

На правах рукописи



**Диковицкий Владимир Витальевич**

**МЕТОДЫ ИНТЕРФЕЙСНОЙ НАВИГАЦИИ И ПОИСКА НОРМАТИВНО-  
СПРАВОЧНЫХ ДОКУМЕНТОВ В КОРПОРАТИВНЫХ  
ИНФОРМАЦИОННЫХ СИСТЕМАХ**

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

**АВТОРЕФЕРАТ**

**диссертации на соискание ученой степени**

**кандидата технических наук**

Санкт-Петербург – 2016

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики» (Университет ИТМО).

**Научный руководитель:** доктор технических наук  
**Максим Геннадьевич Шишаев**

**Официальные оппоненты:** **Печников Андрей Анатольевич**  
Доктор технических наук, доцент  
Федеральное государственное бюджетное учреждение науки Институт прикладных математических исследований Карельского научного центра Российской академии наук, руководитель лаборатории телекоммуникационных систем

**Виноградов Геннадий Павлович**  
Доктор технических наук, профессор  
Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Тверской государственный технический университет», профессор кафедры «Информатика и прикладная математика»

**Ведущая организация:** Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Поволжский государственный технологический университет»

Защита состоится «29» сентября 2016 г. в «13» часов «00» минут на заседании диссертационного совета Д 002.199.01 при Федеральном государственном бюджетном учреждении науки Санкт-Петербургском институте информатики и автоматизации Российской академии наук по адресу: 199178, Санкт-Петербург, В.О., 14 линия, д. 39, ауд.401.

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Санкт-Петербургского института информатики и автоматизации Российской академии наук по адресу: 199178, Санкт-Петербург, В.О., 14 линия, д. 39 и на сайте <http://www.spiiras.nw.ru/dissovet/>.

Автореферат разослан «\_\_» \_\_\_\_\_ 2016 года.

Ученый секретарь  
диссертационного совета Д 002.199.01

кандидат технических наук



**Фаткиева Роза Равильевна**

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность работы** связана с возрастанием требований к эффективности информационного обеспечения организаций, деятельность которых требует аккумуляции и обновления знаний в различных предметных областях. Нормативно-справочные документы — классификаторы материалов и оборудования, регламенты — это стратегический актив компании, который предприятия используют в процессе постоянного информационного обмена. Многочисленность нормативных документов и их версий влечет дублирование и рост объема хранимой информации. Одни и те же процессы и объекты с разных сторон описываются в различных документах и могут рассматриваться специалистами с различных точек зрения, что усугубляет проблему поиска релевантных документов. Это обуславливает необходимость формирования единого информационного пространства предприятия в виде мультипредметной информационной системы предприятия (МИСП). Создание и поддержка функционирования такой информационной системы (ИС) порождает ряд важных задач, требующих научно обоснованного подхода к их решению: формирование семантической модели предметной области (СМПО) МИСП на основе интеграции существующих формализованных знаний, обеспечение актуальности контента единого информационного пространства в условиях динамики предметной области, обеспечение корректного и быстрого восприятия предоставляемой специалистам различных предметных областей информации, релевантности и пертинентности результатов информационного поиска.

Задача построения информационных систем, основанных на формализованных знаниях, известна довольно давно. По данной тематике опубликовано и выпущено множество статей и монографий, разработан целый ряд моделей и методов и спроектированных на их основе информационных систем, которые находят широкое применение в различных областях. Вместе с тем, несмотря на высокий уровень исследований в этой области, создание информационных систем, делающих возможным эффективным (в плане уменьшения времени) доступ к нормативно-справочной информации (НСИ) предприятия различным специалистам, остаётся сложной, до конца не решённой проблемой. Предлагаемым в данной работе решением является разработка новых методов автоматического формирования семантической модели предметной области НСИ, методов интерфейсной навигации и поиска документов с целью повышения эффективности доступа специалистов к НСИ. Повышение эффективности доступа к требуемой информации представляется возможным за счет адаптации пользовательского интерфейса к различным категориям пользователей. Использование автоматически формируемой семантической модели предметной области позволяет разделить семантику и контент информационных баз, что обеспечивает семантическую интеграцию НСИ предприятия.

**Цель работы** состоит в совершенствовании процессов поиска нормативно-справочной информации путем повышения полноты и точности поиска информации за счет автоматизированного формирования интегрированной семантической модели предметной области и разработке в рамках этой модели методов интерфейсной навигации и адаптивного поиска документов.

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

1. Анализ особенностей построения и использования информационных систем, основанных на знаниях;
2. Разработка архитектуры мультипредметной информационной системы

предприятия, основанной на семантической модели предметной области и моделях предпочтений пользователей.

3. Разработка методов формирования семантической модели предметной области информационной системы, интерфейсной навигации и поиска документов, реализующих уточнение семантической модели и адаптированное к различным пользователям представление информации.
4. Программная реализация и проверка эффективности разработанных методов.

**Методы исследований.** Для решения поставленных задач использованы методы системного анализа, теории графов, методы информационного поиска, математического моделирования, модульного и объектно-ориентированного программирования, искусственного интеллекта и инженерии знаний.

**Положения, выносимые на защиту:**

1. Метод автоматизированного формирования семантической модели предметной области, реализующий уточнение автоматически сформированных знаний.
2. Метод поиска документов, обеспечивающий автоматизированное расширение запроса и адаптивное ранжирование документов.
3. Метод интерфейсной навигации, обеспечивающий формирование навигационной структуры интерфейса, соответствующей модели предпочтений пользователя.
4. Комплекс программных средств, реализующий предложенные модели и методы для повышения эффективности доступа к ресурсам информационных систем.

