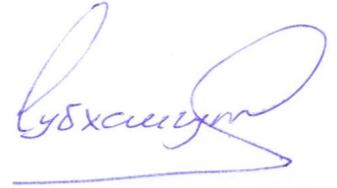


На правах рукописи



Субхангулов Руслан Айратович

ОНТОЛОГИЧЕСКАЯ ИНФОРМАЦИОННАЯ ПОДДЕРЖКА  
ПРОЕКТИРОВАНИЯ В ЭЛЕКТРОННЫХ АРХИВАХ ТЕХНИЧЕСКОЙ  
ДОКУМЕНТАЦИИ

Специальность 05.13.12 – Системы автоматизации проектирования  
(промышленность)

Автореферат  
диссертации на соискание ученой степени  
кандидата технических наук

Ульяновск – 2015

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

Научный руководитель: кандидат технических наук, доцент  
**Наместников Алексей Михайлович**

Официальные оппоненты: **Новикова Светлана Владимировна**  
доктор технических наук, доцент,  
Казанский национальный  
исследовательский технический  
университет им. А.Н. Туполева-КАИ,  
кафедра прикладной математики и  
информатики, профессор

**Радионова Юлия Александровна**  
кандидат технических наук, ФНПЦ АО  
«НПО «Марс», ведущий инженер-  
программист

Ведущая организация: **Институт автоматизации и процессов  
управления ДВО РАН, г. Владивосток**

Защита диссертации состоится «16» декабря 2015 г. в 15 часов на заседании диссертационного совета Д212.277.01 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.

Автореферат разослан «\_\_» \_\_\_\_\_ 2015 г.

Ученый секретарь диссертационного совета, доктор технических наук, профессор



Смирнов Виталий Иванович

## Общая характеристика работы

### Актуальность работы

Современные электронные архивы (ЭА) крупных проектных организаций насчитывают десятки и сотни тысяч документов, относящихся к проектно-технической документации. В таких ЭА сосредоточен значительный опыт проектирования сложных технических систем, к которым можно отнести современные автоматизированные системы (АС). Проектирование новой системы крайне редко осуществляется без обращения к предыдущему опыту разработок и поиска аналогов. Такие исследователи, как Норенков И.П., Тарасов В.Б., Collins Н. отмечают тот факт, что при увеличении объема ЭА затрудняется анализ документов по заранее заданным реквизитам. Вследствие распределенного проектирования от участников проекта часто требуются навыки в области семантического анализа технической документации и выполнение корректных предметно-ориентированных проектных запросов к ЭА.

В основе модели информационной поддержки лежат механизмы выполнения интеллектуальных запросов, которые предполагают способность поисковой системы к самоорганизации, осуществление независимого общения с пользователем, эффективный поиск текстовых документов, реагирующий на изменения информационной потребности пользователя.

Учет специфики проектных знаний приводит к необходимости формирования онтологии электронного архива особой структуры, позволяющей учитывать контекст принимаемых проектных решений, жизненный цикл проектируемой АС и систему индивидуальных предпочтений проектировщиков. Таким образом, система информационной поддержки, как неотъемлемая часть электронного архива проектной организации, должна обладать свойствами интеллектуальной системы. Известные исследователи в области онтологических систем, такие как Загорулько Ю.А., Гаврилова Т.А., Соловьев В.Д., Лукашевич Н.В., Добров Б.В., Ландэ Д.В., Смирнов С.В., Соснин П.И., Gruber T.R., Berners-Lee T., Uschold M. и другие отмечают важность и актуальность исследований,

базирующихся на онтологическом подходе в сложно-структурированных предметных областях. В работах Соснина П.И., Смирнова С.В. отмечается важность применения систем, основанных на знании, в процессе человеко-компьютерного взаимодействия при нахождении проектных решений.

В настоящее время не существует математических методов и алгоритмов, позволяющих выполнять контекстно-ориентированные запросы к ЭА с учетом текущего состояния проекта и индивидуальных предпочтений проектировщика. Следовательно, актуальным является разработка моделей, методов и алгоритмов информационной поддержки, которые составляют теоретическую основу интеллектуальной системы формирования запросов к ЭА технической документации с применением предметно-ориентированной онтологии с учетом принципиальной неполноты средств языка запросов.

### **Цель диссертационной работы**

Целью диссертации является совершенствование процессов взаимодействия субъекта проектирования автоматизированных систем с электронным архивом технической документации, обеспечивающее повышение качества выполнения проектных запросов за счет включения в процесс информационной поддержки дополнительных предметных знаний и учета опыта проектировщика.

### **Предмет исследования**

Модели, методы и средства информационной поддержки процесса проектирования АС при анализе и поиске технической документации в электронном архиве.

### **Объект исследования**

Объектом исследования является электронный архив технической документации крупной проектной организации.

### **Задачи диссертационного исследования**

В соответствии с целью работы актуальными являются следующие задачи исследования:

- Провести сравнительный анализ существующих современных методов, алгоритмов и систем обеспечения доступа к массиву текстовой документации.

Рассмотреть их ограничение в контексте проектирования АС. Исследовать возможность и оценить целесообразность применения прикладной онтологии в задачах информационной поддержки проектирования АС.

- Разработать структурно-функциональную модель онтологии информационной поддержки автоматизированного проектирования АС, способствующую интеграции знаний из внешних профессиональных ресурсов.

- Разработать онтологическую модель профиля проектировщика для обеспечения возможности сохранения опыта взаимодействия субъекта проектирования с электронным архивом.

- Разработать алгоритм формирования контекстно-ориентированных запросов к электронному архиву технических документов с целью улучшения качества информационной поддержки проектирования АС.

- Разработать необходимые программные средства, позволяющие решать задачу информационной поддержки в процессе проектирования АС на основе онтологии, провести вычислительные эксперименты, доказывающие их эффективность, внедрить полученные результаты в практику проектной организации.

### **Методы исследования**

В диссертационной работе применяются методы онтологического анализа, теории графов, теории нечетких систем, теории вероятностей и объектно-ориентированного программирования.

### **Научная новизна**

Научная новизна результатов исследования заключается в следующем:

1. Предложена новая структурно-функциональная модель онтологии информационной поддержки автоматизированного проектирования АС, отличающаяся многоуровневой структурой и позволяющая выполнять проектные запросы в контексте этапов жизненного цикла проектируемого изделия и учитывать текущий контекст проекта.

