

**На правах рукописи**

**ЯКОВЛЕВ Николай Николаевич**

**СИСТЕМНАЯ МОДЕЛЬ КОМПЛЕКСА ТРЕБОВАНИЙ  
К АВТОМАТИЗИРОВАННОЙ ИНФОРМАЦИОННОЙ  
СИСТЕМЕ НА ОСНОВЕ СЕМАНТИЧЕСКОЙ АННОТАЦИИ**

**Специальность 05.13.01 – Системный анализ, управление  
и обработка информации (в промышленности)**

**А В Т О Р Е Ф Е Р А Т**  
**диссертации на соискание ученой степени**  
**кандидата технических наук**

**Уфа–2010**

Работа выполнена на кафедре автоматизированных систем управления  
ГОУ ВПО  
«Уфимский государственный авиационный технический университет»

Научный руководитель                   засл. деят. науки РФ, д-р техн. наук, проф.  
**КУЛИКОВ Геннадий Григорьевич**

Официальные оппоненты               д-р техн. наук, проф.  
**ПАВЛОВ Сергей Владимирович,**  
Уфимский государственный авиационный  
технический университет

канд. техн. наук, доцент  
**ХРИСТОЛЮБОВ Вячеслав Леонидович,**  
ОАО «Уфимское моторостроительное  
производственное объединение»

Ведущая организация                   Республиканский научно-технический  
и информационный комплекс  
**«Баштехинформ»**

Защита состоится \_\_\_\_ \_\_\_\_\_ 2011 года в 10 часов  
на заседании диссертационного совета Д-212.288.03  
при Уфимском государственном авиационном техническом университете  
по адресу: 450000, Уфа-центр, ул. К. Маркса, 12

С диссертацией можно ознакомиться в библиотеке университета

Автореферат разослан 29 декабря 2010 г.

Ученый секретарь  
диссертационного совета  
д-р техн. наук, проф.

**В. В. Миронов**

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Актуальность

Большинство бизнес-процессов любого современного промышленного предприятия подлежат автоматизации и информатизации. Средством автоматизации и информатизации является автоматизированная информационная система (далее АИС), которая формируется, а впоследствии эволюционирует, в соответствии с требованиями, предъявляемыми к ней бизнесом, пользователями, а также законодательной и нормативной средой, окружающей бизнес.

В научно-исследовательской работе СНАОС, выполненной исследовательской организацией Standish Group, указано 10 факторов, которые не позволяют вовремя завершать проект, направленный на создание либо развитие АИС (далее проект АИС), в рамках бюджета и требуемой функциональности. Следует выделить первые три самых распространенных из них:

- недостаток данных от пользователей;
- незаконченные требования и спецификации требований;
- изменение требований и спецификаций требований.

Очевидно, что все три указанных фактора неудач проектов АИС связаны с работой с комплексом требований (КТ), поэтому эффективное управление комплексом требованиями может значительно снизить долю неудавшихся проектов.

Вместе с тем, КТ является одним из ключевых компонентов проекта в соответствии с международным Сводом знаний по управлению проектами PMBoK4. В соответствии с общепринятой методологией IBM Rational Unified Process, регламентирующей разработку АИС, управление требованиями (УТ) к АИС является одним из основных процессов, обеспечивающих качество разработки АИС. Цель УТ состоит в том, чтобы гарантировать документирование, проверку и удовлетворенность потребностей заказчиков от предприятия. Поскольку требование имеет какие-либо парные ему «ответы» (результаты валидации, реализации, тестирования), то актуально повторное использование требования (ПИТ) из одного крупного типового проекта в одном или нескольких других проектах, и, как следствие, использование результатов реализации этого требования. Актуальность ПИТ обусловлена еще и ростом распространенности сервисно-ориентированной архитектуры АИС, одним из принципов которой является повторное использование сервисов как результатов реализации требований.

В соответствии с концепцией Бэзилла и Бриггса, нашедшей свое отражение в международном стандарте системной инженерии ISO/IEC 15288, любая техническая система имеет следующие стадии жизненного цикла (ЖЦ): замысел, разработка, производство, применение, поддержка применения, списание. Принципиальное отличие ЖЦ информационной системы от ЖЦ материальной в том, что на стадии «применение», продукт производства АИС может применяться неограниченное количество раз в любой географической точке, доступной по сети передачи данных. Продукт же производства материальной системы на стадии применения ограничен как количественно, так и в возможности перемещения. Таким образом, информационный продукт может тиражироваться с гораздо меньшими затратами, чем материальный. Масштабное тиражирование при производстве крупных типовых АИС порождает проблему поиска и выбора нужного информационного ресурса среди многообразия доступных.

ПИТ и контроль целостности данных о требовании сталкиваются с такими проблемами как необходимость быстрого поиска требований по семантическим признакам, эргономичного представления данных о требовании и контроля потока работ с требованиями. В данной работе предлагается решение этих проблем на основе системного и семантического моделирования КТ.

Анализ отечественной и зарубежной литературы и источников сети Интернет позволяют сделать вывод, что исследованием и разработкой подходов к решению вышеуказанных проблем занимаются крупные информационные компании, такие как IBM и Borland. Каждая из них предлагает программные средства и технологии, такие как RequisitePro, DOORS и CaliberRM, но ни одна из них не поддерживает работу с семантическими признаками требования, а предлагаемые ими модели потока работ с требованием не удовлетворяют производственным потребностям и целям УТ в полной мере.

Среди российских и зарубежных ученых и IT-специалистов, изучающих проблемы УТ, можно отметить работы А. Левенчука, Б. Мишнева, П. Зильчинского, А. Новичкова, Д. Карлсона, Т.Мунди, Ф. Холсворта, Сойонг Пака, Ян Лин, К. Форсберга, Э. Халл и др. В их работах поднимаются проблемы повторного использования и контроля потока работ, однако не предлагается формализованных, подлежащих автоматизации решений.

Выявленные реалии позволяют констатировать, что проблема организации потока работ с требованием с учетом ПИТ и поиска требований по семантическим признакам представляется недостаточно исследованной, а большинство проектов по организации эффективной работы с требованиями заказчиков не завершаются успехом из-за отсутствия системного подхода к решению данной проблемы. Это определяет актуальность, цель и практическую значимость выбранной темы.

**Объектом исследования** является комплекс требований к автоматизированной информационной системе в течение их жизненного цикла.

**Предметом исследования** является системная модель комплекса требований с формализованными семантическими признаками требования к АИС.

**Целью диссертационной работы** является разработка системной модели КТ, ориентированной на повторное использование требований при реализации новой АИС и повышающей эффективность разработки АИС, а также разработка методики поиска и сравнения требования по его семантическим признакам.

**Задачи исследования**, решение которых необходимо для достижения цели диссертационной работы.

1. Модификация традиционной модели потока работ с требованиями, направленная на их повторное использование.

2. Построение системной модели комплекса требований, ориентированной на их повторное использование, и включающей в себя функциональную, динамическую и информационную модели КТ.