**Научная новизна работы состоит в** создании моделей, алгоритмов и методов формирования и функционирования мультипредметных информационных систем. Применение данных методов позволяет повысить эффективность механизмов информационного поиска нормативно-справочной информации и человеко-машинного взаимодействия. Научной новизной обладают следующие результаты:

1. Разработан метод автоматизированного динамического формирования семантической модели предметной области мультипредметных информационных систем, использующий опыт пользователей для уточнения автоматически сформированных знаний, отличающийся интеграцией существующих формализованных знаний, результатов семантического анализа новых документов и моделей предпочтений пользователей.
2. Разработан метод поиска документов, обеспечивающий автоматизированное расширение запроса и оценку релевантности результатов поиска на основе совместного анализа модели предпочтений пользователя и семантической модели предметной области с учетом субтрактивных отношений.
3. Предложен метод интерфейсной навигации для формирования пользовательских интерфейсов мультипредметной информационной системы, адаптированных для различных категорий пользователей. Повышение эффективности человеко-машинного взаимодействия обеспечивается за счет отображения модели предпочтений пользователей на автоматически формируемую навигационную структуру интерфейса.
4. Создан комплекс программных средств для повышения эффективности доступа к документам организаций, отличающийся использованием методов, способных к автоматическому уточнению и адаптированному представлению информации организаций.

**Практическая значимость работы** состоит в том, что для решения задач исследования создан комплекс программных средств организации эффективного

доступа к ресурсам информационных систем, позволяющий наглядно и обозримо провести систематизацию НСИ предприятия. Применение разработанного комплекса программных средств позволяет усовершенствовать процессы обработки данных и знаний в компьютерных системах и сетях. Разработанный комплекс может успешно применяться в локальных и глобальных информационных системах.

**Апробация работы.** Научные положения и практические рекомендации диссертационной работы в целом, а также отдельные ее разделы докладывались и обсуждались на международных научно-технических конференциях «Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2012)» (г. Минск, 2012 г.), «Современные проблемы прикладной информатики» (г. Санкт-Петербург, 2011 г.), «Интеллектуальные системы и технологии: современное состояние и перспективы» (г. Тверь, 2011 г.), на всероссийских научных конференциях «Прикладные проблемы управления макросистемами» (г. Апатиты, 2010, 2014 гг.), «Теория и практика системной динамики» (г. Апатиты, 2011 г.), а также на семинарах лаборатории региональных информационных систем ИИММ КНЦ РАН.

**Внедрение результатов.** Результаты работы были использованы в информационной системе крупного промышленного предприятия региона (АО «Апатит»), в учебном процессе ФГБУ ВПО КФ ПетрГУ.

**Публикация результатов работы.** По теме диссертационной работы опубликовано 24 статьи, в том числе 4 статьи в изданиях, рекомендованных ВАК, оформлено 3 свидетельства о регистрации программ для ЭВМ.

**Структура и объем работы.** Диссертация объемом 153 машинописных страницы, содержит введение, четыре главы и заключение, список литературы (135 наименований), 4 таблицы, 38 рисунков, одно приложение с копиями актов внедрения.

## **СОДЕРЖАНИЕ РАБОТЫ**

Во введении обосновывается актуальность, научная новизна и практическая значимость диссертационной работы; перечислены положения, выносимые на защиту; сформулированы основные цели, задачи исследования, перечислены методы их решения. Кратко излагается содержание диссертации.

**В первой главе** предлагается общая характеристика решаемой в работе проблемы. Рассматриваются особенности предметной области и существенные проблемы повышения эффективности доступа к информации. Рассмотрены системы управления знаниями организаций, модели представления знаний, источники знаний, а также способы представления формализованных знаний в аспекте повышения эффективности доступа к информации. Выполнена оценка существующих решений, представлены основы формирования, интеграции, представления и использования формальных знаний для повышения эффективности доступа к информации и человеко-машинного взаимодействия. Рассмотрены подходы к интеграции гетерогенных источников информации. Обоснована эффективность применения для их решения методов системного анализа, математического моделирования, методов и моделей интеллектуальных систем. В заключении сформулированы выводы и основные требования к подходам повышения эффективности доступа пользователей корпоративных информационных систем к требуемой информации.

Во второй главе представлен поход к построению и концептуальная модель мультимедийной информационной системы организаций. Концептуальная модель МИСП представлена на рисунке 1.

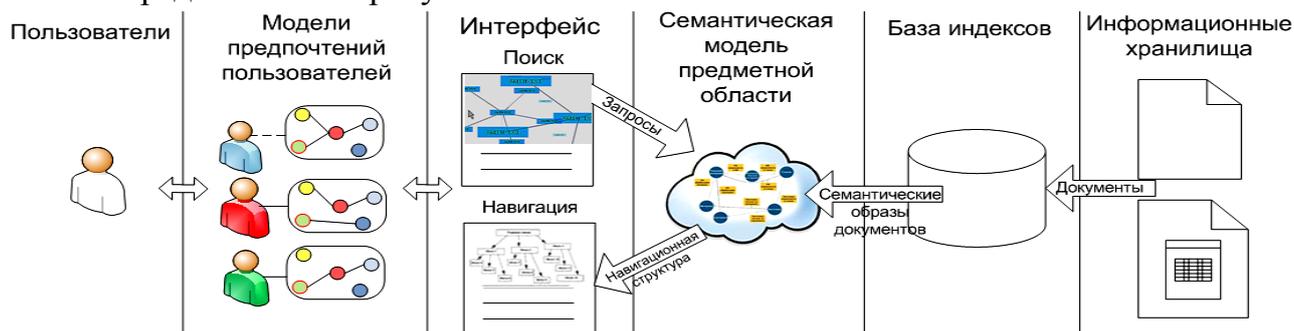


Рисунок 1 — Концептуальная модель мультимедийной информационной системы

Основу МИСП составляют семантическая модель предметной области, модель предпочтений пользователей (МПП), база индексов текстовых документов организации. Входными данными, обеспечивающими функционирование МИСП, являются документы и статистика работы пользователя. Выходными данными являются результаты обработки запросов и навигационная структура интерфейса. СМПО формируется в результате интеграции семантических образов документов организации. Семантический образ документа — семантическая сеть, множество вершин которой составляют понятия СМПО, присутствующие в документе, множество ребер — множество двухместных отношений над понятиями. Запрос — множество понятий предметной области, представленных множеством ключевых слов. Модель предпочтений пользователя — семантическая сеть, множество вершин которой составляют понятия СМПО, которыми оперирует пользователь, множество ребер — множество взвешенных двухместных отношений над понятиями, вес которых характеризует значимость семантического отношения между понятиями для пользователя, определенную на основе статистики его взаимодействия с системой. МПП служит «эталоном» в процессе формирования пользовательского интерфейса и его оценки как степени соответствия структуры интерфейса модели предпочтений пользователя. МПП формируется на основе обработки запросов пользователя и статистики его работы. Взаимодействие пользователя и мультимедийной информационной системы может быть представлено следующим алгоритмом:

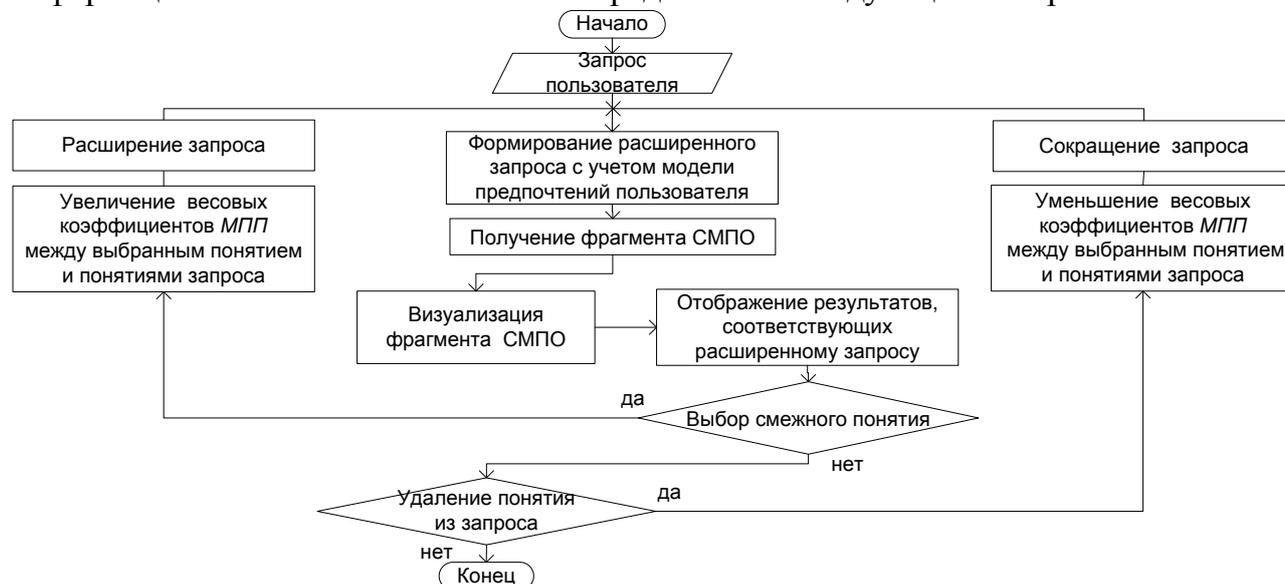


Рисунок 2 — Алгоритм взаимодействия пользователя и ИС

Взаимодействие заключается в итеративном расширении запросов пользователя на основе МПП, обеспечении возможности коррекции запроса средствами адаптивного интерфейса, а также учета пользовательских предпочтений путем коррекции весовых коэффициентов между понятиями МПП. При совместном употреблении в запросе пар понятий, входящих в состав МПП, изменяются весовые коэффициенты отношений между данными понятиями. Величина изменения определяется формулами 3,4. Семантическая модель предметной области, модель предпочтений пользователей и база индексов документов составляют системные информационные базы, используемые в дальнейшем для формирования навигационно-поискового интерфейса, обеспечивающего эффективный (в смысле скорости и релевантности) доступ пользователя к требуемым данным. Диаграмма сценариев использования мультипредметной информационной системы представлена далее:



Рисунок 3 — Диаграмма сценариев использования мультипредметной информационной системы

В третьей главе представлены методы, обеспечивающие формирование и эффективное функционирование компонентов МИСП. Ключевым компонентом системных информационных баз МИСП является модель предпочтений пользователя, используемая для адаптации информационного содержания интерфейса и ранжирования результатов поиска информации. С целью снижения объемов хранимых системных данных модель строится не для каждого пользователя в отдельности, а для групп пользователей. Формально категории пользователей определяются следующим образом:

Пусть  $C$  — некоторое множество понятий (концептов),  $U$  — множество пользователей. Каждый концепт  $c$  имеет множество атрибутов:

$$A(c) = \{a(c)_i\}, a(c)_i \in C, i = \overline{1, N_c}.$$

Упорядочив множество атрибутов по убыванию степени их значимости для пользователя  $u$ , получим последовательность, характеризующую его представление о данном концепте:

$$A^u(c) = \{a^u(c)_i\}, i = \overline{1, N_c} : a^u(c)_i \varphi^u a^u(c)_j, \forall i \leq j,$$

где  $\varphi^u$  — отношение, задающее значимость атрибутов для пользователя  $u$ ;  $a\varphi^u b$  означает, что «для пользователя  $u$   $a$  не менее значим, чем  $b$ ».

Определим группу пользователей, имеющих схожие представления о понятиях из некоторого множества  $C$ . Назовем подобную группу пользовательской *категорией*  $k$ -го порядка на множестве концептов  $C$ , и определим ее следующим образом:

$$U_C^k = \{u \mid \{a^u(c)_i\} = \{a^{u'}(c)_i\}, i = \overline{1, k}, \forall c \in C, \forall u' \in U_C^k\}.$$

Модель предпочтений некоторой  $k$ -й группы пользователей представлена взвешенным мультиграфом:

$$UM_k = \{C, L_k\},$$

где  $C$  — множество вершин графа, представляющих понятия СМПО,  $L_k = \{l_k^{ijm}\}$  — множество взвешенных дуг, вес которых характеризует значимость семантического отношения  $m$ -го типа между  $i$ -м и  $j$ -м понятиями для  $k$ -й категории пользователей.