2. Разработана онтологическая модель профиля проектировщика, которая позволяет специфицировать опыт взаимодействия субъекта проектирования с электронным архивом на концептуальном уровне.

3. Разработан алгоритм формирования контекстно-ориентированных запросов к электронному архиву технических документов на основе байесовского классификатора с целью информационной поддержки деятельности проектировщика с учетом моделируемых информационных потребностей.

4. Предложена методика использования онтологических алгоритмов информационной поддержки в жизненном цикле проектирования АС, отличающаяся возможностью интеграции системы профилей субъектов проектирования, онтологического ресурса с электронным архивом технической документации с привлечением дополнительных знаний, извлеченных из внешних профессиональных wiki-ресурсов.

#### **Практическая значимость работы**

Созданная программная система информационной поддержки проектировщика применяется в процессе проектирования автоматизированных систем и позволяет достичь улучшенных технико-экономических показателей объектов проектирования за счет сокращения времени выполнения опытно-конструкторских работ.

Разработанные модели и алгоритмы реализованы в форме программной системы и внедрены в деятельность ФНПЦ АО «НПО «Марс» (г. Ульяновск). Практическое использование результатов диссертационной работы подтверждено соответствующими документами.

#### **Основания для выполнения работы**

Данная научная работа выполнялась в рамках тематического плана научных исследований Федерального агентства по образованию в 2010 году, была поддержана грантами РФФИ № 10-07-00064-а в 2010, 2011 и 2012 годах, № 14-01-31086 мол\_а в 2014 и 2015 годах.

## **Достоверность результатов диссертационной работы**

Достоверность научных положений, выводов и рекомендаций подтверждена результатами математического моделирования, результатами экспериментов и испытаний, а также результатами использования материалов диссертации в проектных подразделениях организации.

### **Основные положения, выносимые на защиту:**

1. Модель прикладной онтологии информационной поддержки проектирования АС является адекватной для выполнения проектных запросов в контексте этапов жизненного цикла проектируемого изделия и учитывает текущий контекст проекта.

2. Онтологическая модель профиля проектировщика является достаточной для адекватного представления опыта взаимодействия субъекта проектирования с электронным архивом на концептуальном уровне.

3. Алгоритм формирования контекстно-ориентированных запросов к электронному архиву технических документов с целью информационной поддержки проектировщика является эффективным по критериям точности и полноты.

4. Разработанный комплекс программ как подсистема информационной поддержки пользователя электронного архива технической документации в полной мере реализует все описанные теоретические положения и позволяет сократить время выполнения поисковых запросов к электронному архиву технических документов.

### **Апробация работы**

Основные положения и результаты диссертации докладывались, обсуждались и получили одобрение на всероссийской школе-семинаре «ИМАП-2011» (г. Ульяновск, 2011 г.); молодежной научно-технической конференции «Автоматизация процессов управления» (г. Ульяновск, 2011 г.); 46-й научно-технической конференции УлГТУ (г. Ульяновск, 2012 г.); 4-й всероссийской научно-технической конференции аспирантов, студентов и молодых ученых «ИВТ-2012» (г. Ульяновск, 2012 г.); 2-м международном симпозиуме «Гибридные

и синергетические интеллектуальные системы: теория и практика» (г. Калининград, 2012 г.); 13-й национальной конференции по искусственному интеллекту с международным участием «КИИ-2012» (г. Белгород, 2012 г.); всероссийской школе-семинаре «ИМАП-2012» (г. Ульяновск, 2012 г.); 47-й научно-технической конференции УлГТУ (г. Ульяновск, 2013 г.); III-й международной научно-технической конференции «OSTIS-2013» (г. Минск, 2013 г.); VII-й международной научно-практической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте» (г. Коломна, 2013 г.); IV-й международной научно-технической конференции «OSTIS-2014» (г. Минск, 2014 г.); VI-й всероссийской научно-практической конференции «НСМВ-2014» (г. Санкт-Петербург, 2014); 3-м международном симпозиуме «Гибридные и синергетические интеллектуальные системы: теория и практика» (г. Калининград, 2014 г.); 13-й национальной конференции по искусственному интеллекту с международным участием «КИИ-2014» (г. Казань, 2014 г.); V-й международной научно-технической конференции «OSTIS-2015» (г. Минск, 2015 г.); VIII-й международной научно-практической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте» (г. Коломна, 2015 г.).

### **Научные публикации**

По результатам работы были опубликованы 21 статья, из которых 6 – в журналах из перечня ВАК, и 3 тезиса докладов. Получено свидетельство (РОСПАТЕНТ) о государственной регистрации программ для ЭВМ №2012617587.

### **Структура и объем диссертации**

Диссертация состоит из введения, четырех глав, заключения, списка использованной литературы и приложений. Основное содержание работ изложено на 152 страницах, включая 45 рисунков и 8 таблиц. Список использованных источников состоит из 121 наименований.

### **Личный вклад**

Все результаты, составляющие содержание диссертации, получены автором самостоятельно.

## Краткое содержание работы

Во введении рассмотрена актуальность работы, определена ее цель и задачи, сформулированы положения, выносимые на защиту, их научная новизна и практическая ценность. Представлены основания для выполнения работы, ее апробация и структура.

В **первой главе** содержится анализ понятия информационной потребности пользователя электронного архива, представлены способы формализации информационной потребности в интеллектуальных системах. Рассматриваются методы информационной поддержки пользователя для удовлетворения его информационных потребностей в процессе принятия проектных решений. Приведен обзор методов информационного поиска и классификации технических документов электронного архива. Рассмотрено понятие «онтология», представлены формальные модели онтологии и способы применения ее в задачах информационного поиска и классификации документов. Анализируются существующие программные системы поддержки проектных решений в электронных архивах технических документов.

Во **второй главе** представлена формальная модель информационной поддержки, описываются онтологические модели и методы информационной поддержки проектировщика на основе алгоритмов формирования контекстно-ориентированных запросов к электронному архиву технической документации.

Система контекстов, в которых происходит проектирование современной АС, включает в себя: *контекст процесса проектирования*, который является внешним контекстом относительно проектной организации; *контекст проектной организации*, который определяется системой понятий и терминов, используемых внутри организации в процессе проектной деятельности и *контекст проекта*, который образуют термины документа, на основе которого проектируется АС (техническое задание и (или) технико-экономическое обоснование).