3. Разработка методики поиска и сравнения требований к АИС по семантическим признакам, позволяющей находить семантически схожие требования из комплексов требований существующих проектов в интерактивном режиме.

4. Разработка алгоритмического обеспечения ПИТ, отражающего взаимодействие участников проекта АИС с системой УТ.

5. Разработка прототипа программного обеспечения системы УТ, позволяющего реализовать предлагаемые модели и методику, его апробация и анализ эффективности полученных научных результатов на его основе.

#### **Методы исследования**

Для решения поставленных в диссертационной работе задач использованы методы системного анализа, методы общей теории множеств, теория реляционных и многомерных баз данных, методы объектно-ориентированного программирования, в качестве средства моделирования применены методологии объектно-ориентированного проектирования UML и структурного анализа SADT.

#### **Основные научные результаты, выносимые на защиту.**

1. Модель потока работ с требованиями, модифицированная таким образом, чтобы результаты работы с требованиями могли использоваться повторно при возникновении подобных требований в других проектах.

2. Системная модель комплекса требований, ориентированная на повторное использование требования и результатов его реализации. Она включает в себя модифицированный поток работ с требованиями, представляющий собой функциональную и динамическую модели КТ, и информационную модель комплекса требований, сопровождаемую многоаспектным словарем данных (МСД). МСД позволяет обеспечить достоверность и контроль использования данных о требованиях на протяжении их ЖЦ.

3. Методика поиска и сравнения требований по семантическим признакам, включающая правило семантической аннотации требования, позволяющее формализовать семантику требования и формальный критерий сравнения требований на основе такой аннотации. Следствием данной методики явилась многомерная модель данных о требованиях, основополагающими измерениями которой выступают семантические признаки требования к АИС.

4. Алгоритмическое обеспечение ПИТ, отражающее взаимодействие участников проекта АИС с системой УТ при поиске требований в КТ.

5. Прототип системы УТ (СУТ) на основе системной модели комплекса требований и методики поиска и сравнения требований по семантическим признакам. Анализ эффективности полученных результатов, проведенный на основе прототипа.

#### **Научная новизна работы** содержится в следующих результатах.

1. Предложена структура потока работ с требованиями, отличающаяся от традиционной наличием дополнительной подструктуры, обеспечивающей возможность повторного использования результатов проектных работ с требованиями.

2. Построена системная модель комплекса требований, отличающаяся ориентированностью на повторное использование требования и результатов его реализации. Модель позволяет контролировать поток работ с требованием, предусматривающий ПИТ, и вести базу данных результатов работы с этим требованием в течение его ЖЦ, что повышает полноту и достоверность информации о требованиях к АИС.

3. Разработана методика поиска и сравнения требований по его семантическим признакам, отличающаяся возможностью формализации семантики требования за счет правила семантической аннотации требования, основанного на разделении аннотирующих семантических концептов на категории, и критерия

сравнения требований на основе такой аннотации. Методика позволяет дискретно представить семантику требования, количественно оценить степень смыслового сходства требований и за счет этого ввести в УТ семантическую составляющую.

4. Разработан алгоритм поиска и сравнения требований при ПИТ, отличающийся использованием представленной в настоящей работе методики поиска и сравнения требований по его семантическим признакам. Алгоритм позволяет в автоматизированном режиме находить по семантическим признакам требования, похожие на исходное.

**Практическую значимость** имеют следующие результаты.

1. Разработанная системная модель КТ и оригинальная методика поиска и сравнения требований по семантическим признакам позволяют ускорить работу с требованием в течение его ЖЦ. Использование их программной реализации в виде СУТ повышает качество реализуемого комплекса требований, и, как следствие, качество поставляемой предприятию АИС.

2. Разработанный на основе технологии РНР программный прототип СУТ «SemanticReq», реализующий представленные в работе научные результаты, позволяет участникам проекта по созданию АИС в принимать решения для ПИТ на основе имеющегося в проекте комплекса требований.

Результаты данного исследования были использованы в ООО «Дататех» в проектах по внедрению и кастомизации АСУД на платформе «БОСС-Референт 3.2.x» (в ОАО «АНК Башнефть»), что подтверждено актом внедрения, а также были внедрены в учебном процессе технического вуза.

#### **Апробация работы**

Основные научные и практические результаты диссертационной работы докладывались и обсуждались на следующих конференциях: 9-, 11- и 12-й Международных конференциях «Компьютерные науки и информационные технологии» (CSIT), Уфа-Красноусольск, 2007, Крит (Греция), 2009, Уфа-Москва-Санкт-Петербург, 2010; Региональной зимней школе-семинаре аспирантов и молодых ученых, Уфа, 2006 и 2010, на семинарах регионального уровня, а также во время научной стажировки в университете Карлсруэ (Германия) по программе DAAD «Михаил Ломоносов» в 2007–2008 годах.

#### **Публикации**

Основные положения и результаты диссертационной работы опубликованы в 9 источниках, включающих 2 статьи в журналах, реферируемых ВАК, и 1 свидетельство о государственной регистрации программы для ПЭВМ.

#### **Структура и объем работы**

Работа включает введение, 4 главы основного материала, заключение, библиографический список из 104 наименований и 2 приложения. Работа без библиографического списка и приложений изложена на 130 страницах машинописного текста и включает 31 рисунок и 10 таблиц.

#### **Связь с плановыми исследованиями**

Данное исследование проводилось в рамках гранта Президента Российской Федерации № НШ-65497.2010.9.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность диссертационной работы, сформулирована цель и задачи исследования, определены научная новизна и практическая значимость работы.

**Первая глава** посвящена обзору и анализу существующих проблем и особенностей УТ в проектах АИС, касающихся разработки и эволюции АИС. Также рассмотрены общие проблемы управления проектами создания АИС. Даны определения основным понятиям, рассмотрены их философские и онтологические аспекты. Проведена аналогия эволюции АИС с эволюцией естественных систем с применением теории эволюции Дарвина и теории вызова и ответа Тойнби.

В соответствии с Глоссарием терминов программной инженерии IEEE, являющимся общепринятым международным стандартом, требование это:

- 1) условия или возможности, необходимые пользователю для решения проблем или достижения целей;
- 2) условия или возможности, которыми должна обладать система или системные компоненты, чтобы выполнить контракт или удовлетворять стандартам, спецификациям или другим формальным документам;
- 3) документированное представление условий или возможностей.

Также приведены определения требования из стандарта качества ISO 9000:2008 и стандарта разработки требований ISO/IEC 29148.

Комплекс требований – это модель (с определенной степенью адекватности, точности) предполагаемого решения в терминах требований.

УТ начинается с выявления и анализа целей и ограничений заказчика. УТ включает поддержку ЖЦ требований, мониторинг реализации требований и организацию работы с требованиями и сопутствующей информацией, возникающей вместе с требованиями. Рассмотрены практики и положения из стандартов ГОСТ 34, ITIL, RUP и BABoK. На их основе проанализированы место КТ в жизненном цикле АИС, особенности и проблемы УТ и пути их решения, а также приведены классификации требований. Дана авторская классификация требований.