В совокупности, МПП всех пользователей образуют фрагмент базы знаний МИСП, представляющий собой мультиграф с векторными весами дуг. Матрица инцидентности мультиграфа имеет размерность 3:

$$M_I : C \times L \times U_k \rightarrow w_k^{ij},$$

где  $C$  — множество вершин графа, представляющих понятия СМПО,  $L$  — множество дуг, задающих отношения над  $C$ ,  $U_k$  — множество категорий пользователей.

Элементами матрицы являются весовые коэффициенты  $w_k^{ij}$ , задающие вес связи между концептами  $C_i$  и  $C_j$  для  $k$ -й категории пользователей.

**Метод автоматизированного формирования семантической модели предметной области** информационной системы на основе принципа «пользователь как эксперт».

Формально СМПО представлена неоднородной  $n$ -арной семантической сетью:

$$KB = \{C, L, Tp\}, Tp = \{synonymOf, HyponymOf, associateWith, subStruct\} \quad (1)$$

$$L = \{l\}, l = \langle c_i, c_j, tp, \bar{w} \rangle, c_i, c_j \in C, tp \in Tp, \bar{w} = \langle w_1, \dots, w_k, \dots, w_r \rangle \quad (2)$$

где  $C$  — множество концептов,  $L$  — множество отношений над концептами,  $\bar{w}$  — вектор весовых коэффициентов,  $Tp$  — множество типов отношений (синонимии, гипонимии, ассоциации, субтрактивных отношений),  $r$  — количество категорий пользователей.

Процесс формирования семантической модели предметной области предприятия на основе коллекции документов информационной системы и расширяемого тезауруса состоит из следующих этапов:

1. Формирование семантического образа документа. Семантический образ задан семантической сетью, полученной статистическими и лексико-грамматическими методами обработки текста:

$$D = \{C^D, L^D\}, C^D \subset C, L^D \subset L,$$

где  $C^D$  — множество концептов, выделенных в документе,  $L^D$  — множество отношений вида (2), выделенных в документе.

2. Интеграция семантических образов в СМПО на основе модифицированной семантической метрики следующим образом:

1. Вычисление оценки сходства имен концептов документа и СМПО:

$$\forall c_i : Eq(c_i, c_j) = \frac{x/len}{\max(len_i, len_j)}, c_i \in C^D, c_j \in C^{KB}, i = \overline{1, N_D}, j = \overline{1, N_{KB}},$$

где  $Eq$  — функция оценки сходства имен двух концептов,  $x$  — длина эквивалентной цепочки символов в именах понятий,  $len$  — длина имен концептов.

II. Вычисление оценки сходства контекста концептов документа с контекстом СМПО:

$$\forall c_i : Syn(c_i, c_j) = \frac{|C_{syn}^D \cap C_{syn}^{KB}|}{|C_{syn}^{KB}|}, c_i \in C^D, c_j \in C^{KB}, i = \overline{1, N_D}, j = \overline{1, N_{KB}},$$

где  $Syn$  — функция оценки сходства контекста двух концептов,  $C_{syn}^D$ ,  $C_{syn}^{KB}$  — множество синонимов концепта  $c_i$  и  $c_j$  соответственно.

III. Вычисление оценки сходства структурного положения концептов документа с контекстом СМПО:

$$\forall c_i : Poseq(c_i, c_j) = \frac{|C_{Hyp}^D \cap C_{Hyp}^{KB}|}{|C_{Hyp}^{KB}|}, c_i \in C^D, c_j \in C^{KB}, i = \overline{1, N_D}, j = \overline{1, N_{KB}},$$

где  $Poseq$  — функция оценки сходства структурного положения двух концептов,  $C_{Hyp}^D$ ,  $C_{Hyp}^{KB}$  — множество гиперонимов концепта  $c_i$  и  $c_j$  соответственно.

IV. Добавление концептов на основании результатов вычисления пороговой функции от среднего трех оценок:

$$f(c_i, c_j) = \frac{a \cdot Eq(c_i, c_j) + b \cdot Poseq(c_i, c_j) + c \cdot Syn(c_i, c_j)}{3} > z,$$

$$c_i \in C^D, c_j \in C^{KB}, i = \overline{1, N_D}, j = \overline{1, N_{KB}},$$

где  $z$  — значение пороговой функции,  $a, b, c$  — некоторые коэффициенты, которые определяются экспертно, исходя из объема и разнородности предметной области.

3. Уточнение СМПО пользователями осуществляется путем изменения весовых коэффициентов существующих отношений между понятиями. Данный процесс инициируется при совместном использовании двух понятий в одном пользовательском запросе. Величина изменения весового коэффициента определяется согласно формуле (3) для весового коэффициента отношения между совместно использованными понятиями, и формуле (4) для весового коэффициента отношения между остальными (неиспользованными) понятиями, где  $dw(c_i, c_j)$  — величина изменения весового коэффициента,  $w(c_i, c_j)$  — значение весового коэффициента отношения между понятиями  $c_i$  и  $c_j$ ,  $ACW$  и  $CW$  — множества отношений с совместно использованными и неиспользованными понятиями.

$$dw(c_i, c_j) = \frac{\sum_{\langle c_i, c_j \rangle \in (ACW)} (w(c_i, c_j))}{|ACW|} \quad (3)$$

$$dw(c_i, c_j) = - \frac{\sum_{c_i \in ACW, c_j \in CW} (w(c_i, c_j))}{|CW / ACW|} \quad (4)$$

Предложенный метод предлагается использовать для автоматизированного формирования и обеспечения актуального состояния семантической модели предметной области организаций в условиях динамики предметной области. На основе

полученной модели решаются задачи обеспечения эффективного доступа пользователей к НСИ, а именно реализуются методы поиска и интерфейсной навигации.

**Метод интерфейсной навигации** обеспечивает повышение эффективности взаимодействия пользователя с МИСП за счет автоматизированного формирования навигационной структуры интерфейса, адекватной модели предпочтений пользователя.

Пользовательский интерфейс представляет собой пару:

$$UI = \langle I, s \rangle,$$

где  $I$  — множество информационных элементов;  $s$  — навигационная структура.