При формировании профиля проектировщика указанные контексты используются для определения понятий предметной области, которые явно соответствуют или явно не соответствуют текущей информационной потребности.

Формально онтология информационной поддержки записывается в виде кортежа:

$$O = \langle O^{ML}, O^{Dm}, G^D, \{G^P\}, R, F \rangle,$$

где  $O^{ML}$  – онтология методологии проектирования;  $O^{Dm}$  – онтология предметной области проектирования АС;  $G^D$  – концептуальная сеть организации;  $\{G^P\}$  – множество концептуальных сетей реализуемых проектов,  $R$  – множество отношений между фрагментами онтологий (бинарное отношение, имеющее семантику «connected\_to» и связывающее стадию (этап или проектную процедуру) процесса проектирования с подмножеством понятий прикладной онтологии; бинарное отношение, имеющее семантику «connected\_to» и связывающее понятие прикладной онтологии с понятием концептуальной сети организации; бинарное симметричное отношение, имеющее семантику «same\_as», связывающее понятие концептуальной сети проектной организации с понятием из концептуальной сети проектов),  $F$  – множество интерпретирующих функций, представленное в виде:

$$F = \{F_{StC}, F_{DC}, F_{DP}\},$$

где  $F_{StC} : St \rightarrow \{C\}$  – функция, сопоставляющая стадии проектирования (этапу, проектной процедуре) подмножество понятий предметной области, задаваемая алгоритмически;  $F_{DC} : \{T^D\} \rightarrow C$  – функция интерпретации терминов, сопоставляющая каждому концептуальному набору терминов из словаря (формирует терминологическое окружение понятия);  $F_{DP} : T^D \times T^P \rightarrow boolean$  – функция определения идентичных понятий в концептуальной сети проектной организации и в концептуальных сетях проектов, возвращающая значения: *истина* или *ложь*.

Множество стадий проектирования АС  $St$  и множество возможных классов артефактов (документов)  $Tr$  представляются в виде четкого соответствия

$\Gamma_{St/Tp} = (St, Tp, F_{St/Tp})$ , которое задается с помощью матрицы инцидентий  $R_{\Gamma_{St/Tp}}$ , строки которой помечены элементами  $St_i \in St$ ,  $i \in I = \{1, 2, \dots, n\}$ , столбцы – элементами  $tp_j \in Tp$ ,  $j \in J = \{1, 2, \dots, m\}$ , а на пересечении  $St_i$  строки и  $tp_j$  столбца ставится элемент  $r_{ij} = \gamma_{F_{St/Tp}} \langle St_i, tp_j \rangle$ , где  $\gamma_{F_{St/Tp}} = \{0, 1\}$  – характеристическая функция, определяющая, существует ли отношение  $St_i | tp_j$  ( $\gamma_{F_{St/Tp}} = 1$ ) или такого отношения нет ( $\gamma_{F_{St/Tp}} = 0$ ).

Под *текущим состоянием проекта*  $P$  понимается множество типов ТД  $Sm_t^p = \{tp\}$ , где  $t$  – текущий момент времени,  $tp \in \{tp_1, tp_2, \dots, tp_\lambda\}$  – тип ТД в составе проекта из перечня классов документальных артефактов длиной  $\lambda$ .

Каждой информационной потребности  $In_j^i$  ставится в соответствие пара классов понятий онтологии предметной области  $C^+ = \{c_1^+, c_2^+, \dots, c_n^+\}$ ,  $C^- = \{c_1^-, c_2^-, \dots, c_m^-\}$ , определяющие положительные и отрицательные подмножества понятий онтологии, соответственно.

В процессе выполнения проектировщиком информационных запросов к электронному архиву определяется набор ТД, которые соответствуют его информационной потребности ( $D^+$ ) и ТД, не соответствующие ей ( $D^-$ ), с учетом текущей стадии (этапа) проектирования. Для каждого документа определяется его концептуальное представление (производится нечеткое онтологическое индексирование). Определяется нечеткое соответствие между множеством  $C^{+(-)}$  и множеством  $T^D = T_{in}^D \cup T_{ext}^D$  как  $\tilde{\Gamma}_{CT} = (C^{+(-)}, T^D, \tilde{F}_{CT})$ , где  $\tilde{F}_{CT}$  – нечеткое множество в  $C^{+(-)} \times T^D$ . Нечеткое соответствие  $\tilde{\Gamma}_{CT}$  задается в виде ориентированного двудольного графа с множеством вершин  $C^{+(-)} \cup T^D$ , каждой дуге  $\langle c_i^+, t_j \rangle$  ( $\langle c_i^-, t_j \rangle$ ) которого приписываем значение функции принадлежности  $\mu_{F_{CT}} \langle c_i^+, t_j \rangle$  ( $\mu_{F_{CT}} \langle c_i^-, t_j \rangle$ ).

Указанное значение функции принадлежности вычисляется на основе нормализованной частоты встречаемости термина в терминологическом

окружении понятия. Поскольку терминологические окружения понятий формируются из внутреннего источника (терминологического словаря)  $T_{in}^D$  и из внешнего источника (профессиональных wiki-ресурсов)  $T_{ext}^D$ , а также  $T_{in}^D \cap T_{ext}^D \neq \emptyset$ , для определения функции принадлежности применяется следующее правило:

$$\begin{aligned} \mu_{F_{CT}} \langle c_i^+, t_j \rangle &= \mu_{F_{CT}} \langle c_i^+, t_{in(j)} \rangle \in T_{in}^D > \\ \text{if } \mu_{F_{CT}} \langle c_i^+, t_{in(j)} \rangle \in T_{in}^D &= \max(\mu_{F_{CT}} \langle c_i^+, t_{in(j)} \rangle \in T_{in}^D >, \mu_{F_{CT}} \langle c_i^+, t_{ext(j)} \rangle \in T_{ext}^D >) \\ \text{else } \mu_{F_{CT}} \langle c_i^+, t_j \rangle &= \mu_{F_{CT}} \langle c_i^+, t_{ext(j)} \rangle \in T_{ext}^D >. \end{aligned}$$

Образ множества  $T^D$  при соответствии  $\tilde{\Gamma}_{CT}$  фактически представляет собой нечеткое множество, элементами которого являются концепты с соответствующими степенями выраженности:

$$\tilde{\Gamma}_{CT}(T^D) = \{ \mu_{\Gamma_{CT}}(c^{+(-)}) / c^{+(-)} \},$$

где  $\mu_{\Gamma_{CT}}(c^{+(-)}) = \bigvee_{t \in T^D} \mu_{F_{CT}} \langle c_i^{+(-)}, t_j \rangle$ .