Рассмотрены проблемы УТ к АИС на примерах внедрения и кастомизации автоматизированной системы электронного документооборота (далее АСУД) для промышленных предприятий и технических вузов. Выделены следующие проблемные подпроцессы УТ:

- контроль целостности требований проекта АИС;
- повторное использование требований к АИС;
- контроль потока работ по реализации требования к АИС.

Общая проблематика УТ и ПИТ схематично представлена на Рисунок 1.

Выполнен анализ существующих программных решений в УТ, таких как IBM Rational RequisitePro, Borland CaliberRM и Telelogic DOORS. Они предоставляют возможности хранения и изменения требований с атрибутами, трассировку и выполнение реляционных запросов, но не обладают функциональностью, обеспечивающей ПИТ и работу с семантикой требований. Делается вывод о необходимости разработки системной модели КТ (СМКТ), способствующей ПИТ, и методики поиска и сравнения требований по семантическим признакам. Проведен анализ близких к исследуемой теме отечественных и зарубежных научных работ, рассматривающих аналогичные проблемы и подходы, указаны недостатки данных работ, не позволяющие решить вышеозначенные проблемы.

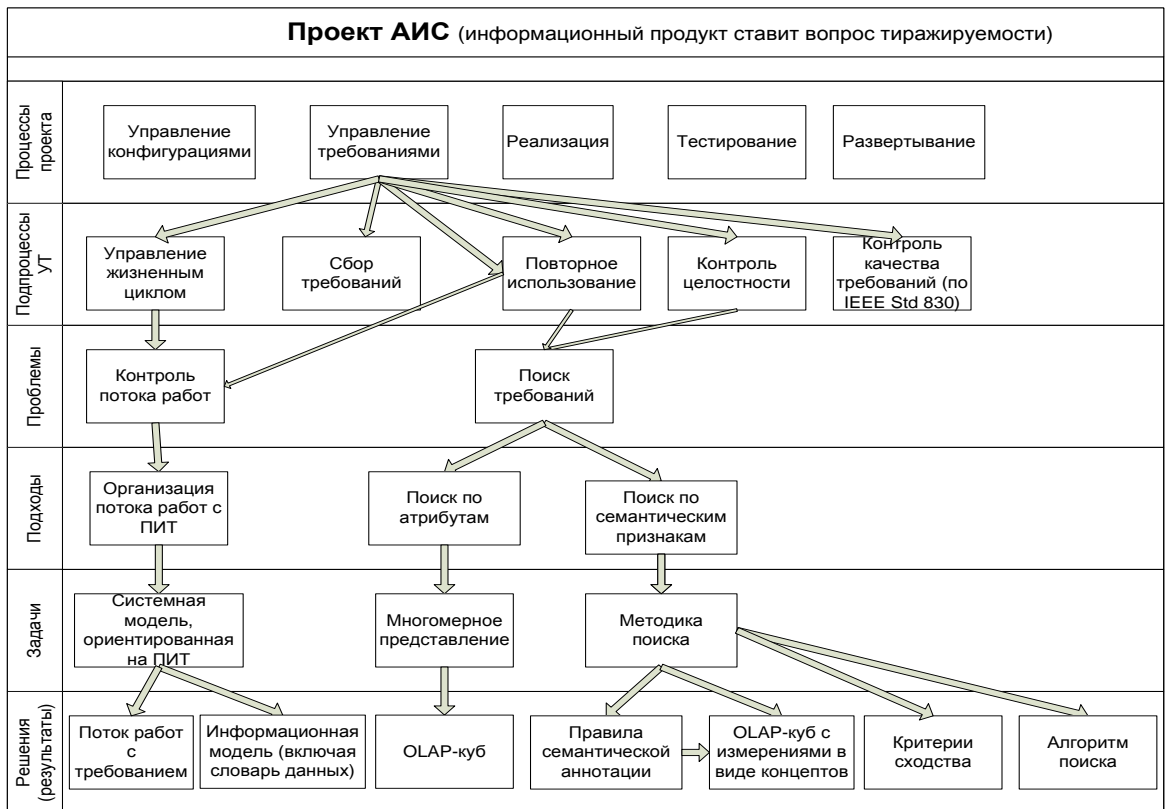


Рисунок 1 – Многоуровневая иерархическая схема проблематики УТ и ПИТ

**Вторая глава** посвящена разработке СМКТ, ориентированной на ПИТ. Системная модель понимается как совокупность функциональной, информационной и динамической модели в терминах SADT.

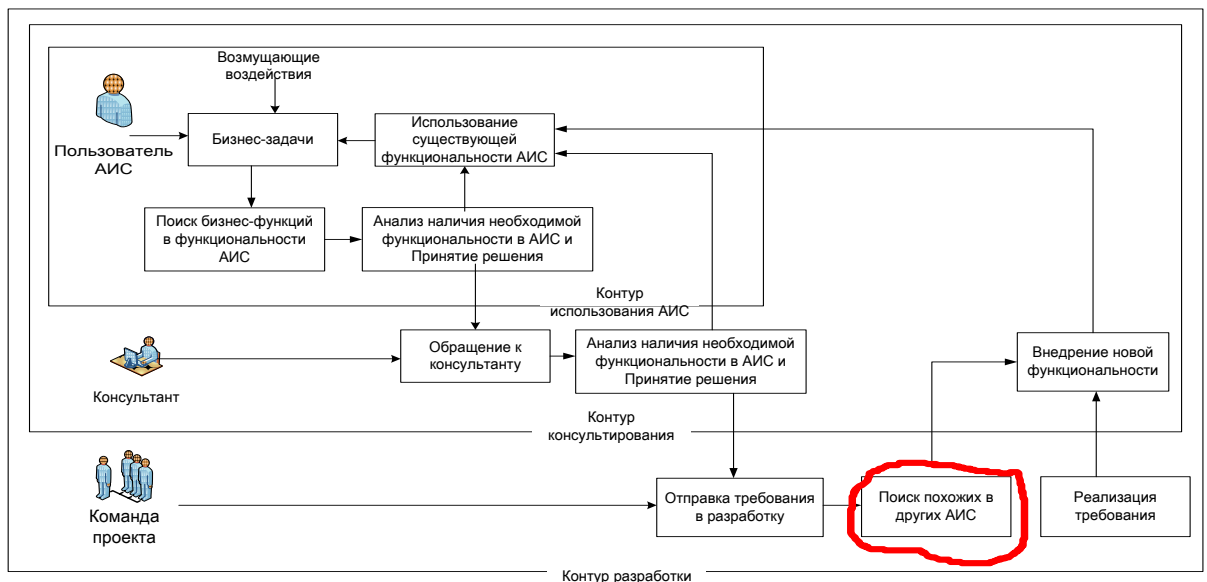


Рисунок 2 – Контур управления требованиями с учетом ПИТ

Представлен тройной контур УТ к АИС по отклонениям (Рисунок 2), в котором объектом управления являются требования и результаты их реализации, субъектом – участники проекта АИС, разделенные по ролям, а управляющим воздействием – решения, принимаемые субъектами с помощью СУТ.

