

На правах рукописи



Мошкин Вадим Сергеевич

**ИССЛЕДОВАНИЕ ПРЕДСТАВЛЕНИЯ
ТЕРМИНОЛОГИИ В ЛИНГВИСТИЧЕСКОМ
ОБЕСПЕЧЕНИИ САПР НА ОСНОВЕ ИНТЕГРАЦИИ
НЕЧЕТКИХ ОНТОЛОГИЙ И ЛОГИЧЕСКОГО
ВЫВОДА**

Специальность: 05.13.12 – Системы автоматизации
проектирования (промышленность)

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Ульяновск – 2017

Работа выполнена на кафедре «Информационные системы» в Ульяновском государственном техническом университете.

Научный руководитель: доктор технических наук, профессор
Ярушкина Надежда Глебовна

Официальные оппоненты: **Ковалев Сергей Михайлович**,
доктор технических наук, профессор,
ФГБОУ ВО «Ростовский государственный
университет путей сообщения», кафедра
«Автоматика и телемеханика на
железнодорожном транспорте», профессор
кафедры

Смагин Алексей Аркадьевич, доктор
технических наук, профессор, кафедра
телекоммуникационных технологии и сетей,
ФГБОУ ВО «Ульяновский государственный
университет», зав. кафедрой

Ведущая организация: Федеральный научно-производственный
центр АО "Научно-производственное
объединение «Марс» (ФНПЦ АО НПО
«Марс»)

Защита диссертации состоится «21» июня 2017 г. в 15 часов 00 минут на заседании диссертационного совета Д 212.277.01 при Ульяновском государственном техническом университете по адресу: 432027, г. Ульяновск, ул. Северный Венец, 32 (ауд. 211, Главный корпус).

С диссертацией можно ознакомиться в библиотеке Ульяновского государственного технического университета. Также диссертация и автореферат размещены в Internet на сайте УлГТУ – <http://www.ulstu.ru>.

Автореферат разослан « » _____ 2017 г.

Ученый секретарь
диссертационного совета,
доктор технических наук,
профессор



Смирнов Виталий Иванович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Проектирование технических систем представляет собой сложный процесс, основанный на последовательности принятия проектных решений, нуждающихся в инструментальных средствах, а также в лингвистическом обеспечении, включающем в себя предметную терминологию.

В процессе проектирования сложных технических систем (СТС) возникает потребность в унификации используемой терминологии с целью обеспечения связности процесса производства. Эта задача усложняется участием в процессе специалистов разных компетенций и уровней подготовки, а также различными форматами используемой документации.

Современные САПР обладают развитым лингвистическим обеспечением, традиционной частью которого являются тезаурусы, предметные словари и онтологии. Лингвистическое обеспечение САПР включает в себя совокупность языков, используемых при автоматизированном проектировании, терминов и определений, правил формализации естественного языка, а также методов сжатия и развертывания текстов (ГОСТ 22487-77).

Терминология предметной области в составе лингвистического обеспечения САПР имеет большое значение в процессе проектирования, т.к. она обеспечивает проект словарной базой и является основной для проектной и сопутствующей документации. В связи с этим, возникает задача извлечения терминологии из текстов по предметной области и ее представления. Представление терминологии имеет целью создание аппарата терминов и понятий, формирующего лингвистическое обеспечение САПР, в виде словарей, тезаурусов и онтологий.

Существенных результатов в области разработки интеллектуальных САПР с использованием онтологических форм представления терминологии добились С.В. Смирнов (разработка среды моделирования инженерных знаний, концептуальное моделирование) Н.М. Боргест (автоматизированное проектирование в области самолетостроения, разработка робота-проектанта), В.А. Витих (разработка методов и средств создания открытых мультиагентных систем для поддержки процессов принятия проектных решений), П.И. Соснин (вопросно-ответные рассуждения в среде автоматизированного проектирования, псевдо-кодовое программное управление в проектировании), В.М. Курейчик (генетические алгоритмы в интеллектуальных САПР, эволюционное проектирование) и др.

Начало разработки интеллектуальных программных систем, основанных на онтологическом подходе, связано с зарубежными исследователями, среди которых Т. Gruber, N. Guarino, M. Uschold, F. Bobillo, U. Straccia, G. Rzevski и т.д. Среди отечественных ученых проблемы использования онтологий и иных семантических моделей представления предметной терминологии в сфере автоматизированного проектирования исследуют Ю.А. Загорулько, Т.А. Гаврилова, Н.Г., В.Ф. Хоросhevский, А.С. Клещев, В.В. Грибова, Д.А. Новиков,

Актуальность темы

В настоящее время постоянный рост требований к эффективности и качеству лингвистического и информационного обеспечения автоматизированного проектирования технических систем в условиях слабой формализации поставленных задач, особенно на ранних этапах проектирования, предполагает необходимость системного решения ряда научных задач:

- необходимость унификации процесса автоматизированной обработки терминологии в лингвистическом обеспечении при решении проектных задач различного рода;
- необходимость разработки семантического базиса анализа представления терминологии в рамках решения задач автоматизированного проектирования;
- отсутствие интегративных концептуальных моделей, использующих различные подходы представления знаний о предметной области в форме тезауруса;
- необходимость одновременного использования разноаспектных описаний особенностей рассматриваемой предметной области;
- необходимость решения проблемы учета опыта решения схожих проектных задач;
- необходимость адаптации и расширения лингвистического и информационного обеспечения проектной деятельности вследствие изменения особенностей ПрО;
- необходимость решения проблемы учета нечеткости в рассуждениях проектировщика.

Поэтому актуальной является тема диссертации, посвященная разработке методов и средств представления терминологии в лингвистическом обеспечении САПР на основе интеграции нечетких онтологий и логического вывода в задачах автоматизированного проектирования.

Цель диссертационной работы

Целью диссертации является разработка и реализация эффективных моделей и алгоритмов представления терминологии в лингвистическом обеспечении САПР на основе интеграции нечетких онтологий и логического вывода, обеспечивающих снижение количества ошибок проектировщика.

Объектом исследования является лингвистическое обеспечение автоматизированного проектирования технических систем. В качестве примера исследованы процессы проектирования автоматизированной системы поточной линии сборки самолета, включающие ЛВС, линии станков с ЧПУ.

Задачи исследования

В соответствии с целью работы актуальными являются следующие задачи диссертационного исследования:

- провести сравнительный анализ современных средств представления

терминологии в лингвистическом обеспечении интеллектуальных САПР;

- разработать онтологически-ориентированную систему извлечения терминологии из проектных документов с использованием сформированного ядра предметной онтологии;

- провести сравнительный анализ современных интеллектуальных методов логического вывода в САПР, выявить их возможности и ограничения в плане решения задач автоматизированного проектирования;

- рассмотреть возможность применения методов онтологического анализа для решения задач оценки состояния сложного технического объекта рассматриваемой предметной области с целью поддержки принятия проектных решений на экспертном уровне;

- исследовать возможность интеграции методов теории нечетких систем и онтологического анализа состояния сложной технической системы;

- разработать онтологическую модель представления набора продукций и алгоритм логического вывода рекомендаций на экспертном уровне в задачах автоматизированного проектирования;

- разработать алгоритм расширения процедурной составляющей базы знаний интеллектуальной САПР с использованием механизма прецедентов;

- разработать алгоритм расширения декларативного описания ПрО с использованием сформированного ядра онтологии интеллектуальной САПР посредством извлечения терминологии из проектных документов;

- разработать автоматизированную систему (АС) поддержки проектирования, реализующую алгоритм интеграции механизмов логического вывода и онтологического представления экспертных знаний с элементами нечеткости, а также использующей механизм учета прецедентов с целью расширения ядра онтологии;

- провести вычислительные эксперименты, позволяющие оценить эффективность предложенных моделей и алгоритмов в процессе проведения концептуального проектирования сложных технических систем;

- внедрить результаты исследований в практику процесса проектирования технических систем предприятий региона.