Навигационная структура определяет иерархию групп информационных элементов (ИЭ) или доступных для пользователя действий. При этом на каждом уровне иерархии исходное множество информационных элементов делится на подмножества в соответствии с одним или несколькими классификационными признаками. В качестве классификационных признаков используются атрибуты понятий предметной области. Очевидно, что при использовании на одном уровне навигационной структуры нескольких признаков, полученные множества ИЭ могут пересекаться. Введем следующие обозначения:

$\Gamma^l = \{G_i^l\}$  — множество разделов  $l$ -го уровня навигационной структуры;

$G_i^l$  —  $i$ -я группа информационных элементов  $l$ -го уровня навигационной структуры;

$P^l = \{p_i^l\}$  — множество классификационных признаков, используемых для формирования групп ИЭ на  $l$ -м уровне навигационной структуры.

Заметим, что использование информационной системы представляет собой, по сути, поиск некоторых информационных элементов по имеющемуся у человека образу. При этом образ, чаще всего, неточный: в нем специфицируются лишь некоторая часть идентифицирующих атрибутов. Вследствие этого, пользователь с разной степенью уверенности может предполагать, в какой из групп ИЭ на некотором уровне навигационной структуры находится искомый элемент. Эта уверенность тем выше, чем более точно представляет пользователь потенциальное содержимое группы. Введем следующую функцию, задающую числовую оценку степени уверенности пользователя  $u$  (чем выше значение, тем выше степень уверенности):

$$p^u : \Gamma^l \rightarrow [0,1] .$$

Оценка времени, требуемого для доступа к искомому информационному элементу в рамках навигационной структуры на  $l$ -м уровне, будет равна

$$O\left(\frac{\max |G_i^l|}{p^u(G_i^l)}\right).$$

Таким образом, при прочих равных, степень уверенности пользователя в принадлежности информационного элемента к той или иной группе определяет качество интерфейса в смысле скорости доступа к требуемой информации.

Пусть навигационная структура интерфейса имеет глубину  $\hat{l}$  уровней. Тогда в качестве количественной оценки эффективности интерфейса для пользователя  $u$  может использоваться сумма:

$$\sum_{l=1}^{\hat{l}} p^u(l).$$

Данная мера может использоваться для оценки уже существующих интерфейсов, когда известно значение  $\hat{l}$ . Для решения же прямой задачи, то есть структуризации исходного множества информационных элементов в рамках навигационной структуры, требуется учитывать дополнительные ограничения. Эти ограничения обусловлены психологией восприятия человека, ограничивающей максимальное количество одновременно эффективно воспринимаемых объектов. Вследствие этого необходимо ограничивать размер группы информационных элементов, а также глубину навигационной структуры.

С учетом сказанного, оптимальная для пользователя  $u$  структура интерфейса есть решение следующей задачи с ограничениями:

$$\max_s \sum_{l=1}^{\hat{l}(s)} p^u(l), g(s) \leq K, \hat{l}(s) \leq K'.$$

Здесь  $\hat{l}(s)$  — количество уровней в навигационной структуре  $s$ ;  $g(s)$  — максимальный размер группы информационных элементов  $\hat{l}(s)$ -го в навигационной структуре  $s$ ;  $K$  — константа, определяющая максимальное число одновременно предъявляемых пользователю информационных элементов для их эффективного восприятия;  $K'$  — константа, определяющая максимальное число уровней навигационной структуры, в рамках которых поиск информации для пользователя остается комфортным.

Для формирования навигационного интерфейса, удовлетворяющего приведенной формулировке задачи синтеза, предлагается процедура, основанная на МПП. Процедура содержит несколько этапов:

1. Определение текущей информационной потребности пользователя на основе модели пользовательских интересов и текущего запроса:

$$UQ = f_m(Q, UM_k), \forall c_i : \exists l : (c_i \in Q) \wedge (c_j \in UM_k) \wedge (\overline{w_k} > 0),$$

$$Q = \{c_i\}, l = \langle c_i, c_j, tp, \overline{w} \rangle, i = \overline{1, N_Q}, j = \overline{1, N_L}, \quad (5)$$

где  $UM_k$  — модель предпочтений  $k$ -ой категории пользователей;  $Q$  — запрос,  $f_m$  — функция, ставящая соответствие запросу фрагмент модели предпочтений,  $\overline{w_k}$  —  $k$ -ый компонент вектора весовых коэффициентов,  $l$  — отношение между концептами  $c_i$  и  $c_j$  в модели  $UM_k$

2. Определение множества информационных элементов интерфейса, соответствующих текущей информационной потребности:

$$G = \{g \mid \exists c_i = g\}, c_i \in UQ, i = \overline{1, N_{UQ}},$$

где  $G$  — множество информационных элементов навигационной структуры;  $c_i$  — концепты текущей информационной потребности пользователя.

3. Разбиение множества информационных элементов интерфейса на подмножества и их ранжирование в соответствии с весовыми коэффициентами модели пользовательских интересов.

$$G_i^d = \{g_k \mid (\forall g_k, g_m \exists l \in L : w(c_k, c_m) > x) \wedge (\exists g_z : \exists l_{\text{Гип}}(c_k, c_z), l_{\text{Гип}}(c_m, c_z) \in L)\},$$

$$k, m, z = \overline{1, N_G}, z = \overline{1, N_L}, G_i^d \subset G, L \in UM, \quad (6)$$

где  $G_i^d$  —  $i$ -я группа информационных элементов  $d$ -го уровня навигационной

структуры;  $l_{\text{гип}}(c_k, c_z)$  — отношение гипонимии в модели интересов пользователя;  $w(c_k, c_m)$  — весовой коэффициент отношения  $l$  над концептами  $c_k, c_m$ ;  $x$  — порог вхождения информационного элемента в навигационную структуру;  $d$  — количество уровней навигационной структуры, задается на основе ограничений на максимальное число информационных элементов.

На основе групп множеств  $G$  формируется интерфейс мультимедийной информационной системы, отражающий информационное содержание ресурсов в терминах СМПО с учетом модели предпочтений пользователей.