Иллюстрируется ЖЦ требования к АИС, предложенный в данной работе на основе практик ITIL 3 и ЖЦ АИС по стандартам RUP и ISO/IEC 12207, разработанных на основе теорий ЖЦ АИС У. Ройса и Б. Бозма. Расписаны способ



представления требования и участник проекта для каждой стадии. Приводится модель потока работ с требованиями, ориентированного на ПИТ, в виде диаграммы активностей по методологии UML, являющейся модификацией сети Петри (Рисунок 3). Овалами представлены активности участников проекта, а прямоугольниками – состояния требования. Эта модель отражает функциональные и динамические свойства требования и КТ.

Ставится задача разработки ИМКТ, которая дает следующие возможности:

- поддерживать КТ с учетом ПИТ и результатов их реализации, т. е. служить информационным обеспечением потока работ с требованием, ориентированного на ПИТ;
- учитывать такие аспекты требования как «стадия ЖЦ», «исполнитель на каждой стадии ЖЦ», «область знаний»;
- получать сводные отчеты о показателях КТ, с помощью которых пользователь-аналитик сможет ориентироваться в информационном пространстве требований по их атрибутивным и семантическим признакам гораздо эффективнее, чем с помощью традиционных реляционных запросов.

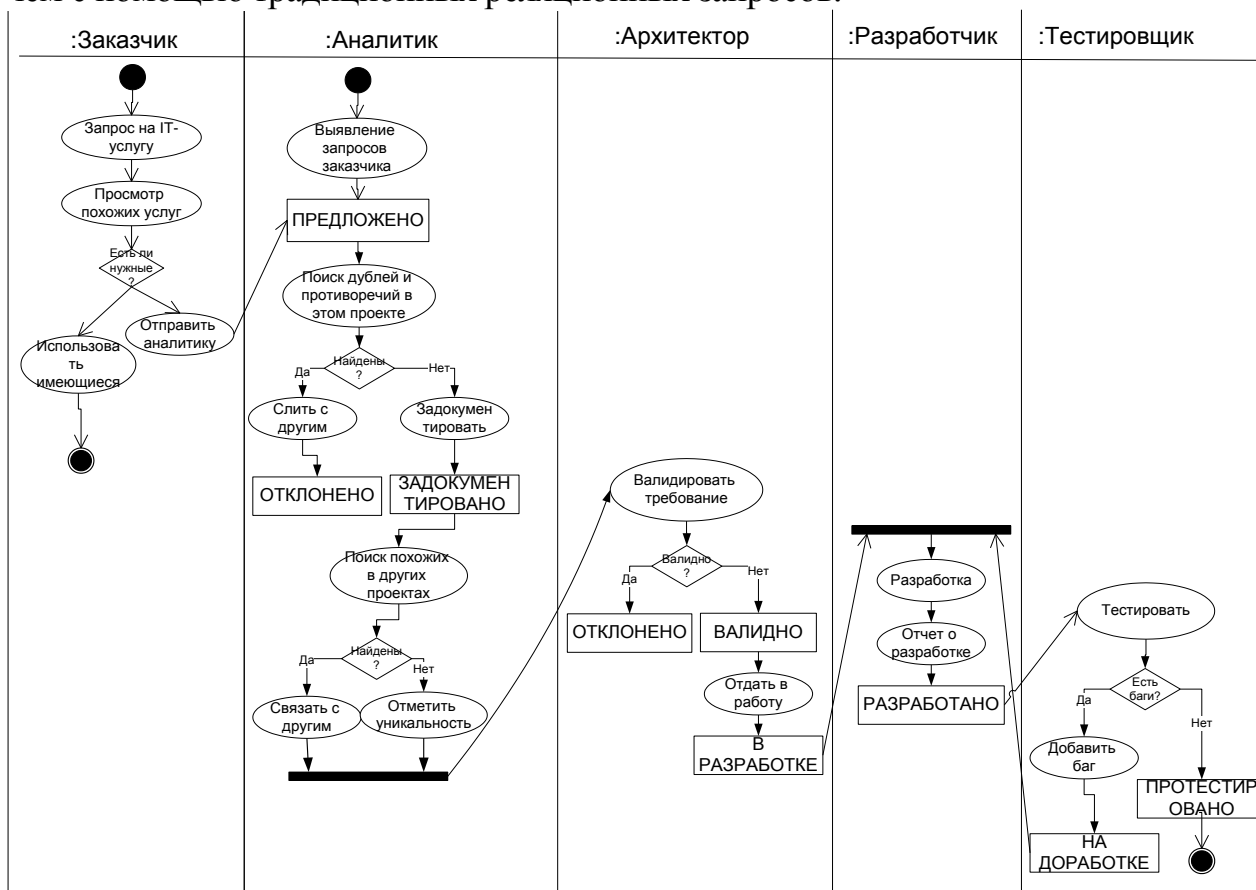


Рисунок 3 – Модель потока работ с требованиями, ориентированного на ПИТ

На основе международных стандартов ITIL и BABoK 1.6 анализируются атрибуты требования, необходимые для решения поставленной задачи. Усовершенствование информационного обеспечения потока работ с требованиями позиционируется как подход к увеличению эффективности УТ.

ИМКТ приводится в виде диаграммы «сущность-связь», построенной в нотации IDEF1X (Рисунок 4). На ее основе разрабатывается многоаспектный словарь данных (МСД) о требовании. Он определяет метаданные требования, которые не позволяет отразить нотация IDEF1X. Эти метаданные определяют

порядок работы с данными о требовании и доступа к ним. Для построения МСД выделено три аспекта, один из которых – область знаний – определяет основание для визуальной группировки атрибутов, а два других – роль пользователя и стадия ЖЦ определяют, может ли текущий пользователь редактировать или вводить тот или иной атрибут на данной стадии ЖЦ. Такой подход служит основанием для многоаспектного разграничения доступа к данным о требованиях и, следовательно, повышает уровень целостности и достоверности данных.