В положительные и отрицательные подмножества понятий онтологии предметной области включаются такие понятия из онтологических представлений технических документов, степень выраженности которых наибольшая.

Формально профиль проектировщика представляется в виде кортежа:

$$Ex_j^i = \langle In_j^i, C^+, C^- \rangle, c^{+(-)} = \mu_{\Gamma_{CT}}(c^{+(-)}) = \max_{c \in D} \left( \bigvee_{t \in T^D} \mu_{F_{CT}} \langle c_i^{+(-)}, t_j \rangle \right), c^{+(-)} \in C^{+(-)},$$

где  $i$  – индекс проектировщика,  $j$  – индекс стадии жизненного цикла проектирования АС.

Контекст проекта записывается как граф вида:

$$G^{PT} = \langle C^{PT}, R^{PT} \rangle,$$

где  $C^{PT}$  – множество вершин-понятий проекта,  $R^{PT}$  – множество дуг, соединяющих вершины-понятия.

Множество понятий проекта определяется как результат функции онтологического индексирования технического задания ( $Tz$ ) на реализуемый проект ( $F_{ol}(Tz)$ ) и функции онтологического доопределения множества  $C^{Tz}$  как результата  $F_{ol}(Tz)$  с применением профессиональных wiki-ресурсов ( $F_{cAdd}(C^{Tz})$ ).

На рис. 1. представлена структурно-функциональная схема уточнения проектных запросов с использованием системы онтологий и профилей проектировщиков.

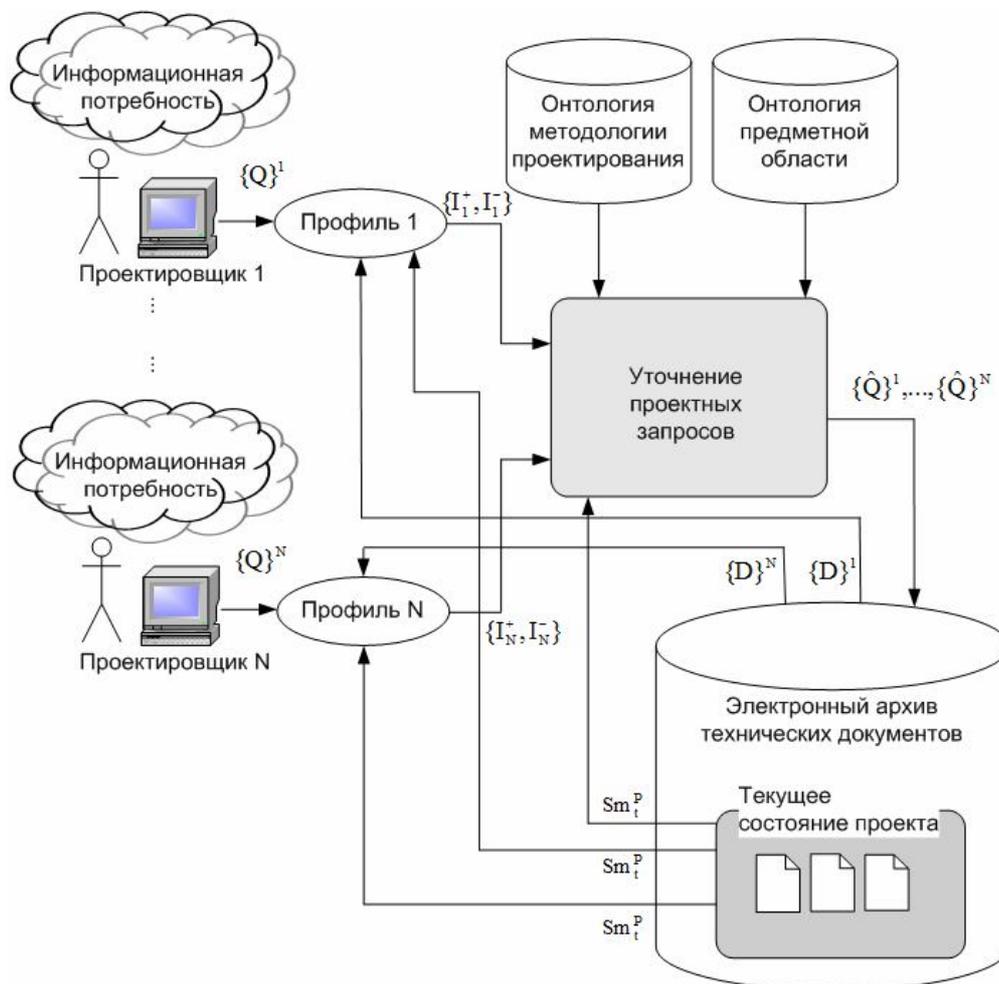


Рис. 1. Структурно-функциональная схема уточнения проектных запросов

Рассмотрим алгоритм формирования проектных запросов, который использует профиль проектировщика для улучшения показателей точности и полноты запросов к электронным архивам технических документов проектной организации.

Пусть множество  $\hat{w} = \{\hat{w}_1, \hat{w}_2, \dots, \hat{w}_n\}$  – есть множество ключевых слов проектного запроса к электронному архиву. Используя функцию  $F_{DC} : \{T^D\} \rightarrow C$ , данное множество отображается в нечеткое множество:

$$\tilde{I}_q = \{\mu_1 / c_1, \mu_2 / c_2, \dots, \mu_m / c_m\}, \quad (1)$$

как результат онтологического преобразования терминов в набор степеней выраженности понятий онтологии.

Для представления модели формирования проектных запросов на концептуальном уровне применяется наивный байесовский классификатор. Вероятность того, что понятие запроса  $c$  принадлежит классу  $k \in \{C^+, C^-\}$ , определяется по формуле Байеса:

$$P(k|c) = \frac{P(c|k) \cdot P(k)}{P(c)},$$

где  $P(c|k)$  – вероятность встретить понятие  $c$  среди всех понятий класса  $k$ ;  $P(k)$  – безусловная вероятность понятия класса  $k$  в онтологии предметной области;  $P(c)$  – безусловная вероятность понятия  $c$  в онтологии предметной области.

