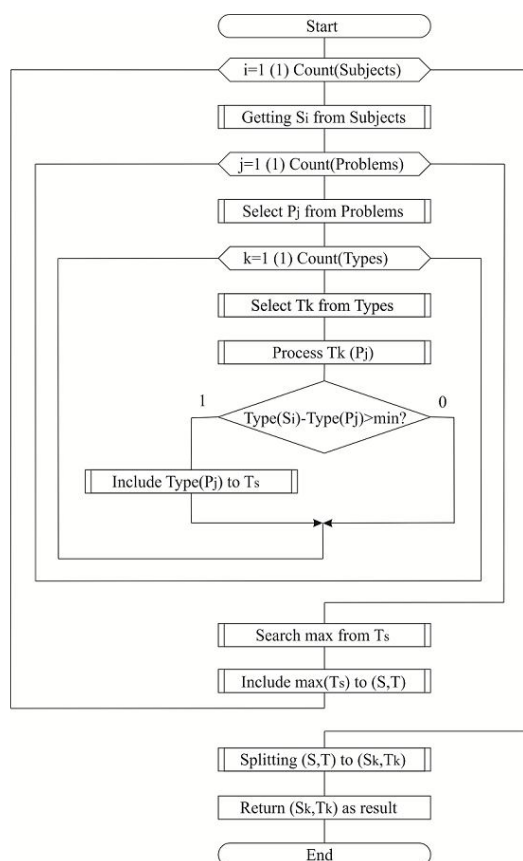


Рисунок 4— Семантическая метрика для планирования обучения (материал)

Как следует из рисунков, существуют точки  $p_i$  и  $m_i$ , обозначающие изначальное состояние знаний обучающегося. Конечная цель курса обучения, определяемая точками  $p_m$  и  $m_m$ . Достижение конечной цели курса обучения возможно при использовании каких-либо промежуточных последовательностей изучения поэтапно новых учебных материалов и контролем с опорой на соответствующие тестовые задачи. Таких последовательностей может быть несколько, они могут не пересекаться или наоборот, иметь общие составляющие. Задачей подбора учебного материала является выбор наилучшей последовательности, которая будет являться индивидуальной трассой обучения.

В главе описан (табл. 3) алгоритм идентификации психологического типа личности (рис. 5) с точки зрения обучающей системы. Алгоритм дает возможность не только определять схожесть обучающихся, но и их различия, а также выделять коллективы обучающихся.

Таблица 3 – Пояснения к алгоритму идентификации



№ п/п	Формальная запись	Семантика выполнения конструкции
1	Getting $S_i$ from Subjects	Получение очередного элемента из множества обучающихся.
2	Select $P_j$ from Problems	Выбор очередной решенной обучающимся $S_i$ задачи из множества задач.
3	Select $T_k$ from Types	Выбор очередного типа задачи.
4	Process $T_k(P_j)$	Определение степени принадлежности решенной задачи очередному типу задачи.
5	$Type(S_i) - Type(P_j) > \min$	Определение разности мер уверенности для соответствующих типов задачи и предыдущего значения типа личности обучаемого.
6	Include $Type(P_j)$ to $T_s$	Включение типа задачи в выделенное множество, на основе которого будет выделен идентифицируемый тип личности
7	Search max from $T_s$	Определение типа личности $T_s$ с максимальным значением уверенности.
8	Include max( $T_s$ ) to ( $S, T$ )	Включение пар (обучающийся, его тип) в соответствующее множество результирующих пар.
9	Splitting ( $S, T$ ) to ( $S_k, T_k$ )	Разбиение множества пар (обучающийся, его тип) на подмножества, выделяемые каждым новым типом личности.
10	Return( $S_k, T_k$ ) as result	Возвращение результата работы алгоритма как множество полученных пар.

Рисунок 5 – Алгоритм идентификации психологического типа личности

В главе рассмотрена взаимосвязь индивидуальных трасс обучения с онтологической системой представления знаний о предметной области. Предлагается выделить в индивидуальной трассе обучения множество контрольных точек  $R_i$ , представляющих собой пару  $(m_i, p_i)$ . Каждой такой точке по содержа-

нию учебного материала будет соответствовать определенное подмножество понятий предметной области  $C_R \subseteq C$ , которые необходимо усвоить перед проведением тестовых испытаний. Таким образом, объективно существует соответствие  $\psi$ , отображающее каждую контрольную точку  $R_i$  в некоторые подмножества понятийной базы  $C$ , что можно записать так:

$$\forall R_i \subseteq R \exists C_R \subseteq C, \psi: R_i \rightarrow C_R.$$

Соответствие  $\psi$  является функциональным и семантически определяется простейшим табличным алгоритмом задания соответствия. Этот алгоритм является первым шагом процедуры выбора учебного материала и тестовых заданий. Вторым шагом процедуры выбора учебного материала является работа с OWL-выражениями, задающими компоненты ABox и TBox онтологии предметной области. Поскольку для любой контрольной точки  $R_i$  существует соответствующее ему множество понятий  $C_R$ , то для идентификации учебного материала как подходящего ресурса для использования в обучении на контрольной точке  $R_i$  применяется алгоритм, доказывающий связность семантической сети понятий  $C_R$  относительно соответствующей онтологии. Алгоритм реализуется следующей последовательностью шагов.