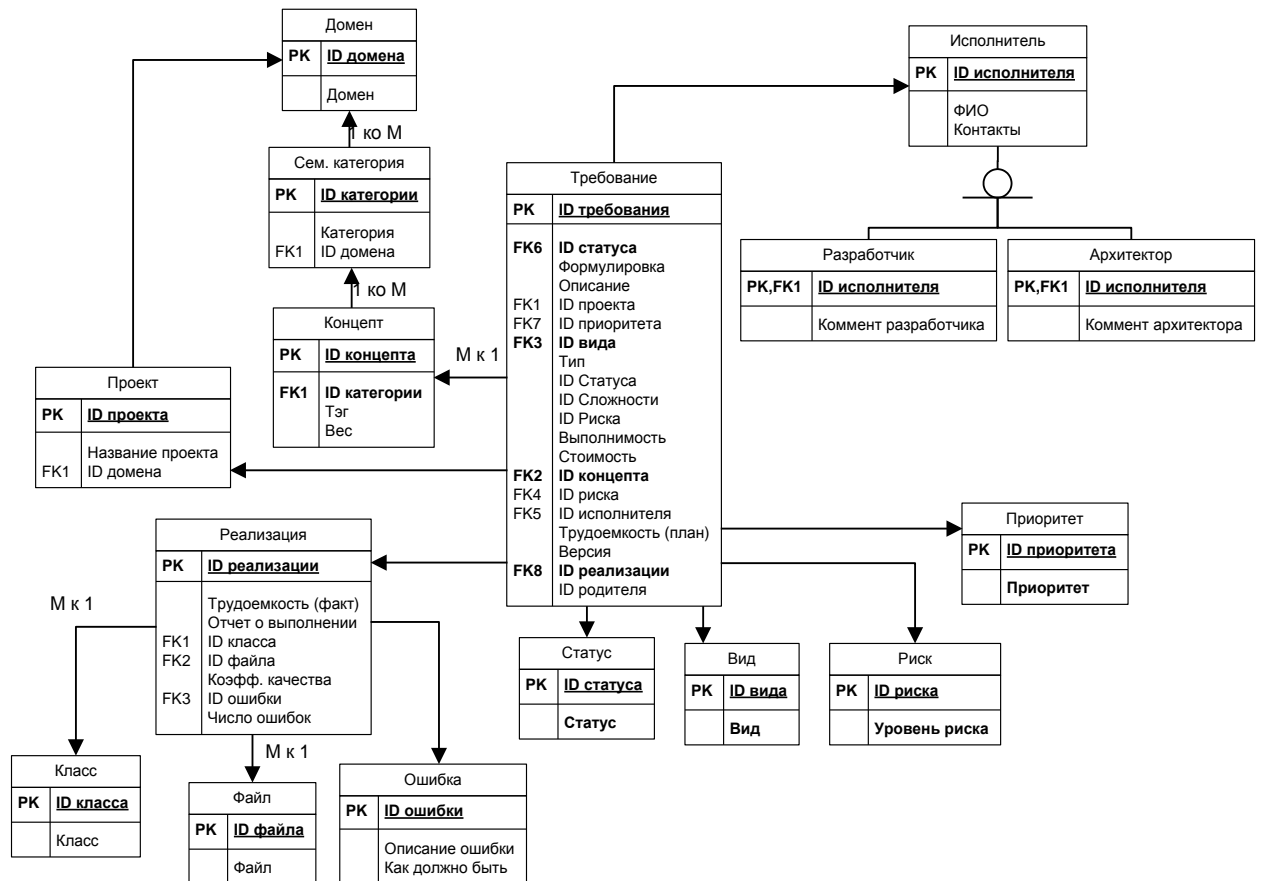


Рисунок 4 – Информационная модель КТ

В основе МСД лежит следующая метамодель. Обозначим множество всех атрибутов требования, показанных на ИМКТ, как  $Attr(R)$ , где  $R$  – это требование.

$$\text{Тогда } A(R, f_j, u_k, s_l) = \{a_1; a_2; \dots; a_n\}, n \in N, (1)$$

– это множество атрибутов каждого требования, которые при визуализации группируются в область знаний  $f_j$ , и их может редактировать пользователь с ролью  $u_k$ , когда требование находится в стадии  $s_l$ .

$f_j \in F$  – область знаний (*field*, для данной работы перечень областей взят из свода знаний по проектному менеджменту исследовательского института РМИ);

$u_k \in U$  – роль пользователя (*User*);

$s_l \in S$  – стадия ЖЦ;

и при  $i = \overline{1, n} \quad \forall a_i \in Attr(R)$ ,

$$\text{а при } \forall f_j \in F \cap \forall u_k \in U \cap \forall s_l \in S \quad A(R, f_j, u_k, s_l) \in Attr(R) (2)$$

Фрагмент МСД представлен в виде матрицы «Стадия ЖЦ \* Область знаний».

Таблица 1 – Фрагмент МСД о требовании (описана часть атрибутов)

Атрибут	Аспект			Обяз	Тип	Мн/ зн
	Область	Стадия	Роль			
Описание	Объем	Выявление	Аналитик		Текст	
Заголовок	Объем	Докум-ние	Аналитик	Да	Текст	
Тип	Объем	Выявление	Аналитик	да	Категория	
Приоритет	Время	Выявление	Аналитик	да	Категория	
Автор	Коммуни- кации	Выявление	Аналитик	да	Категория	
Родительское требование	Объем	Докум-ние	Аналитик		Требова- ние	
Трудоемкость (план)	Стоимость	Валидация	Архитектор		Число	
Аналитик	Ресурсы	Докум-ние	Система		Категория	
Разработчик	Ресурсы	Валидация	Система		Категория	
Уровень риска	Риски	Валидация	Архитектор		Категория	
Классы	Интеграция	Разработка	Разработчик	да	Категория	Да

Представленная СМКТ, ориентированная на ПИТ, включает модель потока работ с требованием, отражающую функциональные и динамические свойства требования, и информационную модель комплекса требований (ИМКТ), предполагающую наличие единого кросс-проектного информационного пространства требований. На основе ИМКТ предлагается многоаспектный словарь данных о требовании (МСД), учитывающий следующие аспекты: 1) состояние требования в соответствии с номенклатурой состояний в потоке работ, 2) роль участника проекта, работающего с требованием, 3) область знаний в соответствии с номенклатурой областей знаний по стандарту PMI.

Применение СМКТ способствует решению описанных в 1-й главе проблем УТ. СМКТ служит основой разработки СУТ с поддержкой ПИТ.

**В третьей главе** предлагается методика поиска и сравнения требований на естественном языке по семантическим признакам, позволяющая сопоставлять семантику полученных требований друг с другом, а также с уже имеющимися требованиями аналогичных проектов АИС, с целью проверки их на дублирование, противоречивость и сходство.

Проблема заключается в том, что в разных проектах, связанных с развитием одной базовой АИС, от разных заказчиков могут поступать похожие, а иногда даже одинаковые требования, а аналитики и разработчики могут даже не подозревать о том, что такое же или аналогичное требование уже поступало к их коллегам, и реализовывать уже готовую функциональность заново. Также могут возникнуть несколько внешне различных требований, имеющих одинаковую семантику с функциональной точки зрения. Поэтому важным элементом управления проектом АИС становится правильно организованная СУТ, базирующаяся на предложенной СМКТ и методике.

Одним из подходов к решению вышеозначенной проблемы УТ, облегчающим работу с комплексом требований к АИС и взаимодействие между участниками проектов и повышающим эффективность работ за счет снижения трудоемкости,

является разработка методики поиска и сравнения требований на естественном языке (ЕЯ) по семантическим признакам.

Автоматизированное сравнение формулировок требований на естественном языке (ЕЯ) представляется очень сложным в контексте ПИТ и нецелесообразным. Поэтому предлагается формализовать семантику требования к АИС.