Наиболее вероятный класс для понятия запроса определяется, используя оценку апостериорного максимума:

$$k_{map} = \arg \max_{k \in K} [\log P(k) \cdot \sum_{i=1}^n \log P(w_i | k)].$$

Оценка вероятностей  $P(k)$  и  $P(w_i | k)$  выполняется на основе обучающей выборки, сформированной для каждой информационной потребности  $In_j^i$ .

Вероятность класса записывается как:

$$P(k) = \frac{D_k}{D},$$

где  $D_k$  – количество документов, принадлежащих классу  $k$  и определяемое на основе результатов выполнения запросов проектировщика к электронному архиву;  $D$  – общее количество документов в обучающей выборке.

Величина  $P(w_i | k)$  определяет вероятность встретить термин  $w_i$  среди терминов документов, принадлежащих классу  $k$ . Значение данной величины определяется с учетом того, что термин из окружения понятия, включенного в проектный запрос, может отсутствовать в документах анализируемого класса:

$$P(w_i | k) = \frac{f_{ik} + I}{\sum_{i' \in V} (f_{ik'} + I)},$$

где  $f_{ik}$  – частота встречаемости  $i$ -го термина в документах класса  $k$ ;  $V$  – терминологический словарь проектной организации (список всех уникальных терминов).

В результате классификации понятий запроса (1) понятия, принадлежащие к положительному классу понятий информационной потребности проектировщика, остаются в составе запроса. Понятия, которые классифицированы как отрицательные, исключаются из исходного запроса.

Поисковый запрос, сформулированный в терминах на естественном языке, преобразуется к онтологическому виду (1), но прежде выполняется ряд предобработок:

- удаление стоп-слов;
- стемминг (выделение основы слова, получение термов).

С полученным запросом  $Q = \{\hat{w}\}$  возможны два варианта обработки:

1)  $Q = C$  – поисковый запрос полностью совпадает с названием концепта предметной онтологии.

2)  $Q = W$  – поисковый запрос совпадает с терминами предметной онтологии.

В первом случае проектный запрос расширяется, используя функцию интерпретации концептов, т.е. запрос дополняется терминами из терминологического окружения концепта:

$$Q = Q \cup F_{DC}(c_i).$$

К полученному множеству концептов применяется функция интерпретации концептов, дополняя запрос терминами, семантически связанные с концептами:

$$Q = Q \cup F_{DC}(F_{DC}^{-1}(\hat{w})).$$

Сформированный запрос представляется в виде графа:

$$G^q = \langle C^q, R^q \rangle, \quad (2)$$

где  $C^q$  – множество вершин-понятий проекта,  $R^q$  – множество дуг, соединяющих вершины-понятия. Понятия в контекстно-ориентированном проектном запросе соединены двумя видами отношения: часть-целое и род-вид (рис. 2, а).