**Метод поиска информации с учетом субтрактивных отношений** позволяет увеличить эффективность поиска, в смысле релевантности результатов, за счет автоматического включения в запрос ограничений. Ограничения формируются на основе субтрактивных отношений искомого концепта, заданных в МПП. Субтрактивные отношения — это отношения между концептами, имеющие отрицательный весовой коэффициент. Использование субтрактивных отношений позволяет автоматизировать процесс добавления ограничений в расширенный запрос.

Метод поиска включает 3 составляющих: модель документа, модель запроса и функцию соответствия между ними. Документ представлен фрагментом СМПО и множеством ключевых слов в базе индексов, выделенных семантическим анализатором на этапе индексации. Запрос представлен в виде множества ключевых слов. Процесс поиска документов, соответствующих запросу состоит из следующих этапов:

1. Формирование запроса в терминах СМПО:

1.1. Формирование расширенного запроса, содержащего отношения и соответствующие запросу концепты СМПО:

$$EQ = f_q(Q, KB) = \{C^Q, L^Q \mid (Eq(c_i^Q, c_j^{KB}) > 1 - \varepsilon)\},$$

$$C^Q \subset C, L^Q \subset L \quad i = \overline{1, N_Q}, j = \overline{1, N_{KB}}, \quad (7)$$

где  $KB$  — СМПО,  $C^q$  — множество концептов СМПО, содержащихся в запросе,  $L^Q$  — множество отношений над концептами  $C^q$ ,  $f_q(Q, KB)$  — функция, ставящая соответствие запросу фрагмент СМПО,  $Eq(c_i^Q, c_j^{KB})$  — функция оценки сходства имен двух концептов,  $\varepsilon$  — погрешность оценки сходства концептов.

1.2. Расширение запроса с учетом весовых коэффициентов отношений и субтрактивных отношений, ограничивающих контекст запроса:

$$EQ = \{C^Q, L^Q\} \cup \{C', L' \mid l: c_i \in C^Q, c_j \in C', |\overline{w_k}| > x\},$$

$$C' \subset C, L' \subset L, l \in L', \quad (8)$$

где  $C'$  — множество концептов СМПО, связанных с концептами множества  $C^Q$  отношениями вида (2) из множества  $L'$ ,  $\overline{w_k}$  —  $k$ -ая компонента вектора весовых коэффициентов отношения  $l$ ,  $x$  — коэффициент включения отношения в расширенный запрос.

2. Получение множества документов  $D$ , соответствующих расширенному запросу:

$$D = \{d_i \mid C^{d_i} \cap C^Q \neq \emptyset\}, \quad i = \overline{1, n}, \quad (9)$$

где  $C^{d_i}$  — множество концептов СМПО, присутствующих в документе  $d_i$ ,  $C^Q$  — множество концептов СМПО, присутствующих в запросе  $EQ$ .

3. Ранжирование множества документов с учетом весовых коэффициентов отношений:

$$R(d_k) = \sum_{L_{d_k}} (f_u(\bar{w}_k, r)) - \sum_{L'_{d_k}} (f_u(\bar{w}_k, r)), \quad (10)$$

$$L_{d_k} = \{l^d \mid (c_i, c_j \in d_k) \wedge (tp \in \{synonymOf, HyponymOf, associateWith\})\}$$

$$L'_{d_k} = \{l^d \mid (c_i, c_j \in d_k) \wedge (tp \in \{subStract\})\}, i, j = \overline{1, n}, k = \overline{1, m}, tp \in Tp,$$

где  $f_u(\bar{w}_k, r)$  — функция получения компоненты вектора весовых коэффициентов отношений из множества  $L_{d_k}$  между концептами  $c_i$  и  $c_j$ , присутствующими в документе  $d_k$ , для категории пользователей  $r$ .  $Tp$  — множество типов отношений. Таким образом, документы, в которых присутствуют субтрактивные отношения, будут иметь меньший приоритет после ранжирования. Результатом ранжирования является упорядоченное по убыванию оценки  $R$  множество документов, представляющих результаты поиска.

В четвертой главе представлено прикладное программное обеспечение (ППО), реализующее предложенные методы. Приводятся результаты вычислительных экспериментов с использованием разработанного ППО.

Архитектура программного комплекса мультипредметной информационной системы предприятия представлена на рисунке 4.

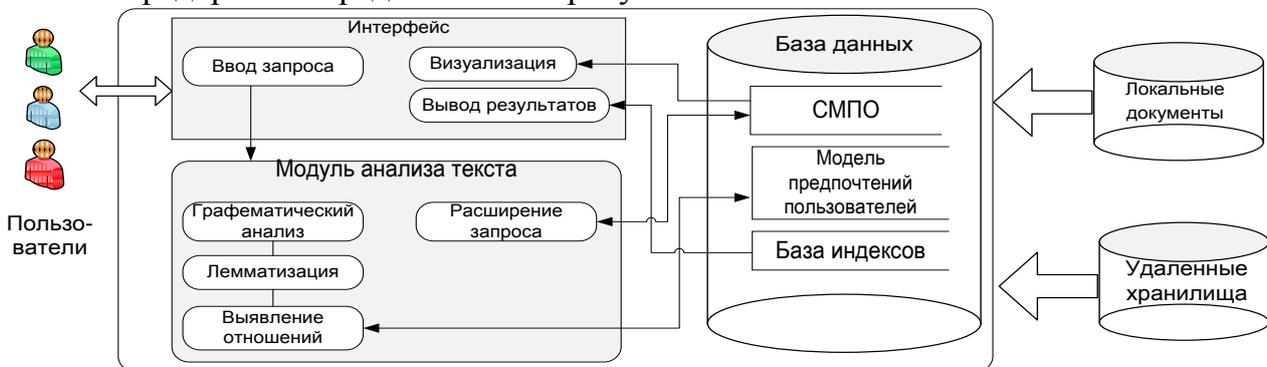


Рисунок 4 — Архитектура программного комплекса

Предложенные методы были апробированы в информационной системе крупного промышленного предприятия региона — АО «Апатит», а также в рамках документооборота Кольского филиала ФГБОУ ВПО «Петрозаводский государственный университет». Семантическую модель предметной области мультипредметной информационной системы составляет русскоязычный тезаурус WordNet 3.0, расширенный результатами работы семантического анализатора над коллекцией нормативно-справочных документов организаций. Эффективность предложенных методов была проверена путем натурального эксперимента с привлечением тестовой выборки пользователей, которым предлагалось решить схожие информационно-поисковые задачи с помощью существующих систем поиска с использованием интерфейса строки поиска и с помощью адаптивного интерфейса мультипредметной информационной системы.