При решении задачи оценки эффективности предложенных моделей и алгоритмов необходима адаптация условий проведения экспериментов под специфику решаемых задач.

Методы исследования

В диссертационной работе применяются методы онтологического анализа, дескрипционной логики, теории нечетких систем, а также объектно-ориентированного программирования при построении программного комплекса.

Научная новизна

Научная новизна результатов исследования заключается в следующем:

1. Предложена онтологически-ориентированная методика извлечения

терминологии из проектных документов, отличающаяся от известных совместным использованием статистических методов, лингвистических шаблонов и базового ядра предметной онтологии;

2. Разработан алгоритм расширения онтологического представления терминологии предметной области технической системы, отличающийся от известных построением вложенных связей термина-кандидата, извлеченного из проектного документа, до опорного класса базового ядра онтологии;

3. Разработан алгоритм расширения процедурной составляющей базы знаний интеллектуальной САПР, отличающийся от известных использованием в качестве прецедента совокупности SWRL-правил и правил решения задачи проектирования;

4. Предложен метод интеграции нечеткого логического вывода и нечеткого онтологического представления терминологии предметной области технической системы, отличающийся от известных использованием иерархического нечеткого вывода и модели онтологии класса FuzzyOWL;

5. Разработана программная система поддержки проектирования, отличающаяся наличием специальной компоненты интеллектуальной САПР для представления терминологии лингвистического обеспечения.

Теоретическая значимость работы

Теоретическая значимость работы заключается в разработке и реализации новых эффективных моделей и алгоритмов представления терминологии в лингвистическом обеспечении САПР на основе интеграции нечетких онтологий и логического вывода, обеспечивающих снижение количества ошибок проектировщика.

Практическая значимость работы

Разработанная интеллектуальная компонента САПР, реализующая предложенные модели и алгоритмы, была использована в рамках проекта разработки проектной документации Автоматизированной системы управления (АСУ) поточной линии сборки АО «Авиастар-СП». Также данная компонента используется при построении универсальной базы знаний учебного центра «Юнитех» агентства переводов «Юнитранс».

Основания для выполнения работы

Результаты диссертационной работы использовались в ряде НИОКР, выполненных в Ульяновском государственном техническом университете, направленных на решение научно-технических задач. К наиболее важным результатам следует отнести:

1. Участие в выполнении гранта РФФИ №13-01-00324 «Исследование формальных методов грануляции слабоструктурированных информационных ресурсов на основе онтологии предметной области».

2. Участие в выполнении гранта РФФИ №15-41-02413 «Интеллектуальный анализ временных рядов на основе нечетких онтологий, извлекаемых из Интернет-ресурсов».

3. Участие в выполнении гранта РФФИ № 16-47-730742 «Интеграция онтологических моделей и проектных диаграмм при концептуальном проектировании сложных информационных систем».

4. Участие в реализации гранта по программе УМНИК в области информационных технологий при поддержке Фонда содействия развитию малых форм предприятий в научно-технической сфере.

5. Участие в реализации государственного задания №2014/232 на выполнение государственных работ в сфере научной деятельности Минобрнауки России по проекту «Разработка нового подхода к интеллектуальному анализу слабоструктурированных информационных ресурсов».

Достоверность результатов диссертационной работы

Достоверность научных положений, выводов и рекомендаций подтверждена результатами вычислительных экспериментов, а также результатами использования материалов диссертации в работе компаний ООО «Юнитранс» и АО «Авиастар – СП».

Основные положения, выносимые на защиту

1. Онтологически-ориентированная методика извлечения терминологии из проектных документов с использованием базового ядра предметной онтологии;

2. Алгоритм расширения онтологического представления терминологии предметной области технической системы с использованием базового ядра онтологии интеллектуальной САПР посредством извлечения терминологии из проектных документов;

3. Алгоритм расширения процедурной составляющей базы знаний интеллектуальной САПР с использованием механизма прецедентов;

4. Метод интеграции нечеткого логического вывода и нечеткого онтологического представления терминологии предметной области технической системы;

5. Программная система поддержки проектирования, являющаяся компонентой интеллектуальной САПР для представления терминологии лингвистического обеспечения.

Апробация работы

Основные положения и результаты диссертационной работы докладывались, обсуждались и получили одобрение на следующих конференциях, семинарах и симпозиумах: VII и VIII Международных научно-практических конференциях «Интегрированные модели и мягкие вычисления в искусственном интеллекте» (г. Коломна, 2013, 2015 гг.); I Международной Поспеловской летней школе-семинаре для студентов, магистрантов и аспирантов «Методы и технологии гибридного и синергетического искусственного интеллекта» (г. Светлогорск, 2014 г.); II Международном Поспеловском симпозиуме «Гибридные и синергетические интеллектуальные системы» (г. Светлогорск, 2014 г.); VI Всероссийской научно-практической конференции «Нечеткие системы и мягкие вычисления» (г. Санкт-Петербург,

2014 г.); XIV национальной конференции по искусственному интеллекту с международным участием «КИИ-2014» (г. Казань, 2014 г.); V, VI и VII Международных научно-технических конференциях «Открытые семантические технологии проектирования интеллектуальных систем» (г. Минск, 2015, 2016, 2017 гг.); I Международной научной конференции «Интеллектуальные информационные технологии в технике и на производстве» (г. Сочи, 2016 г.), XV национальной конференции по искусственному интеллекту с международным участием «КИИ-2016» (г. Смоленск, 2016 г.).

Научные публикации

По результатам работы было опубликовано 35 статей, 10 из которых в журналах из перечня ВАК, а также 1 статья в издании, индексируемом в Scopus. Получены 3 свидетельства о государственной регистрации программ для ЭВМ.

Личный вклад

Все результаты, составляющие содержание диссертации, получены автором самостоятельно. Подготовка к публикации некоторых результатов проводилась совместно с соавторами, причем вклад соискателя был определяющим.

Структура и объем диссертации

Диссертация состоит из введения, четырех глав, заключения, списка использованной литературы и приложений. Основное содержание работы изложено на 190 страницах, включая 44 рисунка и 15 таблиц. Список использованных источников состоит из 141 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** рассмотрена актуальность выполненного исследования, формулируются цель и задачи работы, определяются теоретическая значимость и практическая ценность результатов исследования, а также положения, выносимые на защиту.

В **первой главе** приводится определение и структура лингвистического обеспечения САПР, а также описание современных тенденций интеллектуализации САПР. Помимо этого, рассмотрена архитектура экспертных систем как элемента интеллектуальных САПР, и приведены основные средства и алгоритмы представления предметной терминологии в лингвистическом обеспечении САПР.

Выявлена необходимость введения нечеткости в процесс логического вывода, приведены основные алгоритмы нечеткого логического вывода, а также проведен анализ современных САПР, реализующих рассмотренные алгоритмы. В качестве отдельного класса представления предметной терминологии рассмотрен класс нечетких онтологий. Помимо этого, приведена методика разработки нечеткой онтологии, формируемой из различных источников, а также рассмотрена методика представления нечетких онтологий FuzzyOWL. Также представлены основные статистические и лингвистические алгоритмы

извлечения терминологии из текстовых, в том числе и проектных, документов при решении задачи расширения ядра онтологии интеллектуальной САПР.

Во **второй главе** приводится описание модели онтологии, используемой в качестве формы представления предметной терминологии в процессе проектирования сложных технических систем (СТС).

Предметная область проектирования сложных систем накладывает определенные требования к структуре прикладной онтологии. Онтологию предметной области запишем в виде кортежа:

$$O = (T, R^D, F),$$

где T – термины прикладной области, которую описывает онтология. Множество терминов представлено в виде:

$$T = \{C, In_C\},$$

где C – множество классов онтологии, In_C – множество объектов классов онтологии.