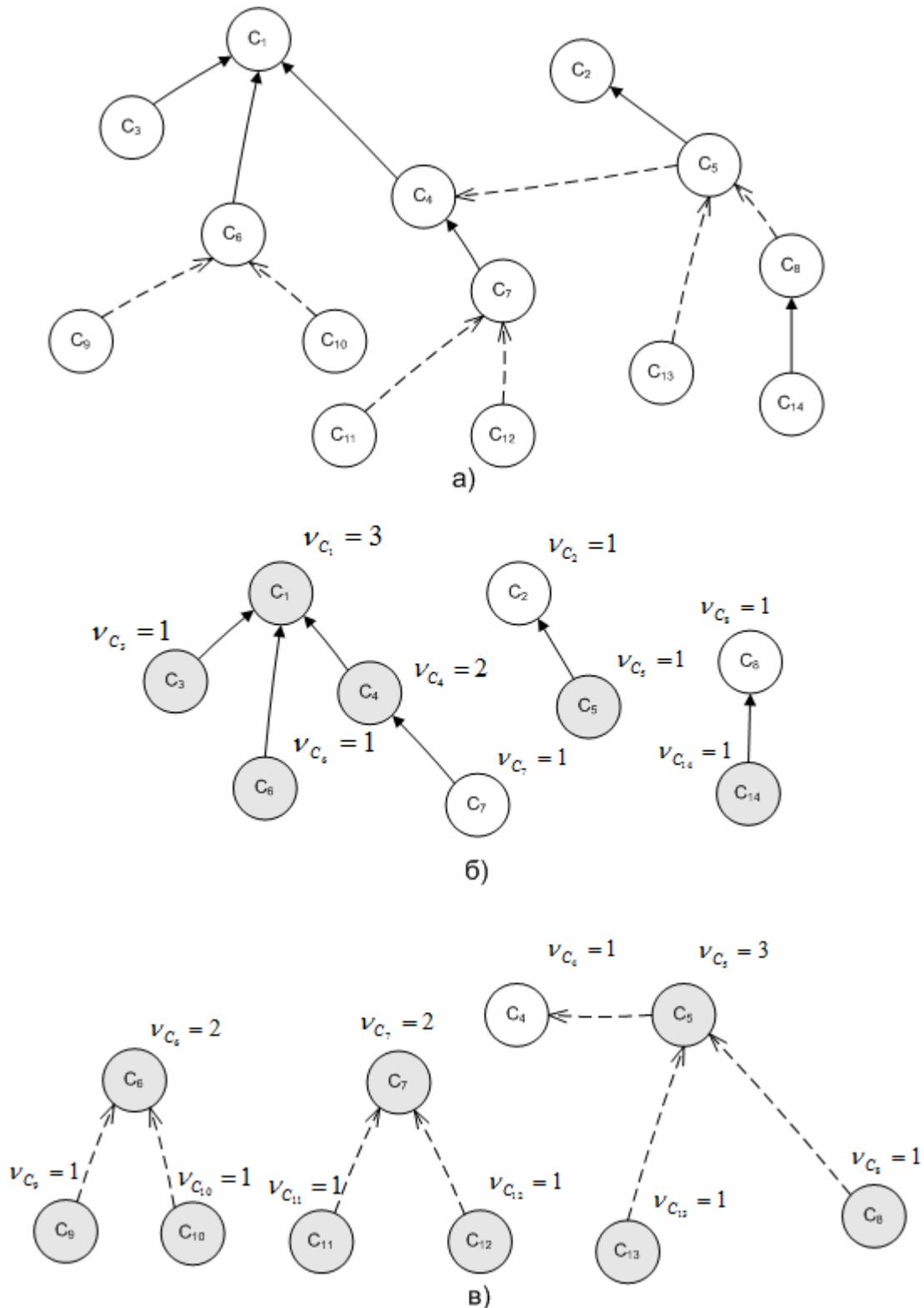


Рис. 2. Схема редуцирования понятий контекстно-ориентированных запросов

Способ редуцирования понятий проектного запроса основывается на разделении исходного графа (2) на несколько подграфов, каждый из которых содержит только вершины, соединенные дугами одной семантической категории (рис. 2, б, в). Редуцирование понятий графа проектного запроса предполагает выполнения ряда следующих шагов.

*Шаг 1.* Разбиение графа проектного запроса на несколько подграфов с учетом семантических категорий («isA» или «part\_of»).

*Шаг 2.* Для каждой вершины подграфов определяется ее степень  $v_{c_j}$  (количество входящих и исходящих дуг).

*Шаг 3.* В редуцированное множество понятий отдельно взятого подграфа проектного запроса включаются понятия согласно заранее определенным правилам.

*Шаг 4.* В редуцированное множество понятий проектного запроса включаются пересечение множеств понятий.

Для нахождения меры сходства между нечетким контекстно-ориентированным запросом и документом используется следующее выражение:

$$similarity = \cos(\theta) = \frac{\sum_{i=1}^n \mu(c_i)_{I_q} \times \mu(c_i)_{I_d}}{\sqrt{\sum_{i=1}^n \mu(c_i)_{I_q} \times \sum_{i=1}^n \mu(c_i)_{I_d}}},$$

где  $c_i$  – понятия из словаря концептов  $S$  предметной области предметно-ориентированной онтологии;  $\mu_{I_q}(c_i), \mu_{I_d}(c_i)$  – степень выраженности концепта в поисковом запросе и документе, соответственно. Алгоритм формирования нечетких контекстно-ориентированных запросов к электронному архиву представлен на рисунке 3.

Основные процедуры методики интеллектуальной информационной поддержки проектирования АС включают в себя: формирование онтологии, разработку индивидуальных профилей проектировщиков и интеграцию разработанных онтологических ресурсов в систему электронного архива проектной организации.

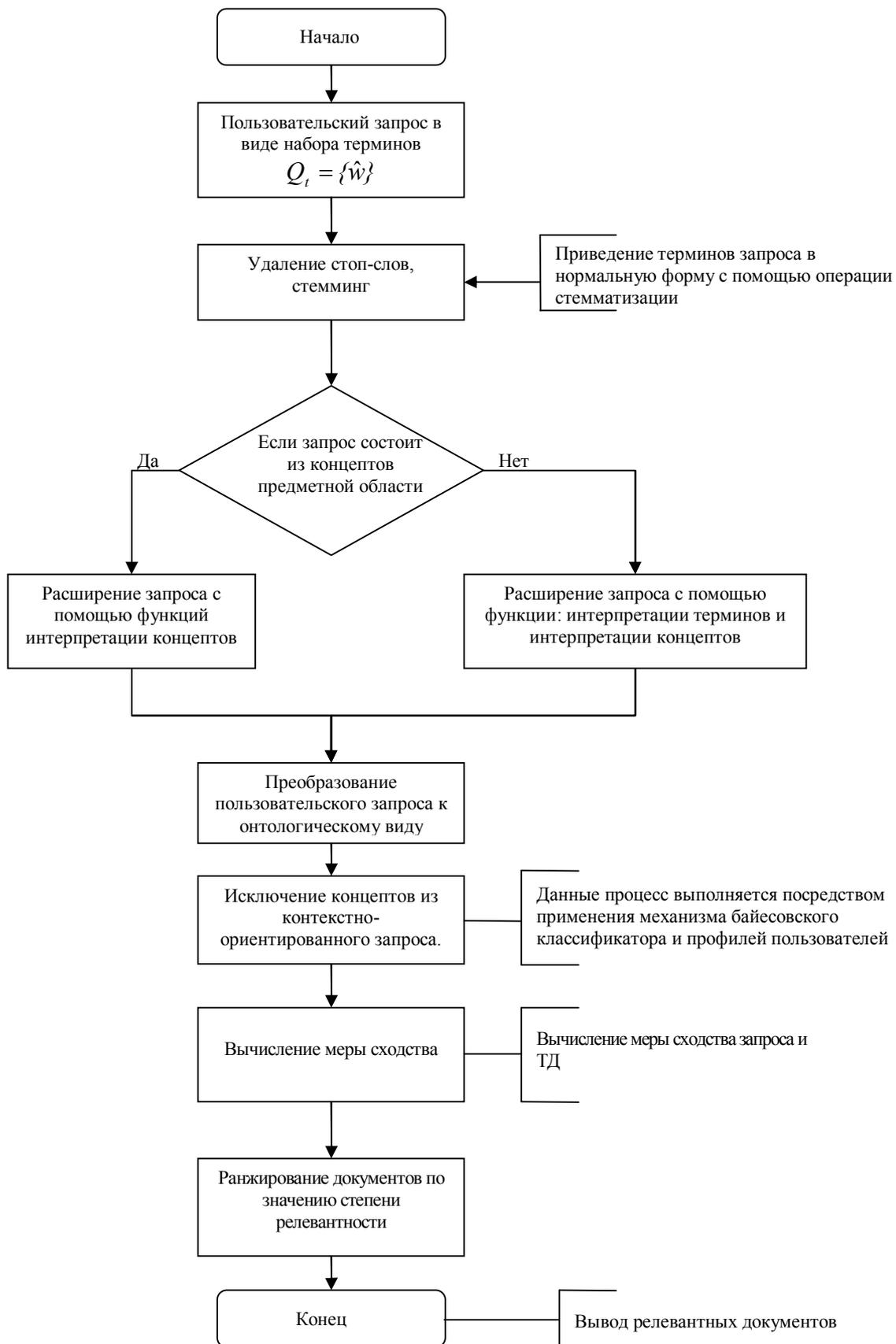


Рис. 3. Алгоритм формирования нечетких контекстно-ориентированных запросов

В третьей главе представлено описание разработанной программной системы информационной поддержки автоматизированного проектирования на основе онтологии.