Шаг 1. Строятся одноуровневые кластеры для каждой из вершин, соответствующих в семантической сети онтологии понятиям тезауруса.

Шаг 2. Проверяет возможность связывания всех кластеров в единую сеть, добавляя их по одному к общей сети на основе существования общих вершин у кластера и предыдущей сформированной сети.

Шаг 3. Если при связывании множество рассматриваемых понятий становится пустым, выдается признак идентификации семантической подсети как связной, что свидетельствует об идентификации материала по своему содержанию как пригодного для изучения по выбранной тематике предметной области.

**В четвертой главе** рассматривается программный комплекс (ПК) поддержки индивидуализации дистанционного обучения SSSIDL v.1.1 (Software System to Support Individualization of Distance Learning). В главе приведена архитектура ПК SSSIDL v.1.1 (рис. 6) и схема взаимодействия с СДО.

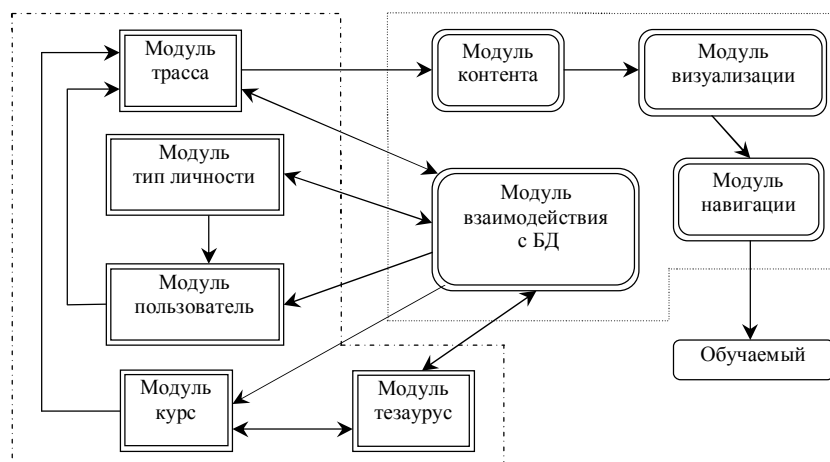


Рисунок 6 – Архитектура SSSIDL



В главе приведена структура, каждого из компонентов ПК и алгоритмы определения типа личности обучаемого и формирования индивидуальной трассы обучения.

Рассмотрена эффективность работы ПК, оцененная с двух точек зрения: преподавателя и обучаемого. Проанализированы результаты экспериментальных исследований, доказывающие эффективность применения на практике разработанных методов и алгоритмов, реализованных в ПК SSSIDL v.1.1.

**В заключении** приводится обобщение основных результатов работы.

### **Основные результаты работы**

1. Определены недостатки существующих систем дистанционного обучения, дифференцированные по разновидностям этих систем.
2. Показана актуальность задачи создания автоматизированных средств формирования баз знаний на основе имеющихся хранилищ готовых учебных материалов и тестовых примеров.
3. Разработана оригинальная онтологическая модель знаний, позволяющая адекватно описывать знания о предметных областях, предназначенные для последующего дистанционного обучения.
4. Впервые введено понятие изоморфизма подразделов онтологических описаний как двунаправленного гомоморфизма алгебраических термов, описывающих подразделы. Показано, что изоморфизм содержательной и тестирующей составляющих онтологии является необходимым требованием для использования такой онтологии в обучении.
5. Сформулирована постановка задачи синтеза обучающих ресурсов, ориентированных на психологическую модель обучаемого. В новой постановке формально описаны принципы подбора учебного материала, достаточного для эффективного усвоения понятий предметной области обучающимся.
6. Создан новый алгоритм идентификации психологического типа личности, позволяющий классифицировать типы личностей обучающихся в  $n$ -уровневой иерархии, где  $n$  заранее неопределенно.
7. Разработана процедура формирования учебных материалов и тестовых заданий, использующая онтологию, как модель знаний о предметной области, и индивидуальную траекторию обучения, учитывающую особенности личности обучаемого.
8. Программно реализован ПК SSSIDL v.1.1, использующий основные результаты теоретической части диссертации.