R^D – множество отношений между объектами онтологии, F – множество функций интерпретации (аксиоматизации), заданных на терминах и/или отношениях онтологии, при этом:

$$F = \{F_C, F_{In}, F_R\},$$

где $F_C: \{C_{Rule}\} \rightarrow \{C\}$ – функция, сопоставляющая набору классов, используемых в системе продукционных SWRL-правил онтологии множество классов OWL-онтологии;

$F_{In}: \{In_{Rule}\} \rightarrow \{In\}$ – функция, сопоставляющая набору объектов классов, используемых в системе продукционных SWRL-правил онтологии множество классов объектов OWL-онтологии;

$F_R: \{R_{Rule}\} \rightarrow \{R^D\}$ – функция, сопоставляющая набору отношений объектов и типа данных, используемых в системе продукционных SWRL-правил онтологии множество отношений классов OWL-онтологии.

Нечеткие SWRL-правила представляют собой дизъюнкты Хорна. Приведем пример нечеткого правила:

*Ситуация(?x) \wedge включаетПоказатель(?x, Фоновая_нагрузка) \wedge
имеетТенденцию(Фоновая_нагрузка, Рост) \wedge
включаетПоказатель(?x, Утилизация_канала_связи) \wedge
имеетТенденцию(Утилизация_канала_связи, Спад) \wedge
включаетПоказатель(?x, Число_коллизий_в_сети) \wedge
имеетТенденцию(Число_коллизий_в_сети, Рост) \rightarrow предполагает(?x,
Проблема_архитектуры)*

Взаимосвязь онтологии ПрО и набора SWRL-правил в процессе логического вывода осуществляется посредством формирования запросов к онтологии, генерируемых системой анализа при выполнении набора правил (рис. 1).

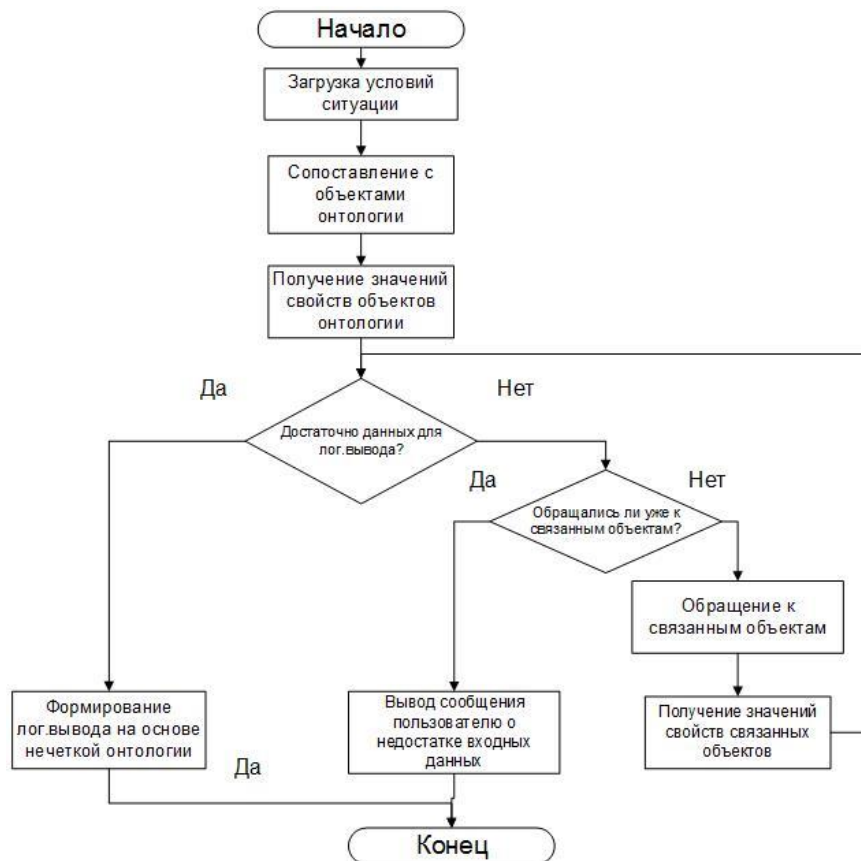


Рисунок 1 Алгоритм логического вывода на основе онтологии

Логический вывод проектных рекомендаций с использованием онтологии, содержащей нечеткие отношения, классы, а также объекты классов, имеет существенные отличия по сравнению с извлечением знаний из четкой онтологии (рис. 2).

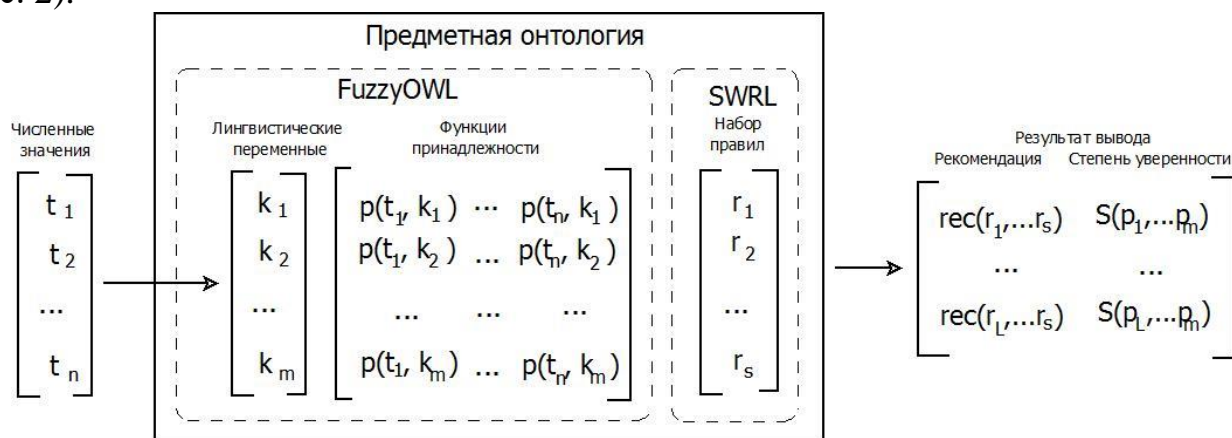


Рисунок 2 Схема логического вывода на основе нечеткой онтологии

Алгоритм взаимодействия нечеткой FuzzyOWL-онтологии и системы нечетких SWRL-правил включает в себя следующие этапы:

1. На вход алгоритма поступают значения характеристик анализируемой сложной системы, носящие качественный (в некоторых случаях – нечеткий) или количественный характер («большой», «малый», «быстрый рост» и т. д.).
2. Композиция входных значений и условных частей SWRL-правил.

3. Импликация множества пар – «рекомендация – показатель степени уверенности».

При этом показатель степени уверенности является функцией от поступающих значений функций принадлежности, относящихся к заложенным в соответствующем правиле значениям предметной области:

$$S_{Rule} = \min \mu(t_i; k_j),$$

где t – термин проблемной области, присутствующий в сработавшем SWRL-правиле; k – лингвистическая переменная; $j=[1..m]$; $\mu(t,k)$ – значение соответствующей функции принадлежности.

В следующем пункте **второй главы** описан вариант модификации алгоритмов поддержки принятия проектных решений путем учета опыта решения подобных проектных задач, посредством использования механизма учета прецедентов. Формальная модель системы вывода, основанного на прецедентах, может быть представлена в виде упорядоченной тройки:

$$CBR = \{Cases, I, S_{CASE}\},$$

где Cases – база прецедентов, I – онтология ПрО (нечеткая), S_{CASE} – алгоритм поиска подходящего прецедента. Структура прецедента может быть представлена в следующем виде:

$$Case = \{Index_{CASE}, D(Index_{CASE}), Eff(D(Index_{CASE}))\},$$

где $Index_{CASE}$ – индекс прецедента, т.е. описание начальной ситуации, $D(Index_{CASE})$ – множество решений поставленной задачи, $Eff(D(Index_{CASE}))$ – множество оценок эффективности принятого решения задачи.