Интеллектуальная система информационной поддержки проектирования АС – это программная система, предназначенная для систематизации и автоматизации работы с электронным архивом ТД, учитывающая состояние онтологии предметной области.



Рис. 4. Структура системы информационной поддержки

Основными функциями данной системы являются:

- онтологически-ориентированное индексирование ТД. Формирование онтологических представлений ТД;
- хранение онтологии предметной области и онтологических представлений ТД;

- формирование и хранение профилей пользователей, учитывающие информационную потребность субъекта проектирования на разных стадиях проектирования АС;
- формирование контекстно-ориентированных проектных запросов к электронному архиву;
- извлечение ТД из электронного архива с применением онтологических профилей пользователей.

Разработанная система включена в состав электронного архива ФНПЦ АО «НПО «Марс». В качестве хранилища предметно-ориентированной онтологии и профилей проектировщиков выступает сервер онтологий Sesame. Программная система реализована на языке программирования Java в среде разработки Eclipse.

Система информационной поддержки имеет структуру, представленную на рисунке 4. Результат работы подсистемы предметно-ориентированного редактора онтологии сохраняется в формате RDF. На рисунке 5 представлена графическая модель фрагмента RDF-схемы.

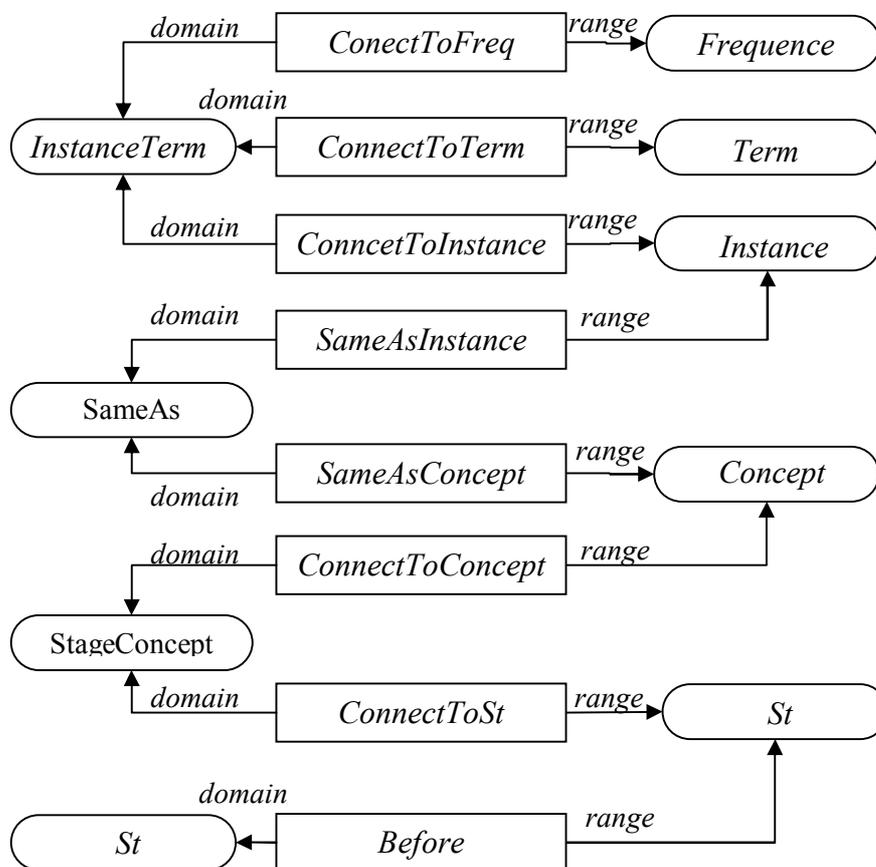


Рис. 5. Графическое представление фрагмента RDF-схемы

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

Вычислительные эксперименты проводились в центре обработки данных ФНПЦ АО «НПО «Марс». Данный вычислительный ресурс под управлением операционной системы Red Hat Enterprise Linux с виртуализацией Red Hat Enterprise Virtualization обладает следующими характеристиками: 8-ми ядерный процессор Intel Xeon MP E7-2830 Westmere-EX, размер оперативной памяти – 256 Гб. План вычислительного эксперимента включал следующие этапы:

1. Построение концептуальной сети проектов с использованием концептов предметной онтологии электронного архива и wiki-ресурса.

2. Создание профилей пользователей системы информационной поддержки электронного архива.

3. С применением профилей, созданных на втором этапе, выполнялись запросы к электронному архиву. Сформирован набор пользовательских запросов, состоящих из терминов, относящихся к предметной области разработки АС.

4. Выполнение экспериментов, отличительная особенность которых заключалась в том, что не использовались пользовательские профили в процессе формирования контекстно-ориентированных запросов.

5. Выполнение вычислительных экспериментов с онтологией, разработанной экспертом предметной области. Полученные результаты сравнивались с результатами экспериментов с онтологией, которая содержит концептуальную сеть проектов, полученную автоматически из внешних профессиональных ресурсов.

6. Данный этап представляет собой выполнение вычислительных экспериментов с 1-го по 4-ый этап с рабочим проектом, выполняемый коллективом проектировщиков ФНПЦ АО «НПО «Марс».

Для сравнительного анализа результатов применялись следующие программные продукты: Яндекс. Персональный поиск (ЯПП), 1С:Документооборот 8 (1С), Архивариус 3000 (А300), Copernic Desktop Search (CDS).

В ходе эксперимента была автоматически построена концептуальная сеть проекта, которая включает в себя 205 концептов и около 22 000 уникальных терминов. Кроме того, была разработана онтология предметной области, включающая 81 понятие и 10078 уникальных терминов, построена онтология методологии проектирования, которая содержит около 50 концептов.

Сформированы пользовательские профили электронного архива проектировщика, программиста и инженера. Экспертом ФНПЦ АО «НПО «Марс» была подготовлена выборка, состоящая из 5850 ТД. Были проведены эксперименты и вычислены оценки качества полученных результатов.

Онтологическая модель показала лучшие результаты, чем существующие системы поиска, если запрос состоял из более чем трех терминов (рис. 6).