Для оценки характеристик адаптивного пользовательского интерфейса был использован метод GOMS, заключающийся в сравнении временных характеристик выполнения пользователями производственных задач в рамках традиционного и разрабатываемого интерфейса. Целью пользователей являлось получение доступа к требуемому документу. Среднее время выполнения операции представлены в таблице 1.

Таблица 1 — Оценка показателей использования методов навигации

Оцениваемые характеристики	Навигация с использованием традиционного интерфейса	Навигация с использованием адаптивного интерфейса
Среднее время навигационного поиска, с.	31,8	20,85

Для оценки эффективности предложенного метода поиска осуществлялся поиск по заранее проиндексированной коллекции объемом 14 тыс. документов. Экспертами оценивались результаты выполнения 10 запросов. В качестве критериев оценки выступали скорость поиска – время, затраченное на удовлетворение информационной потребности, выраженной одним запросом; точность — соответствие результатов запросу; и полнота результатов — полнота охвата документов с упоминанием об объекте поиска.

$$Precision = \frac{|D_{rel} \cap D_{retr}|}{|D_{retr}|}, \quad Recall = \frac{|D_{rel} \cap D_{retr}|}{|D_{rel}|},$$

где  $D_{rel}$  — множество релевантных документов в базе индексов,  $D_{retr}$  — множество документов, найденных системой.

Для оценки альтернатив экспертам была предложена лингвистическая шкала измерений.

Оценка  $i$ -й альтернативы производилась  $j$ -м экспертом по формуле:

$$v_{ij} = 1 - \frac{(l-1)}{k},$$

где  $l$  — индекс значения лингвистической шкалы;  $k$  — количество значений этой шкалы.

Для оценки  $i$ -й альтернативы  $n$  экспертами используется формула:

$$s_i = \sum_{j=1}^n v_{ij}$$

Результаты оценки характеристик приведены в таблице 2.

Таблица 2 — Оценка показателей использования методов поиска

Оцениваемые характеристики	Используемый метод поиска документов	
	Поиск по ключевым словам с использованием интерфейса строки ввода	Поиск с использованием метода поиска информации с учетом субтрактивных отношений
Оценка скорости выполнения поиска	0,5	0,8
Точность результатов	0,9	0,9
Полнота результатов	0,7	0,9
Среднее значение оценок	0,7	0,9

Высокая точность поиска с использованием интерфейса строки ввода обусловлена знакомством пользователей с предметной областью и, как следствие, малой неопределенностью при формировании запроса, а также относительно малым объемом коллекции документов. Тем не менее, результаты экспериментов позволяют сделать вывод о корректности и обоснованности использования динамической автоматически формируемой и верифицируемой СМПО для реализации адаптивного интерфейса и метода поиска информации с учетом субтрактивных отношений в корпоративных информационных системах.

## **ЗАКЛЮЧЕНИЕ**

В диссертационной работе содержится решение научно-технической задачи разработки методов интерфейсной навигации и поиска нормативно-справочных документов в корпоративных информационных системах. В ходе исследования получены следующие результаты:

1. Разработан метод автоматизированного формирования семантической модели предметной области информационной системы организаций на основе принципа «пользователь как эксперт», заключающийся в интеграции на основе модифицированной составной семантической метрики разнородных источников знаний и последующего уточнения знаний пользователями;
2. Разработан метод поиска информации на основе формализованных знаний, учитывающий весовые коэффициенты отношений семантической модели предметной области для различных категорий пользователей и субтрактивные отношения, ограничивающие область поиска;
3. Разработан метод интерфейсной навигации, реализующий динамическое формирование адаптивного интерфейса, реализующего обратную связь с пользователем;
4. Создан комплекс программных средств для повышения эффективности доступа к документам организаций, отличающийся использованием методов, способных к автоматическому уточнению и адаптированному представлению информации организаций. Результаты проведенного анализа эффективности разработанных методов мультипредметных корпоративных информационных систем с использованием разработанного программного обеспечения показали, что при осуществлении информационного поиска обеспечивается сокращение времени, необходимое для доступа к необходимой информации примерно в 1,5 раза, и увеличение полноты результатов примерно в 1,3 раза при сохранении точности результатов информационного поиска.

Аспекты научной новизны соответствуют областям исследований п.3 «Модели, методы, алгоритмы, языки и программные инструменты для организации взаимодействия программ и программных систем» и п.7 «Человеко-машинные интерфейсы; модели, методы, алгоритмы и программные средства машинной графики, визуализации, обработки изображений, систем виртуальной реальности, мультимедийного общения» специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

## **ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

**В рецензируемых журналах из списка ВАК и изданиях, приравненных к ним:**

1. Ломов, П.А. Онтологическая модель государственного и муниципального управления для проведения семантической интеграции информационных ресурсов /