Так как прецеденты, в отличие от правил, оперируют не переменными, а конкретными объектами классов, а также значениями свойств этих объектов, то, начальная ситуация описывается следующим множеством:

$$Index_{CASE} = \{I_f, P_f\},$$

где I_f – множество объектов классов онтологии, P_f – множество значений свойств соответствующих объектов.

Главное особенностью модифицированного алгоритма вывода рекомендаций является проведение параллельного и независимого логического вывода результатов анализа на основании базы правил и базы прецедентов.

Также **во второй главе** представлено описание двух разработанных алгоритмов расширения ядра онтологии предметной области посредством анализа проектной документации.

Тезаурусный алгоритм извлечения терминологии предполагает непосредственный поиск вхождений лемм поступающих на вход слов и их сочетаний среди терминов, определенных в онтологии. Алгоритм определения степени близости слов/сочетания слов терминам проблемной области согласно тезаурусному алгоритму предполагает:

1. Оценку степени близости поступающего на вход алгоритма слова/сочетания слов каждому объекту онтологии без учета онтологического критерия оценки.

2. Определение опорного объекта онтологии, наиболее близко ассоциирующегося с входным одно-/многословием. Схема разработанного

алгоритма представлена на рисунке 3.

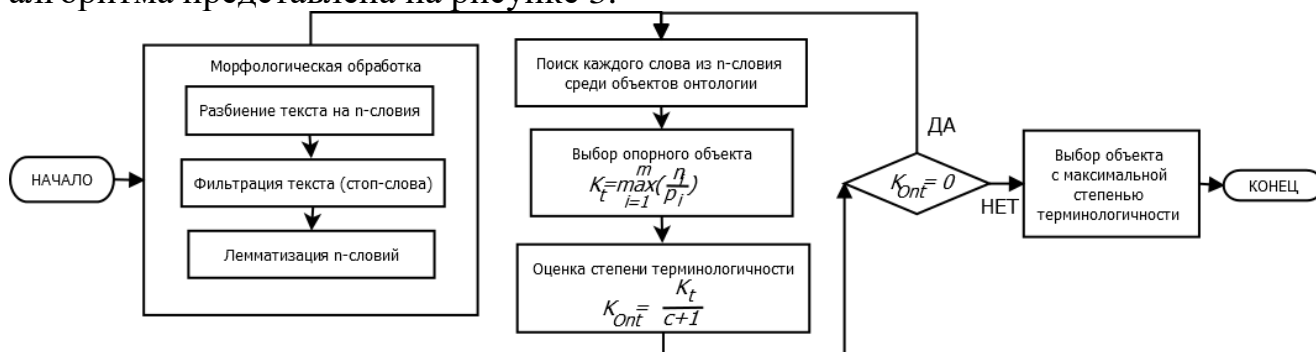


Рисунок 3 Схема тезаурусного алгоритма извлечения терминологии

Опорный объект ядра расширяемой онтологии имеет следующую степень близости по отношению к входному слову/сочетанию слов:

$$k_t = \max_{i=1}^m \left(\frac{n_i}{p_i} \right),$$

где m – количество всех объектов ядра онтологии; n_i – число слов из леммы входного многословия, найденных в лемме объекта онтологии; p_i – общее число слов в лемме объекта онтологии.

Общая схема оценки степени близости слов/сочетания слов терминам проблемной области согласно тезаурусному алгоритму приведена на рисунке 4.



Рисунок 4 Тезаурусный алгоритм. Поиск опорного объекта

При этом порядок следования слов многословия в опорном объекте должен сохраняться.

Степень близости слова/сочетания слов терминам рассматриваемой предметной области оценивается по следующей формуле:

$$k_{Ont} = \frac{k_t}{c+1},$$

где k_t – результат первого этапа анализа; p_i – число отношений, связывающих опорный объект онтологии с ближайшим объектом, имеющим истинное значение свойства «является Термином».

Алгоритм вложенных связей позволяет извлечь термины из текста посредством их сопоставления с имеющимися объектами и сочетаниями лемм соответствующих объектов с помощью отношений R_{add} . Пример: свойства «имеет Отношение» и «является Частью», посредством которых алгоритм

позволяет формировать словосочетания естественным образом.

$$T_1 + R_1 + t_2 + R_2 + \dots + t_i + R_j + \dots + t_m + R_n,$$

где $R_i \in R_{add}$, $t_j \in T$, T — термины прикладной области, которую описывает онтология. Схема разработанного алгоритма представлена на рисунке 5.

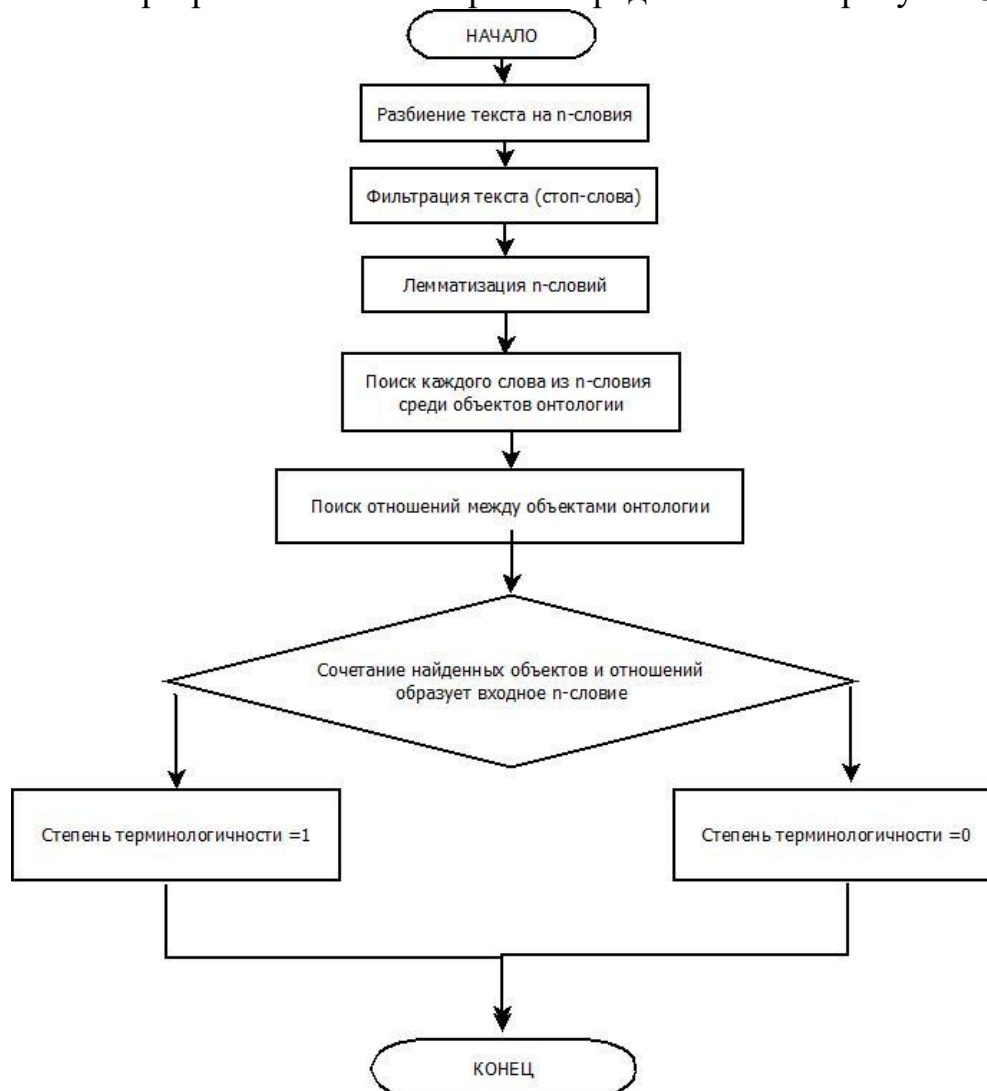


Рисунок 5 Схема алгоритма вложенных связей извлечения терминологии из текста

Пример формирования многословий с помощью свойства «имеетОтношение»:

1. Найденная цепочка объектов: «Вращение» + «имеетОтношение» + «Двигатель» + «имеетОтношение» + «ПеременныйТок».
2. Объединение лемм объектов онтологии: «вращение двигатель переменный ток».
3. Термин, извлекаемый из обрабатываемого текста: «вращение двигателя переменного тока».