### **Публикации по теме диссертации**

#### Статьи в изданиях по списку ВАК

1. Каширин И. Ю., Медведев Р.Е. Онтологическое накопление учебных знаний на основе сервиса информационных сетей // Научно-технические ведомости СПбГПУ. 2013. №174. С. 75-85.
2. Медведев Р.Е. Изоморфизм в онтологической модели знаний интеллектуальной системы дистанционного обучения // Вестник РГРТУ. 2013. № 45. С. 62-65.
3. Медведев Р.Е. Моделирование индивидуальной трассы изучения материала на основе базы знаний системы дистанционного обучения // Фундаментальные исследования. 2013. №10. С. 1001-1004.

*Статьи в сборниках научных трудов*

4. Медведев Р.Е. Анализ существующих подходов организации систем дистанционного обучения // Математическое и программное обеспечение вычислительных систем: межвузовский сборник научных трудов. Рязань: РГРТУ. 2010. С. 133-137.

5. Медведев Р.Е. Ключевые критерии выбора платформы для системы дистанционного обучения // Информационные технологии: межвузовский сборник научных трудов. Рязань: РГРТУ. 2011. С. 101-104.

6. Медведев Р.Е. Модели представления знаний в интеллектуальных системах дистанционного обучения // Математическое и программное обеспечение вычислительных систем: межвузовский сборник научных трудов. Рязань: РГРТУ. 2011. С. 36-38.

7. Медведев Р.Е. Проблема выбора платформы для дистанционной обучающей среды // Математическое и программное обеспечение вычислительных систем: межвузовский сборник научных трудов. Рязань: РГРТУ. 2011. С. 108-110.

8. Медведев Р.Е. Стратегии проверки знаний в адаптивных системах дистанционного обучения // Программные информационные системы: межвузовский сборник научных трудов. Рязань: РГРТУ. 2011. С. 4-6.

9. Каширин И.Ю., Медведев Р.Е. Методы контроля знаний в системах дистанционного обучения // Информационные технологии в учебном процессе и научных исследованиях: межвузовский сборник научных трудов молодых ученых, специалистов и студентов. Рязань: РГРТУ. 2012. С. 28-31.

10. Медведев Р.Е. Модель обучаемого как ключевой инструмент адаптации в интеллектуальных системах дистанционного обучения // Информационные технологии в образовании: межвузовский сборник научных трудов. Рязань: РГРТУ. 2012. С. 106-109.

11. Медведев Р.Е. Способ организации знаний в интеллектуальных системах дистанционного обучения // Методы и средства обработки и хранения информации: межвузовский сборник научных трудов. Рязань: РГРТУ. 2012. С. 125-128.

12. Медведев Р.Е. Структура современных интеллектуальных систем дистанционного обучения // Математическое и программное обеспечение вычислительных систем: межвузовский сборник научных трудов. Рязань: РГРТУ. 2012. С. 33-35.

*Тезисы докладов на международных и всероссийских конференциях*

13. Медведев Р.Е. Анализ существующих технологий построения систем дистанционного обучения // Проблемы передачи и обработки информации в сетях и системах телекоммуникаций: материалы 16-ой международной науч.-техн. конф. Рязань: РГРТУ. 2010. С. 128-129.

14. Медведев Р.Е. Интеллектуальные системы дистанционного обучения // Наука и образование в 21 веке: материалы международной науч.-прак. конф. Тамбов: ТРОО "Бизнес - наука - общество". 2012. С. 97-98.

15. Медведев Р.Е. Использование open source платформы для построения интеллектуальной системы дистанционного обучения // Новые информационные технологии в научных исследованиях: материалы 17-ой международной науч.-техн. конф. молодых ученых и специалистов. Рязань: РГРТУ. 2012. С. 28-29.

16. Медведев Р.Е. Классификация программных обучающих сред // Наука и образование в 21 веке: материалы международной науч.-прак. конф. Тамбов: ТРОО "Бизнес - наука - общество". 2013. С. 85-86.

17. Медведев Р.Е. Онтологическая модель представления знаний в интеллектуальных обучающих системах // Новости передовой науки: материалы 9-ой международной науч.-практ. конф. София: ООД "Бял ГРАД-БГ" 2013. С. 27-29.

18. Медведев Р.Е. Применение internet технологий в системах дистанционного обучения // Новые информационные технологии в научных исследованиях: материалы 16-ой всероссийской науч.-техн. конф. студентов, молодых ученых и специалистов, посвященная празднованию юбилея РГРТУ. Рязань: РГРТУ. 2011. С. 171-172.

**Медведев Роман Евгеньевич**

**Алгоритмы автоматизированного формирования баз знаний  
для систем дистанционного обучения**

Автореферат  
диссертации на соискание ученой степени  
кандидата технических наук

Подписано в печать 04.09.2013 г. Формат бумаги 60x84/16.  
Бумага офсетная. Печать трафаретная. Усл. печ. л. 1.0.  
Тираж 100 экз.

Рязанский государственный радиотехнический университет.  
390005, г. Рязань, ул. Гагарина, 59/1.  
Редакционно-издательский центр РГРТУ.