- М.Г. Шишаев, П.А. Ломов, В.В. Диковицкий // Труды ИСА РАН: Том 59, - М. : КРАСАНД : URSS, 2010. С. 118-132.
2. Шишаев, М.Г. Использование концепции «User as an expert» в разработке мультипредметных веб-ресурсов, основанных на онтологиях/ М.Г. Шишаев, П.А. Ломов, В.В. Диковицкий // Труды ИСА РАН: Информационные технологии. Системное моделирование. Численные методы решения. Компьютерный анализ текстов. Том 62. Выпуск 3. – М. КРАСАНД, 2012, С. 40-47.
  3. Диковицкий, В.В. Технология формирования адаптивных пользовательских интерфейсов для мультипредметных информационных систем промышленных предприятий / В. В. Диковицкий, М.Г. Шишаев // Информационные ресурсы России. 2014. № 1(137). С. 23–26.- ISSN 0204–3653.
  4. Диковицкий, В.В. Методы интеллектуальной обработки и представления информации в мультипредметных информационных системах промышленных предприятий // Труды СПИИРАН. 2015. Вып. 42. С. 56-76.
- В других изданиях:**
5. Шишаев, М.Г. Использование онтологий для независимого от реализации представления бизнес-логики прикладной ИС / М.Г. Шишаев, В.В. Диковицкий, Л.П. Попова // Информационные технологии в региональном развитии – Апатиты: изд-во КНЦ РАН, 2009. – Вып. IX. - С. 51-55
  6. Диковицкий, В.В. Концепция двунаправленного семантического поиска / В.В. Диковицкий, М.Г. Шишаев, П.А. Ломов // Прикладные проблемы управления макросистемами: мат. VIII Всерос. Школы - семинара, 29 марта - 2 апреля 2010 г.
  7. Диковицкий, В.В. Система интеграции ВЕБ-ресурсов: Модуль динамического формирования поисковых запросов: тез. докл. XII Межрег. научно-практ. конф., 15-17 апр. 2009 г. - Апатиты: КФ ПетрГУ, 2009. –Ч.1. – С.8.
  8. Ломов, П.А. Онтологическая модель государственного и муниципального управления для проведения семантической интеграции информационных ресурсов/ П.А. Ломов, В.В. Диковицкий, М.Г. Шишаев, // мат. VIII Всерос. Школы - семинара, 29 марта - 2 апреля 2010 г.
  9. Диковицкий, В.В. Обработка текстов естественного языка в моделях поисковых систем / В.В. Диковицкий, М.Г. Шишаев // Труды Кольского научного центра РАН. Информационные технологии. – Апатиты, 2010. – Вып. 1. – С. 29-34.
  10. Диковицкий, В.В. Российская Арктика в современном веб-пространстве / В.В. Диковицкий, М.Г. Шишаев, М.В. Мальков, Р.Р. Сепеда-Эррера // Материалы докладов IV-я Всероссийская научная конференция «Теория и практика системной динамики». – Апатиты, 2011. – С. 10-12.
  11. Диковицкий, В.В. Метод визуального семантического поиска на базе семантической сети с субтрактивными связями/ В.В. Диковицкий, М.Г. Шишаев // Современные проблемы прикладной информатики: сб. науч. трудов междунар. науч.-практ. конф. 25-27 мая 2011 г. / отв. ред. И.А. Брусакова, И.Л. Андреевский. - СПб.: Изд-во Политехн. Ун-та, 2011 г. С.158-161
  12. Диковицкий, В.В. Применение метода семантического поиска на основе семантической сети с субтрактивными связями для реализации сервисов интернет-портала / В.В. Диковицкий, М.Г. Шишаев // Интеллектуальные системы и технологии: современное состояние и перспективы. Сборник научных трудов Международной летней школы-семинара по искусственному интеллекту для студентов, аспирантов и молодых ученых (Тверь – Протасово, 1-6 июля 2011 г.) – Тверь: Изд-во Тверского

государственного технического университета, 2011. – 292 с. сс. 196-200

13. Шишаев, М.Г. Современные методы создания мультимедийных веб-ресурсов на базе визуализации и обработки формализованной семантики / М.Г. Шишаев, В.В. Диковицкий, П.А. Ломов, Р.Р. Сепеда-Эррера Р.Р. // Вестник КНЦ 2011-сс.62-72.

14. Диковицкий, В.В. Веб-сервис представления сетевых структур на основе методов визуального анализа информации. Свидетельство о регистрации электронного ресурса № 17472 10.10.2011

15. Диковицкий, В.В. Веб-сервис динамического формирования навигационных структур. Свидетельство о регистрации электронного ресурса № 17472 17.11.2011

16. Диковицкий, В.В. Извлечение знаний пользователя и верификация знаний самоорганизующихся информационных систем с обратной связью// Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2012): материалы II Междунар. научн.-техн. конф. / редкол. : В. В. Голенков (отв. ред.) [и др.]. – Минск : БГУИР, 2012. – 548 с.

17. Ломов, П.А., Преобразование OWL-онтологий для визуализации и использования в качестве основы пользовательского интерфейса / П.А.Ломов, М.Г. Шишаев, В.В. Диковицкий // Научный журнал «Онтология проектирования» - №3-2012. – Самара: Новая техника, 2012, сс. 49-61 ISSN 2223-9537

18. Ломов, П. А. Технология упрощенного представления OWL онтологий для их использования в графических пользовательских интерфейсах / П.А.Ломов, М.Г. Шишаев, В.В. Диковицкий // Сборник трудов конференции «Инженерия знаний и технологии семантического веба - 2012». - СПб: НИУ ИТМО, 2012. - с. 54 - 66.

19. Диковицкий, В.В., Метод информационного поиска на основе динамической расширяемой базы знаний // Труды Кольского научного центра РАН. 4/2012(11). Информационные технологии. Выпуск.3, с. 85-88.

20. Диковицкий, В.В., Ломов П. А., Программный комплекс формирования ментальной модели пользователя сети интернет. Свидетельство о регистрации электронного ресурса в ФГАНУ "ЦИТиС" номер 50201450815 от 04.12.2014.

21. Диковицкий, В.В., Семантическое профилирование пользователей в задаче информационного поиска // Труды Кольского научного центра РАН. Информационные технологии. – Вып.6. - 3/2015(29) -Апатиты: Изд-во КНЦ РАН, 2015. – с.54-58 - ISBN 978-5-91137-317-7

22. Shishaev, M.G. Architecture and Technologies of Knowledge-Based Multi-Domain Information Systems for Industrial Purposes / V.V. Dikovitsky, M. G. Shishaev , N. V. Nikulina // Automation Control Theory Perspectives in Intelligent Systems. Proceedings of the 5th Computer Science On-line Conference 2016 (CSOC2016), Vol 3. – 2016. Pp. 359 – 369.