Пример формирования многословий с помощью свойства «являетсяЧастью»:

1. Найденная цепочка объектов: «Подшипник» + «являетсяЧастью» + «Шпиндель»;
2. Объединение лемм объектов онтологии: «подшипник шпиндель»;
3. Термин, извлекаемый из обрабатываемого текста: «подшипник шпинделя».

В третьей главе приведена архитектура и функциональные возможности разработанной интеллектуальной компоненты САПР, реализующей алгоритм интеграции механизмов логического вывода и онтологического представления терминологии с элементами нечеткости, а также использующей механизм учета прецедентов и анализа проектной документации с целью расширения ядра онтологии.

В четвертой главе проводится анализ адекватности разработанных моделей и методов на основе вычислительных экспериментов.

На рис. 6 выделены фрагменты отдельных технологических процессов сборки самолетов для производства тяжелого военно-транспортного самолета Ил-76МД- 90А на АО «Авиастар-СП», к которым в рамках данного исследования, были применены разработанные методы и алгоритмы.

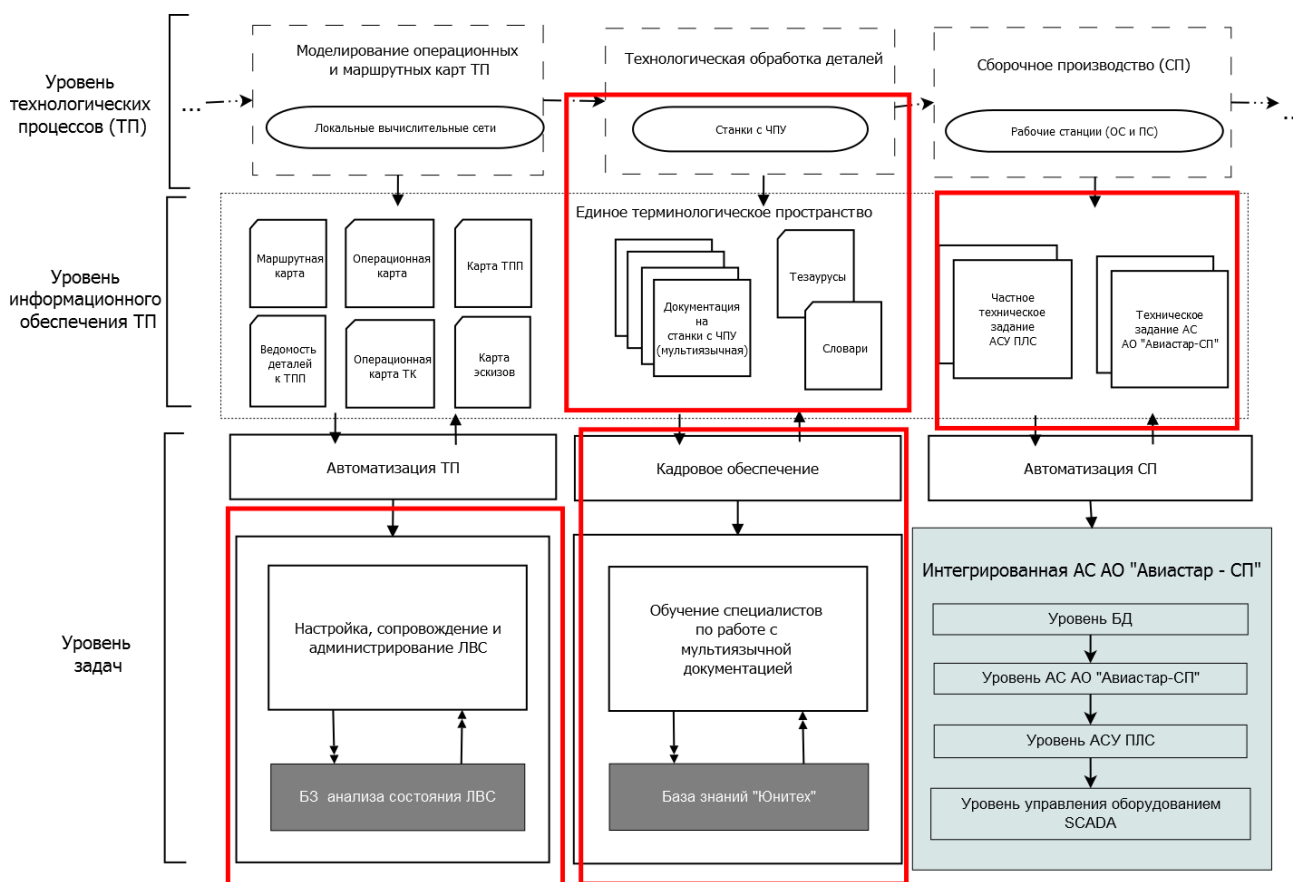


Рисунок 6 Фрагмент схемы СТС производственного типа на примере поточной линии сборки самолетов Ил-76МД- 90А на АО «Авиастар-СП»

В первом пункте главы представлены результаты экспериментов по анализу состояния ЛВС в условиях искусственного повышения трафика с использованием разработанной интеллектуальной компоненты САПР, реализующей предложенные алгоритмы интеграции онтологических и продукционных моделей представления предметной терминологии.

Сравнительный анализ результатов экспериментов с использованием четких и нечетких онтологий представлен на рис. 7.

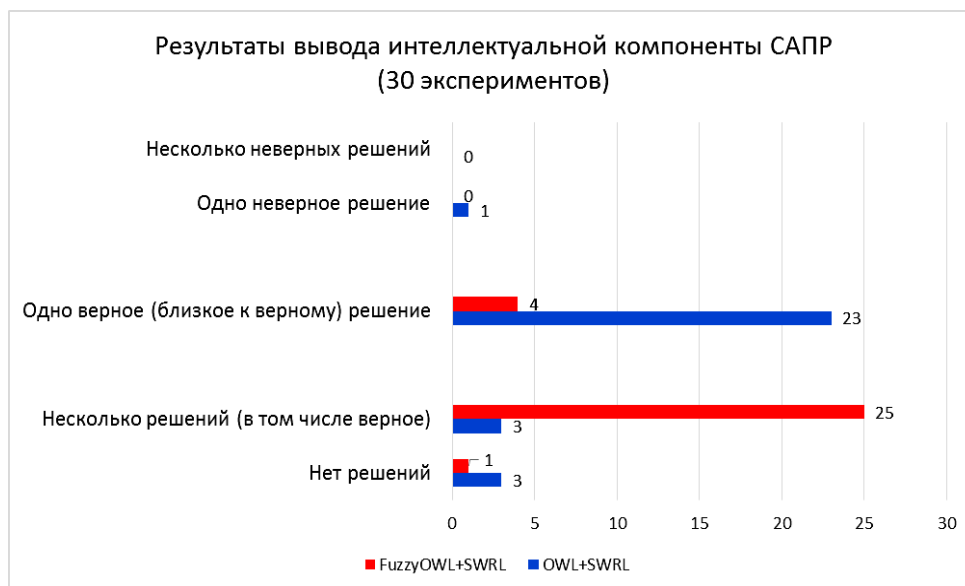


Рисунок 7 Сравнение применения алгоритма с использованием четких и нечетких онтологий

Как видно из результатов проведенных экспериментов, благодаря включению нечеткости в описание предметной области, значительно снижаются риски потери возможных результатов работы блока логического вывода, увеличивается гибкость процесса вывода, а пользователю предоставляется гораздо широкий выбор вариантов, которые при этом упорядочены по мере их релевантности с точки зрения совокупности экспертных знаний и предметной терминологии, заложенных в онтологии и наборе правил.