Для разработки методики поиска и сравнения требований по семантическим признакам ставятся следующие задачи:

- 1) формализация семантики требований (выделение их формальных семантических признаков) для семантической аннотации (СА);
- 2) разработка формальных критериев сравнения требований по их семантическому (смысловому) содержанию;
- 3) разработка алгоритма аннотации, поиска и сравнения требований.

Приводится теоретико-множественное описание КТ. Вводится понятие «домен» – группа проектов, объединенных по какому-либо признаку, например, предметной области.

$D = \{d_1, d_2, \dots, d_n\}$  – множество доменов (предметных областей),  $n \in N$  ;

$P_i = \{p_1^i, p_2^i, \dots, p_m^i\}$  – множество проектов  $i$ -го домена, где  $i = \overline{1, n}$  ;

$R_{ji} = \{r_1^{ji}, r_2^{ji}, \dots, r_l^{ji}\}$  – множество требований  $j$ -го проекта  $i$ -того домена, где  $i = \overline{1, n}$  и  $j = \overline{1, m}$ .

Требование – это утверждение, имеющее свою структурную специфику, а его структура в соответствии с известными практиками разработки АИС (ИПЛ, ВАВоК) сводится к шаблону «[обстоятельства][субъект][действие][объект]». Пример: «Подписант [субъект] должен иметь возможность открыть на редактирование [действие] поручение [объект] на стадии исполнения [обстоятельства]». Таким образом, требование описывается несколькими категориями, каждая из которых несет в себе часть его семантики – концепт (ключевое слово), причем каждый концепт принадлежит определенной категории.

$Concepts_i$  – множество концептов  $i$ -го домена, где  $i = \overline{1, n}$ , с точки зрения векторного анализа является базисом СА.

В качестве решения задачи формализации семантики требования разрабатывается **правило категоризованной СА**. СА состоит из концептов, каждый из которых принадлежит к одной из категорий. Каждая категория имеет только один концепт. Концепт и является семантическим признаком требования.

Аналитик заполняет predetermined для каждого проекта семантические атрибуты требования (категории). Помимо вышеприведенных основных четырех категорий для каждого домена могут добавляться специфичные для него дополнительные категории. Домен обладает набором семантических категорий, актуальных для этих всех его проектов, и репозиторием всех концептов, уже имеющихся во всех проектах домена.

Представлена теоретико-множественная модель комплекса требований.

$r_k^{ji} = \{text^{kji}, C^{kji}, A^{kji}\}$  – вектор атрибутов  $k$ -го требования (где  $k = \overline{1, l}$ )  $j$ -го проекта  $i$ -го домена, где  $text^{kji}$  – это текстовая формулировка требования,

$C^{kji} = \{c_1^{kji}, c_2^{kji}, \dots, c_q^{kji}\}$  – множество концептов (т. е. СА), а каждый концепт  $c_x^{kji}$  принадлежит суммарному множеству, т. е.  $c_x^{kji} \in Concepts_i$ , а  $A^{kji} = \{a_1^{kji}, a_2^{kji}, \dots, a_s^{kji}\}$  – множество несемантических атрибутов требования, управление которыми рассмотрено во второй главе.

Правило категоризованной СА позволяет аналитику получить для всех требований  $r_k^{ji}$  множество концептов  $C^{kji}$  на основании его заголовка  $text^{kji}$ .

Аннотация может производиться как вручную (вводом концепта с клавиатуры или выбором из списка уже имеющихся), так и автоматически (система обрабатывает текст требования и, если, находит в тексте требования слово, уже зафиксированное ранее как концепт этого домена, то значение подставляется в категорию автоматически).

В рамках определенного домена набор категорий можно расширить, например, для АСУД можно добавить такие категории как «документ», «адресат». Так как концепты разбиваются по категориям, то в рамках одного домена число категорий, и, следовательно, число аннотирующих концептов всегда будет одинаково. Это дает возможность представлять аннотацию требования в виде «строк» с одинаковым набором букв, причем для каждой позиции (категории) предполагается уникальный алфавит из концептов.

На основании этой возможности в качестве **критерия сравнения** и меры смыслового сходства требований предлагается расстояние Хэмминга (далее РХ).

$|C^{kji} \cap C^{k'j'i}|$  – число элементов пересечения двух множеств концептов, определяет сходство требований  $r_k^{ji}$  и  $r_{k'}^{j'i}$ : чем больше пересечений, тем сильнее сходство. Следует обратить внимание, что литера  $i$  не имеет штриха – домен для обоих требований один и тот же.

Критерий сравнения требований позволяет при заданном для исходного требования  $r_{k'}^{j'i}$  множестве  $C^{k'j'i}$  найти все требования  $r_k^{ji}$  для  $k = \overline{1, l}$  при  $j = \overline{1, m}$ , для которых  $|C^{kji} \cap C^{k'j'i}| \rightarrow \max \left( |C^{kji}|, |C^{k'j'i}| \right)$  с целью получить доступ к атрибутам  $A^{kji}$ , включающим, в том числе, и описание реализации.

Результат использования данного критерия является основой для принятия участником проекта решения о семантической схожести требований.

Предлагаемый **алгоритм поиска** семантически похожих требований на основе категоризованной СА и РХ включает следующие операции:

- 1) Начало;
- 2) в  $j$ -том проекте  $i$ -го домена вводится требование  $r_{k'}^{j'i}$ ;
- 3) для требования  $r_{k'}^{j'i}$  задается множество концептов  $C^{k'j'i} = \{c_1^{k'j'i}, c_2^{k'j'i}, \dots, c_{q_i}^{k'j'i}\}$ , по одному концепту на каждую из  $q_i$  категорий, где  $q_i$  – число категорий  $i$ -го домена;

4) для каждого  $k = \overline{1, l}$  при  $j = \overline{1, m}$  вычисляется РХ  $d(C^{k'j^i}, C^{kji}) \in [0; q_i]$  и если  $d(C^{k'j^i}, C^{kji}) \leq d_0$ , где  $d_0$  – условный уровень значимости, заданный аналитиком, то утверждается, что  $r_k^{ji}$  семантически похоже на  $r_{k'}^{j^i}$ ;

5) если  $j = j' \cup k \neq k'$ , то требование  $r_{k'}^{j^i}$  позиционируется системой как дубликат либо противоречие требованию  $r_k^{ji}$ ; аналитик принимает решение;

6) если  $j \neq j'$ , то требование  $r_{k'}^{j^i}$  позиционируется системой как аналогичное требование из другого проекта, т.е его атрибуты  $A^{kji}$  могут быть использованы при реализации  $r_{k'}^{j^i}$ ; аналитик принимает решение;

7) Конец.

В результате Пользователь-аналитик получает выборку требований, упорядоченную по степени убывания похожести на исходное (возрастания РХ).

Благодаря правилу категоризованной СА появляется возможность строить многомерное пространство концептов, в котором каждое требование является точкой с заданными координатами. Это служит основанием для построения OLAP-куба как средства повышения качества поиска данных о требованиях.