В таблице 1 представлены результаты экспериментов с использованием модифицированного алгоритма с интеграцией механизма прецедентов.

Таблица 1 – Результаты экспериментов с использованием модифицированного алгоритма с интеграцией механизма прецедентов

Номер/ (Партия экспериментов)	Количество экспериментов			Количество использованных прецедентов из базы
	Результат эксперимента - нет решений	Результат эксперимента - несколько решений	Результат эксперимента - одно верное решение	
1 (1-10)	1	8	1	1
2 (11-20)	-	9	1	3
3 (21-30)	-	8	2	4

Как видно из результатов, с каждой следующей партией экспериментов вывод правильного решения на основе прецедентов становится чаще, так как система проходит обучение, благодаря внесению верных решений в CBR-базу.

Несмотря на то, что включение нечеткости в описание предметной области значительно снижает риски потери возможных результатов работы блока логического вывода и увеличивает гибкость процесса вывода, наличие большого количества вариантов решения задачи не дает пользователю в полной мере

положиться на какой-либо из предложенных вариантов.

Помимо этого, представлены результаты экспериментов по извлечению терминологии из проектной документации на основании разработанных алгоритмов при построении универсальной базы знаний учебного центра «Юнитех» агентства переводов «Юнитранс».

Проведена оценка полученных результатов в сравнении с результатами применения статистических методов извлечения терминологии в соответствии с метриками, применимыми к задаче классификации (Precision, Recall, F-мера). Результаты оценки приведены в таблице 2.

Таблица 2 – Сравнительная характеристика результатов извлечения терминологии онтологическим и статистическими алгоритмами

Кол-во слов в словоупотреблении	Термины	Кандидаты	Верно	P	R	F ₁
Алгоритм вложенных связей						
1	294	168	134	0,80	0,46	0,58
2	631	431	372	0,86	0,59	0,70
3	361	370	327	0,88	0,91	0,89
Frequency						
1	294	134	123	0,92	0,42	0,58
2	631	469	347	0,74	0,55	0,63
3	361	334	267	0,80	0,74	0,77
TF*IDF						
1	294	147	138	0,94	0,47	0,63
2	631	456	328	0,72	0,52	0,60
3	361	277	166	0,60	0,46	0,52
C-Value						
1	294	120	112	0,93	0,38	0,54
2	631	789	316	0,40	0,50	0,44
3	361	295	162	0,55	0,45	0,50

Исходя из полученных выше результатов, следует отметить, что статистические методы показали значительно более высокие результаты при извлечении однословий. Это связано с тем, что онтологический метод в первую очередь извлекает связанные с имеющимися в базе знаний термины. В случае же двух- и трехсловий, онтологический метод позволяет извлекать меньше неверных терминов. В первую очередь, это связано с тем, что статистические методы больше ориентируются на частоту употреблений словосочетаний, вне зависимости от привязки к особенностям ПрО, и могут извлекать общенаучные термины, а также термины других ПрО.

Также в главе 4 приведены результаты отраслевой подготовки специалистов по мультязычной проектной документации с использованием универсальной базы знаний «Юнитех», полученной в результате применения предложенной методики извлечения терминологии из проектных документов.

В течение 6 месяцев специалисты изучали «Учебный справочник технического переводчика: станки с ЧПУ», эксплуатационную документацию, видеоматериалы и дополнительную информацию в сети Интернет, выполняли

рабочие переводы. Одним из пунктов отраслевой подготовки было изучение базы знаний «Юнитех», формируемой на основе применения предложенных семантических алгоритмов. В начале стажировки и по прошествии каждого месяца обучения стажеры выполняли перевод контрольных образцов, что позволило в динамике отследить показатели качества перевода.

Результаты зависимости количества приобретенных отраслевых знаний от месяца подготовки представлены на рисунке 8.

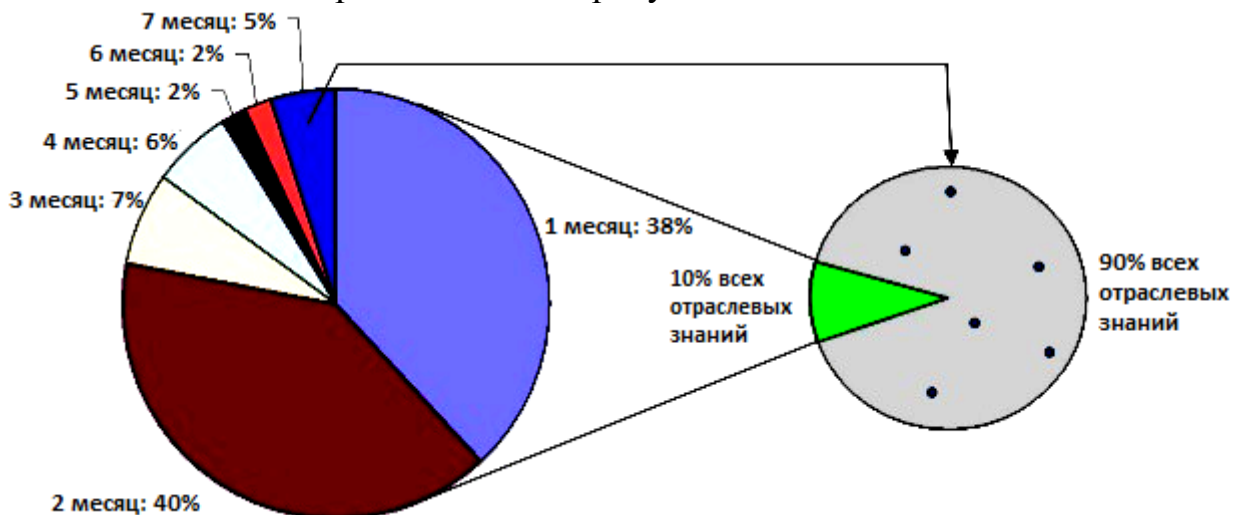


Рисунок 8 Зависимость количества приобретенных отраслевых знаний от месяца подготовки

Как видно из результатов, использование в процессе обучения специалистов по работе с мультязычной проектной документацией унифицированной базы знаний, полученной путем извлечения терминологии предложенными онтологическими алгоритмами, позволяет обеспечить специалиста 78% приобретаемых отраслевых знаний за первые два месяца обучения.

Помимо этого, были проведены эксперименты по применению предложенных алгоритмов извлечения терминологии с использованием предметной онтологии в процессе разработки технической документации Автоматизированной системы управления Поточной линии сборки АО «Авиастар-СП».

Порядок проведения экспериментов:

1. Без помощи разработанной системы была составлена начальная версия ТЗ по соответствующему проекту.
2. Было разработано ядро OWL-онтологии по предметной области АСУ Поточной линии сборки тяжелых самолетов. Ядро состоит из 140 классов и объектов классов, а также 26 отношений и имеет 3 уровня иерархии.
3. Было выбраны наиболее часто используемых в процессе обучения специалистов в данной области ресурсы и учебные пособия.
4. С целью исправления семантических ошибок и расширения терминологической составляющей ТЗ, к данным ресурсам были применены разработанные алгоритмы онтологического расширения ядра онтологии на основе извлечения терминов. Лучшие результаты показал алгоритм вложенных

связей, эффект использования результатов применения которого в процессе разработки промежуточных и конечных версий ТЗ приведен в табл. 3. В качестве характеристик, определяющих семантическую адекватность разработанного ТЗ были выделены следующие показатели: количество определений и число исправленных смысловых ошибок.

Таблица 3 – Результаты экспериментов по использованию тезаурусного алгоритма в процессе разработки ТЗ АСУ Поточной линии сборки

Характеристика	Начальная версия	v4	v9	v13	v16
Количество определений	9	15	16	18	20
Количество смысловых ошибок	13	3	2	3	-

Как видно из результатов экспериментов, применение разработанных алгоритмов уже после первой итерации позволило значительно снизить количество семантических ошибок в тексте ТЗ, а количество определений, включенных в структуру ТЗ, возросло в 1,7 раза.

В заключении сформулированы основные выводы и результаты диссертационной работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В ходе диссертационного исследования получены следующие результаты:

- 1) проведен сравнительный анализ современных средств представления терминологии в лингвистическом обеспечении интеллектуальных САПР;
- 2) разработана онтологически-ориентированная система извлечения терминологии из проектных документов с использованием сформированного ядра предметной онтологии;
- 3) проведен сравнительный анализ современных интеллектуальных методов логического вывода в САПР, выявить их возможности и ограничения в плане решения задач автоматизированного проектирования;
- 4) рассмотрена возможность применения методов онтологического анализа для решения задач оценки состояния сложного технического объекта рассматриваемой предметной области с целью поддержки принятия проектных решений на экспертном уровне;
- 5) исследована возможность интеграции методов теории нечетких систем и онтологического анализа состояния сложной технической системы;
- 6) разработана онтологическая модель представления набора продукций и алгоритм логического вывода рекомендаций на экспертном уровне в задачах автоматизированного проектирования;
- 7) разработан алгоритм расширения процедурной составляющей базы знаний интеллектуальной САПР с использованием механизма прецедентов;

8) разработан алгоритм расширения декларативного описания предметной области (ПрО) с использованием сформированного ядра онтологии интеллектуальной САПР посредством извлечения терминологии из проектных документов;

9) разработана АС поддержки проектирования, реализующую алгоритм интеграции механизмов логического вывода и онтологического представления экспертных знаний с элементами нечеткости, а также использующей механизм учета прецедентов с целью расширения ядра онтологии;

10) проведены вычислительные эксперименты, основными результатами которых являются:

- применение разработанных алгоритмов интеграции продукционной и онтологической форм представления знаний с элементами нечеткости позволяет снизить риски потери возможных результатов работы блока логического вывода, а также увеличивает гибкость процесса вывода рекомендации, предоставляя пользователю более широкий выбор вариантов, упорядоченных по степени их релевантности (более, чем в 5 раз расширен список возможных решений);

- использование в процессе обучения специалистов по работе с мультязычной проектной документацией унифицированной базы знаний, полученной путем извлечения терминологии предложенными онтологическими алгоритмами, позволяет обеспечить специалиста 78% приобретаемых отраслевых знаний за первые два месяца обучения;

- использование алгоритмов извлечения терминологии из проектной документации по тематике станков с ЧПУ с использованием предметной онтологии повышает эффективность извлечения двух- и трехсловных терминологических сочетаний на 20,5% и 33% соответственно.

11) результаты исследований внедрены в практику процесса проектирования технических систем предприятий региона.

СПИСОК ПУБЛИКАЦИЙ

Статьи, опубликованные в журналах, рекомендованных Перечнем ВАК России:

1. Мошкин В.С., Пирогов А.Н., Тимина И.А., Шишкин В.В., Ярушкина Н.Г. Интеллектуальный анализ проектных и терминологических метрик в управлении проектами // Автоматизация процессов управления. – 2016. – № 4 (46). – С. 84–91.

2. Гуськов Г.Ю., Мошкин В.С., Наместников А.М., Филиппов А.А., Ярушкина Н.Г. Разработка многоагентной системы извлечения знаний из гетерогенных источников // Радиотехника. – 2016. – № 9. – С. 57–63.

3. Армер А.И., Мошкин В.С. Подход к формированию наборов эталонов речевых команд с использованием онтологии // Онтология проектирования. – 2016. – Т. 6, №3(21). – С. 270–277.

4. Армер А.И., Мошкин В.С., Ярушкина Н.Г. Использование онтологии для формирования наборов эталонов речевых команд в задаче распознавания речевых команд на фоне шумов // Радиотехника. – 2016. – № 9. – С. 72–76.

5. Ярушкина Н.Г., Мошкин В.С. Подход к обучению онтологии на основе гибридизации алгоритмов извлечения знаний из текстов и механизма прецедентов // Вестник Ростовского государственного университета путей сообщения. – 2016. – № 2 (62). – С. 78–83.

6. Барабанова Е.Ю., Башаев В.А., Клейн В.В., Мошкин В.С. Построение информационной поддержки автоматизированного проектирования управляющих программ для станков с ЧПУ // Радиотехника. – 2015. – № 6. – С. 63–67.

7. Ярушкина Н.Г., Мошкин В.С. Применение алгоритма логического вывода на основе FuzzyOWL-онтологии // Радиотехника. – 2015. – № 6. – С. 68–72.

8. Андреев И.А., Башаев В.А., Клейн В.В., Мошкин В.С., Ярушкина Н.Г. Семантическая метрика терминологичности на основе онтологии предметной области // Автоматизация процессов управления. – 2014. – № 4 (38). – С. 76–84.

9. Ярушкина Н.Г., Мошкин В.С. Применение онтологического подхода к анализу состояния локальной вычислительной сети // Радиотехника. – 2014. – № 7. – С. 120–124.

10. Мошкин В.С., Ярушкина Н.Г. Система онтологического анализа временных рядов // Автоматизация процессов управления. – 2014. – № 2 (36). – С. 78–85.

приндексированные в SCOPUS

11. Yarushkina N., Moshkin V., Klein V., Andreev I, Beksaeva E.: Hybridization of Fuzzy Inference and Self-learning Fuzzy Ontology-Based Semantic Data Analysis. In: Proceedings of the First International Scientific Conference “Intelligent Information Technologies for Industry” (IITI’16), pp. 277–285 (2016)

в иных изданиях

12. Moshkin V.S., Filippov A.A., Yarushkina N.G. Integration of an applied ontology and wiki-resources in the context of the unified knowledge base // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2017): материалы VII Междунар. научн.техн. конф. (Минск, 16-18 февраля 2017 г.) / редкол.: В. В. Голенков (отв. ред.) [и др.]. – Минск: БГУИР. – 2017. – С. 107–110.

13. Андреев И. А., Бексаева Е. А., Клейн В. В., Мошкин В. С., Серков И. П. Лингвистический подход к автоматизированному построению предметной онтологии // Прикладные информационные системы: третья Всероссийская НПК (г. Ульяновск, 30 мая – 12 июня 2016 г.): сборник научных трудов / под ред. Е. Н. Эгова. – Ульяновск : УлГТУ, 2016. – С. 256–263.

14. Мошкин В.С., Ярушкина Н.Г. Гибридизация алгоритмов извлечения знаний из текстов и механизма прецедентов в процессе расширения онтологии // Пятнадцатая национальная конференция по искусственному интеллекту с