Категоризованное множество концептов, принадлежащее определенному домену, представляет собой дискретное семантическое пространство требований, измерениями которого являются семантические категории. Это пространство конечно и имеет постоянную размерность, равную числу семантических категорий требования в данном домене.

Правило категоризованной СА вводит также множество категорий домена  $Cat^i = \{cat_1^i, cat_2^i, \dots, cat_{q_i}^i\}$ , где  $q_i$  – число категорий в домене, тогда  $\forall t = \overline{1, q_i} \Leftrightarrow c_t^{kji} \in V(cat_t^i)$ , т. е. концепт t-й категории i-го домена j-го проекта k-го всегда принадлежит к какой-либо категории.

При чем  $V(cat_t^i)$  – это множество концептов t-й категории i-го домена, т. е. словарь категории.

Имеем множество категорий i-го домена  $Cat_i = \{cat_1^i, cat_2^i, \dots, cat_{q_i}^i\}$ , где  $q_i$  – число категорий в домене, тогда

$$\forall t = \overline{1, q} \Leftrightarrow c_t^{kji} \in Cat_i^i \quad (3).$$

Так как таксономия концептов иерархична, то любая категория в домене имеет множество уровней таксономии. Обозначим множество уровней j-й категории i-го домена как  $LEV_j^i$ . Следовательно, для каждого i-го домена имеем множество пар  $\{(cat_1^i, LEV_1^i), \dots, (cat_q^i, LEV_q^i)\}$ .

Такое множество имеет вид, совпадающий с видом традиционной математической схемы OLAP-куба.  $t = \overline{1, q_i} \mid V(cat_t^i) = (v_1^{it}, v_2^{it}, \dots, v_{zq}^{it})$ , где  $zq$  – число элементов в q-й категории.

Любой концепт любой категории принадлежит множеству концептов домена  $\forall v_k^{it} \in Concepts_i$ , а все концепты всех категорий домена представляют собой множество концептов домена  $\bigcup_{j=1}^q V(cat_j^i) = Concepts_i$ . Любой концепт любого требования любого проекта принадлежит одной из категорий домена:

$$(l = \overline{1, q}) \cap \forall k \cap \forall j | c_l^{kij} \in V(cat_j^i) \quad (4)$$

Таким образом, любой концепт  $c_l^{kij}$  является координатой в семантическом пространстве требований домена  $D_i$ , а полный набор концептов, среди которых каждый и только он принадлежит к своей категории, определяет подмножество

требований  $i$ -го домена  $R(D_i, \bigcup_{j=1}^{q_i} c_j^i)$ , где  $\forall j = \overline{1, q} | c_j^i \in V(cat_j^i)$ .

Цель участника проекта – найти количественные атрибуты из множества  $A^{kji}$ , если  $r_k^{ij} \in R(D_i, \bigcup_{j=1}^{q_i} c_j^i)$ , которые являются мерами OLAP-куба.

На основе вышеприведенной теории, а дополненной ИМКТ, представлена многомерная модель данных требования, включающей концепты требования в качестве измерений OLAP-куба.

С точки зрения многомерного представления атрибутивные данные подразделяются на *категориальные данные*, качественно характеризующие требование, не имеющие количественного выражения, и *показатели* – количественные данные, подлежащие агрегации на основе категориальных данных. Это позволяет построить OLAP-куб и с его помощью в интерактивном режиме генерировать OLAP-отчеты. Это дает возможность участникам проекта искать требования визуально по категориальным атрибутам требования, например, сложность, уровень риска, а также анализировать показатели УТ (трудоемкость, число требований).

Таким образом, представлена методика поиска и сравнения требований к АИС на естественном языке по семантическим признакам, учитывающая структурную специфику требования и предусматривающую единообразную категоризацию аннотирующих концептов для всех требований в рамках проекта. Данная методика включает 1) правило категоризованной семантической аннотации требования, основанное на разделении аннотирующих семантических концептов на категории; 2) критерий сравнения смыслового сходства требований на основе этой аннотации. Представлен алгоритм аннотации, поиска и сравнения требований.

Набор категорий одинаков для всех требований одного проекта либо группы проектов, определенной пользователем. Кроме того, представлено расширение ИМКТ и многомерного OLAP-куба, основополагающими измерениями которого выступают семантические концепты требования к АИС.

**Четвертая глава** посвящена разработке и апробации прототипа программного обеспечения SemanticReq, реализующего описанные во второй и третьей главах модели и методику, а также анализу их эффективности на основе апробации разработанного прототипа. Прототип разработан на основе технологии

PHP и каркасной системы управления сайтом CMF Drupal с использованием концепции AJAX и СУБД MySQL 5, и является динамическим Web-приложением. Предлагается классовая архитектура СУТ, включающей сервис семантической аннотации и поиска требований к АИС по семантическим признакам.

Благодаря реализации предложенной в работе СМКТ, участники проекта самостоятельно, не вступая в контакт с коллегами, с помощью СУТ получали ответы на следующие вопросы, связанные с ПИТ (Таблица 2).

Таблица 2 – Вопросы участников проекта, на которые отвечает СУТ

Участник	Вопросы
Разработчик	– Кто оформил требование? – Сколько времени отведено на разработку?
Тестировщик	– Кто реализовывал требование? (программно) – По какому сценарию тестировать требование?
Аналитик	– Какие требования уже зафиксированы в других проектах? – Какие из них реализованы? – Какие требования в данном проекте дублируются?
Менеджер проекта	– Сколько времени затрачено на разработку или тестирование?

Полученные ответы способствовали более быстрому и качественному выполнению работ по реализации требований к АИС.

Эффективность предлагаемых моделей и методики определялась на основе анализа комплекса требований (около 120) проекта АСУД ОАО «Башнефть», реализованного в ООО «Дататех». Требования, собранные аналитиками проектов, сравнивались между собой. Результаты анализа приведены в Таблица 3.

Таблица 3 – Сравнение результатов анализа требований в IT-проектах

Параметр	Эксперт	SemanticReq	Сравнение
Число дубликатов требований в проекте	6	8	0,75
Число противоречивых требований в проекте	1	2	0,5
Число требований других проектов, пригодных для ПИТ	2	8	0,25
<b>Итого</b> отмечено похожих требований	9	24	0,375
Затраты на реализацию требований, имеющих аналоги в других проектах, чел-час	46	9	5,11
Затраты на нахождение дубликатов и аналогов, чел-час	11	1	11
<b>Итого</b> трудозатрат	47	10	4,7

Всего на реализацию всех требований проектов было затрачено 290 человеко-часов, таким образом, сэкономленные 47 часов составили 16,2% от общего объема работ. Очевидно, что применение СУТ позволяет находить по семантическим признакам больше требований в единицу времени и в несколько раз снижает трудозатраты, связанные с ПИТ. Для эксперта-аналитика пропадает необходимость рутинной работы по сравнению текстов требований и удержания в своей памяти описаний других похожих проектов. При увеличении числа похожих проектов и требований в них эти показатели могут увеличиваться в факториальной зависимости



от числа парных сочетаний требований в соответствии с формулой  $C_2^n = \frac{n!}{(n-2)!*2!}$ , где  $n$  – число требований в проекте.