- международным участием КИИ – 2016 : труды конференции. Т.1. – Смоленск : Универсум, 2016. – С. 80–87.
15. Мошкин В.С., Филиппов А.А., Ярушкина Н.Г. Разработка многоагентной базы знаний в рамках технологической платформы Athene // Пятнадцатая национальная конференция по искусственному интеллекту с международным участием КИИ – 2016 : труды конференции. Т.3. – Смоленск : Универсум, 2016. – С. 261–269.
16. Филиппов А.А., Мошкин В.С., Ярушкина Н.Г., Шалаев Д.О. Разработка агента для управления знаниями с использованием графовой базы данных // Гибридные и синергетические интеллектуальные системы: материалы III Всероссийской Поспеловской конференции с международным участием. Под редакцией А. В. Колесникова. – 2016. – С. 360-367.
17. Мошкин В.С., Ярушкина Н.Г. Интеграция нечетких онтологий и базы прецедентов в единой семантической среде // В сборнике: Гибридные и синергетические интеллектуальные системы: материалы III Всероссийской Поспеловской конференции с международным участием. Под редакцией А. В. Колесникова. – 2016. – С. 411–419.
18. Мошкин В.С., Ярушкина Н.Г. Модифицированный метод логического вывода знаний на основе нечеткой онтологии и базы прецедентов // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2016) : материалы VI Междунар. научно-техн. конф. (Минск, 18-20 февраля 2016 г.) / редкол.: В. В. Голенков (отв. ред.) [и др.]. – Минск : БГУИР. – 2016. – С. 265–270.
19. Филиппов А.А., Мошкин В.С., Шалаев Д.О., Ярушкина Н.Г. Единая онтологическая платформа интеллектуального анализа данных // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2016) : материалы VI Междунар. научно-техн. конф. (Минск, 18-20 февраля 2016 г.) / редкол.: В. В. Голенков (отв. ред.) [и др.]. – Минск: БГУИР. – 2016. – С. 77–82.
20. В.С. Мошкин Особенности построения нечетких онтологий группой экспертов// Информатика и вычислительная техника : сборник научных трудов 7-й Всероссийской научно-технической конференции аспирантов, студентов и молодых ученых ИВТ-2015 / под ред. Н. Н. Войта. – Ульяновск : УлГТУ, 2015. – С. 368–371.
21. Клейн В.В., Мошкин В.С., Ярушкина Н.Г. Разработка семантической метрики терминологичности лексических единиц в решении задачи формирования словарей // Прикладные информационные системы : Вторая Всероссийская НПК (г. Ульяновск, 25 мая – 7 июня 2015 г.): сборник научных трудов / под ред. Е.Н. Эгова. – Ульяновск : УлГТУ, 2015. – С. 199–206.
22. Мошкин В.С., Ярушкина Н.Г. Построение нечеткой онтологии тестирования состояния ЛВС группой экспертов // Шестая международная конференция «Системный анализ и информационные технологии» САИТ – 2015 (15–20 июня 2015 г., г. Светлогорск, Россия) : труды конференции. В 2-х т. – Т. 2. М. : – С. 107–111.

23. Андреев И.А., Башаев В.А., Клейн В.В. Мошкин В.С. Определение вероятности терминологичности словоупотреблений в текстах конкретной предметной области // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник научных трудов VIII Международной научно-практической конференции (Коломна, 18-20 мая 2015 г.). В 2-х томах. Т.2 – М. : Физматлит, 2015. – С. 764–773.

24. Мошкин В.С., Ярушкина Н.Г. Логический вывод на основе нечетких онтологий // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник научных трудов VIII Международной научно-практической конференции (Коломна, 18-20 мая 2015 г.). В 2-х томах. Т1. – М. : Физматлит, 2015. – С. 259–267.

25. Андреев И.А., Башаев В.А., Клейн В.В., Мошкин В.С., Ярушкина Н.Г. Оценка терминологичности лексических единиц на основе онтологии предметной области // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2015) : материалы V Междунар. научно-техн. конф. (Минск, 19-21 февраля 2015 г.) / редкол.: В. В. Голенков (отв. ред.) [и др.]. – Минск : БГУИР. – 2015. – С. 395–400.

26. Мошкин В.С., Ярушкина Н.Г. Методики построения нечетких онтологий сложных предметных областей // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2015) : материалы V Междунар. научно-техн. конф. (Минск, 19-21 февраля 2015 г.) / редкол.: В. В. Голенков (отв. ред.) [и др.]. – Минск : БГУИР. – 2015. – С. 401–406.

27. Мошкин В.С., Ярушкина Н.Г. Особенности интеграции механизмов логического вывода в онтологическую модель представления знаний с помощью SWRL-правил // Четырнадцатая национальная конференция по искусственному интеллекту с международным участием КИИ – 2014 : труды конференции. Т.1. – Казань: Изд-во РИЦ «Школа». – 2014. – С. 173–181.

28. Мошкин В.С., Ярушкина Н.Г. Взаимодействие онтологического анализа и механизмов логического вывода // Гибридные и синергетические интеллектуальные системы: материалы II Международного Пospelовского симпозиума / под ред. д-ра техн. наук, проф. А.В. Колесникова. – Калининград : Изд-во БФУ им. И.Канта, 2014. – С.239–245.

29. Мошкин В.С., Андреев И.А., Башаев В.А., Клейн В.В. Семантическая метрика «термин/не термин» на основе онтологии проблемной области // Методы и технологии гибридного и синергетического искусственного интеллекта: материалы I Международной Пospelовской летней школы-семинара для студентов, магистрантов и аспирантов. Калининград : Изд-во БФУ им. И. Канта, 2014. – С. 67–73.

30. Ярушкина Н.Г., Мошкин В.С. Онтология как основа семантического анализа данных в задачах автоматизированного проектирования // Нечеткие системы и мягкие вычисления (НСМВ-2014): труды Шестой всероссийской научно – практической конференции. – Т. 1. – СПб. : Политехника-сервис, 2014. – С. 174–182.

31. Башаев В.А., Мошкин В.С., Андреев И.А., Клейн В.В. Использование семантической метрики для решения задачи извлечения терминологии из текста проблемной области // Информатика и вычислительная техника : сборник научных трудов 6-й Всероссийской научно-технической конференции аспирантов, студентов и молодых ученых ИВТ-2014 / под ред. Н. Н. Войта. – Ульяновск : УлГТУ, 2014. – С.72–78.

32. Мошкин В.С. Интеграция знаний продукционного характера в онтологическую модель предметной области//Вузовская наука в современных условиях: сборник материалов 48-й научно-технической конференции. Ч.2. – Ульяновск : УлГТУ, 2014. – С. 164–167.

33. Ярушкина Н.Г., Мошкин В.С. Онтологический подход к анализу временных рядов // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сб. научных трудов VII Международной научно-практической конференции. – Коломна – 2013. – № 2. – С. 529–537

34. Мошкин В.С. Использование резонеров в процессе резюмирования временных рядов на основе онтологии проблемной области // Информатика и вычислительная техника: сборник научных трудов 5-й Всероссийской научно-технической конференции аспирантов, студентов и молодых ученых ИВТ-2013 / под ред. Н. Н. Войта. – Ульяновск : УлГТУ, 2013. – С. 126–129.

35. Мошкин В.С. Онтологический анализ временных рядов // Вузовская наука в современных условиях: сборник материалов 47-й научно-технической конференции. Ч. 2. – Ульяновск : УлГТУ, 2013. – С. 272–275.

Свидетельства на регистрацию программы для ЭВМ:

1. Свидетельство № 2017611364 Российская Федерация. Интеллектуальная экспертная система на основе нечеткой онтологии: свидетельство о государственной регистрации программы для ЭВМ / Ярушкина Н.Г., Мошкин В.С.; заявитель и правообладатель Ульян. гос. техн. ун-т. – № 2016660651, заявл. 11.10.2017; зарегистр. 02.02.2017.

2. Свидетельство № 2014616701 Российская Федерация. Онтологически-ориентированная система логического вывода: свидетельство о государственной регистрации программы для ЭВМ / Ярушкина Н.Г., Мошкин В.С.; заявитель и правообладатель Ульян. гос. техн. ун-т. – № 2014614071, заявл. 05.05.2014; зарегистр. 02.07.2014.

3. Свидетельство № 2014616248 Российская Федерация. Онтологически-ориентированная система извлечения терминологии: свидетельство о государственной регистрации программы для ЭВМ / Ярушкина Н.Г., Башаев В.А., Мошкин В.С.; Клейн В.В., Андреев И.А., заявитель и правообладатель Ульян. гос. техн. ун-т. – № 2014613574, заявл. 21.04.2014; зарегистр. 18.06.2014.

Мошкин Вадим Сергеевич

Исследование представления терминологии в лингвистическом
обеспечении САПР на основе интеграции нечетких онтологий и логического
вывода

Автореферат

Подписано в печать. _____ . Формат 60x80/16

Усл. печ. л. 1,17

Тираж 100 экз. Заказ

ИПК «Венец» УлГТУ, 432027, г. Ульяновск, Северный Венец, 32.