Игнорирование проблемы устранения противоречивых требований снижает вероятность успеха проекта и увеличивает репутационные риски исполнителя. А игнорирование проблемы дублей и повторного использования – повышает трудоемкость реализации требований: в рассмотренном примере – на 19 человеко-часов. Таким образом, авторская СУТ SemanticReq дает возможность получить как измеримый (в человеко-днях), так и неизмеримый эффект в виде повышения качества КТ и самой как АИС как продукта.

В работе также рассматривается область применения полученных результатов, приводится анализ преимуществ (возможность СА требования, управление потоком работ с разграничением доступа и ПИТ) и недостатков (невозможность интеграции на прикладном уровне с другими СУТ). Описаны возможные направления дальнейших исследований в этой области.

## ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ РАБОТЫ

1. Модифицирована модель потока работ с требованием, таким образом, чтобы результаты работы с требованием могли использоваться повторно при возникновении подобных требований в других проектах.

2. Разработана системная модель КТ, ориентированная на повторное использование требования и результатов его реализации. Она включает в себя модифицированный поток работ с требованием, представляющий функциональную и динамическую модели комплекса требований, и информационную модель комплекса требований, сопровождаемую многоаспектным словарем данных (МСД). Модель позволяет контролировать поток работ с требованием в течение его ЖЦ, предусматривающий ПИТ, и вести учет результатов работы с этим требованием в течение его ЖЦ, а МСД определяет режим доступа к этим результатам.

3. Разработана оригинальная методика поиска и сравнения требований по семантическим признакам, включающая правило семантической аннотации требования, позволяющее формализовать семантику требования, формальный критерий сравнения требований на основе такой аннотации и алгоритм поиска семантически похожих требований. Методика позволяет в автоматизированном режиме находить похожие по смыслу требования в различных проектах АИС. Представлено расширение OLAP-куба, позволяющего анализировать показатели КТ и визуализировать требования к АИС с учетом их семантики.

4. Разработан алгоритм поиска и сравнения требований при ПИТ на основе представленной в настоящей работе методики поиска и сравнения требований по его семантическим признакам. Позволяет находить в других проектах требования, результаты реализации которых могут быть использованы при реализации исходного требования

5. На основе предложенной системной модели комплекса требований и методики поиска и сравнения требований по семантическим признакам реализован прототип семантико-ориентированной СУТ. Анализ результатов работы прототипа показал, что применение научных результатов данного исследования ускоряет принятие решений в УТ, снижает трудозатраты на ПИТ в 4,7 раза, общие

трудозатраты на реализацию требований в проекте на 16,2% и повышает качество УТ к АИС.

## **ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

### **В рецензируемых журналах из перечня ВАК**

1. Семантическое и многоаспектное моделирование в управлении требованиями к математическому и программному обеспечению / Сулейманова А. М., Н. Н. Яковлев // Вестник Башкирского Университета, Уфа, 2010. Т.15. №3. С. 553-556.
2. Использование OLAP-технологии для комплексного анализа основных показателей бизнес-процессов кафедры вуза / Г. Г. Куликов, Г. В. Старцев, В. А. Суворова, Н. Н. Яковлев // Вестник УГАТУ. Уфа, 2007. Т9, №7 (25). С. 60-66.

### **В других изданиях**

3. Методика обследования хозяйственной деятельности организации и существующей в ней учетной OLTP-системы для проектирования и построения OLAP-системы / Г. Г. Куликов, О. М. Куликов, Н. Н. Яковлев // Информационные технологии в науке, образовании и производстве: сборник материалов II Международной научно-технической конференции, Орел: изд-во Орл. гос. ун-та, 2006. С. 102-107.
4. Управление контентом на основе OLAP-технологий для создания процессно-ориентированных образовательных программ // Г. В. Старцев, А. М. Сулейманова, В. А. Суворова, Н. Н. Яковлев // Компьютерные науки и информационные технологии: матер. 9-й Междунар. конф. (CSIT'2007). Уфа: Изд-во «Диалог», 2007. Т. 2. С.143-146 (Статья на англ. яз.)
5. Выбор компьютерных архитектур посредством интеллектуального анализа данных на основе метрик программного кода / Н. Н. Яковлев // Сборник материалов научного семинара стипендиатов программы «Михаил Ломоносов» 2007/08 года. Москва: Изд-во «DAAD», 2008. - С. 73-75 (статья на англ. яз.)
6. Поддержка повторного использования требований при кастомизации ПО АИС на основе поиска по тэгам / Н. Н. Яковлев, О. Р. Хабибуллина // матер. 11-й Междунар. конф. (CSIT'2009). Уфа: Изд-во «Диалог», 2009 Т.3, С. 93-97. (Статья на англ. яз.)
7. Свид. о рег. программ для ЭВМ № 2010613872. Система управления требованиями IT-проекта SemanticReq // Н. Н. Яковлев, Р. Р. Кашаев, Л. Ш. Иванова. РосПатент. 15.06.2010.
8. Семантико-ориентированный подход к моделированию требований при развитии ПО АИС / Н. Н. Яковлев, Н. Д. Торопова // матер. 12-й Междунар. конф. (CSIT'2010). Уфа: Изд-во «Диалог», 2010. Т.3, С. 103-107. (Статья на англ. яз.)
9. Системная модель комплекса требований к автоматизированной информационной системе, ориентированная на повторное использование требований / Н. Н. Яковлев // Альманах современной науки и образования. Тамбов: Грамота, 2010. № 11 (42): в 2-х ч. Ч. 2. С.115-117.
10. OLAP-куб и многомерная модель данных о требованиях к автоматизированной информационной системе на основе методики семантической аннотации и сравнения требований / Н. Н. Яковлев, Н.Д. Торопова // Альманах современной науки и образования. Тамбов: Грамота, 2010. № 11 (42): в 2-х ч. Ч. 2. С.117-119.

Яковлев Николай Николаевич

СИСТЕМНАЯ МОДЕЛЬ КОМПЛЕКСА ТРЕБОВАНИЙ  
К АВТОМАТИЗИРОВАННОЙ ИНФОРМАЦИОННОЙ  
СИСТЕМЕ НА ОСНОВЕ СЕМАНТИЧЕСКОЙ АННОТАЦИИ

Специальность 05.13.01 – Системный анализ, управление  
и обработка информации (в промышленности)

АВТОРЕФЕРАТ  
диссертации на соискание ученой степени  
кандидата технических наук

Подписано в печать 27.12.2010 Формат 60x84 1/16  
Бумага офсетная. Печать плоская. Гарнитура Times.  
Усл.печ.л. 1.0. Уч. – изд.л. 0,9  
Тираж 100 экз. Заказ № 535

ГОУ ВПО Уфимский государственный авиационный  
технический университет.  
Центр оперативной полиграфии  
450000, Уфа-центр, ул. К.Маркса, 12