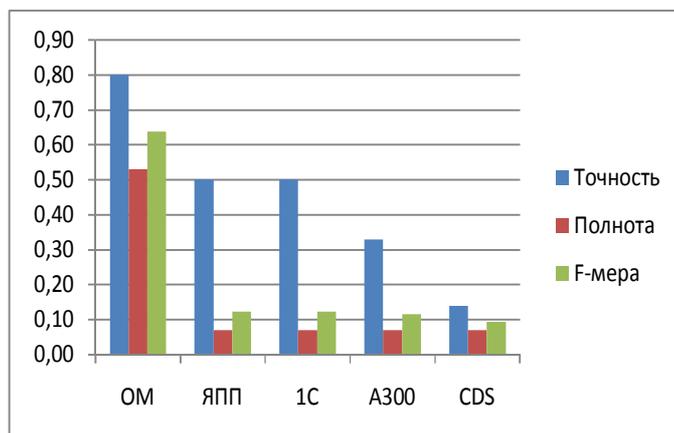


Рис. 6. Сравнение результатов экспериментов с длинными запросами

В тоже время низкие оценки качества онтологическая модель показала, если в процессе поиска использовались короткие запросы (рис. 7).

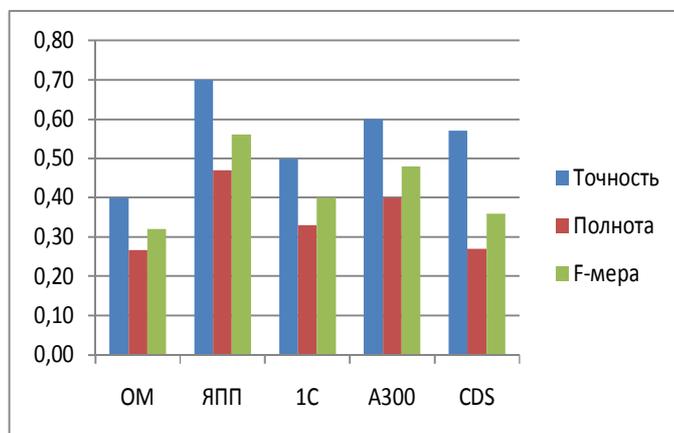


Рис. 7. Сравнение результатов экспериментов с короткими запросами

С запросами, которые явно определяли предметную область, существующие программные системы показали более качественный результат (рис. 8), но при появлении смысловой размытости запроса наблюдалось падение качества результата поиска у существующих систем, в то время как с использованием онтологической модели смысловая неопределенность запроса практически не влияла на результат (рис. 9).

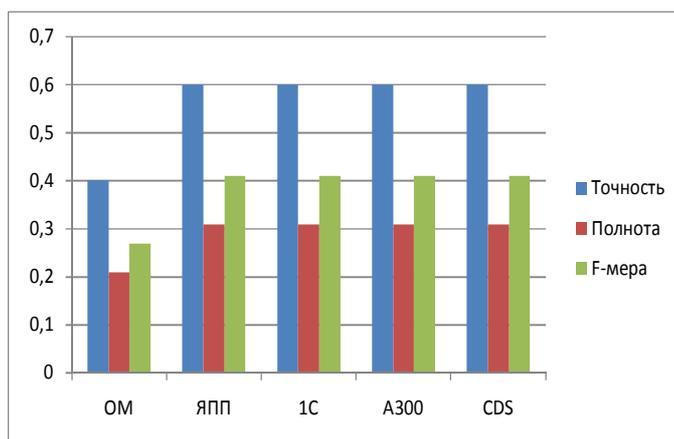


Рис. 8. Сравнение результатов экспериментов с запросами, семантически явно определяющими предметную область

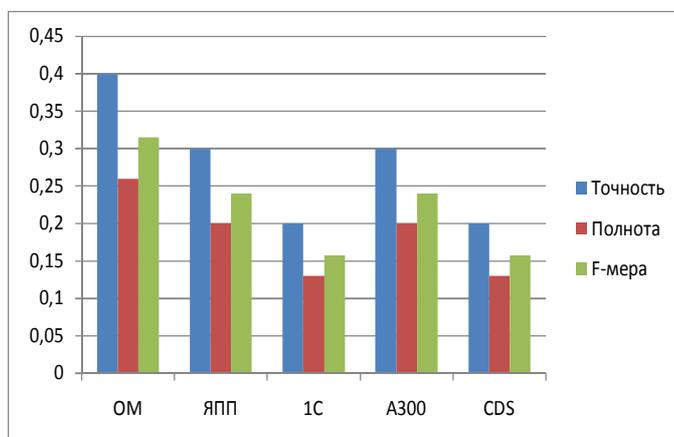


Рис. 9. Сравнение результатов экспериментов с запросами, семантически не явно определяющими предметную область

Онтологическая модель, использующая профили пользователей показывает более качественный результат, чем поиск, который такие профили не учитывает (рис. 10).

Результаты вычислительных экспериментов показывают, что формирование концептуальной сети проекта автоматизированным способом и применение ее в

процессах информационной поддержки субъекта проектирования позволяют улучшить качество поиска по сравнению с экспертной онтологией (рис. 11).

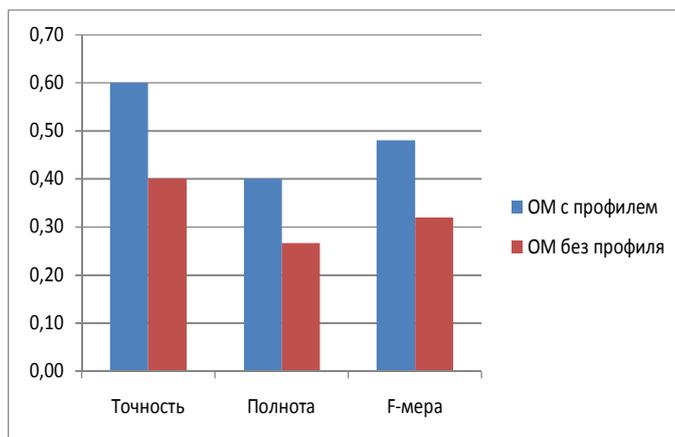


Рис. 10. Сравнение результатов экспериментов онтологических моделей поиска

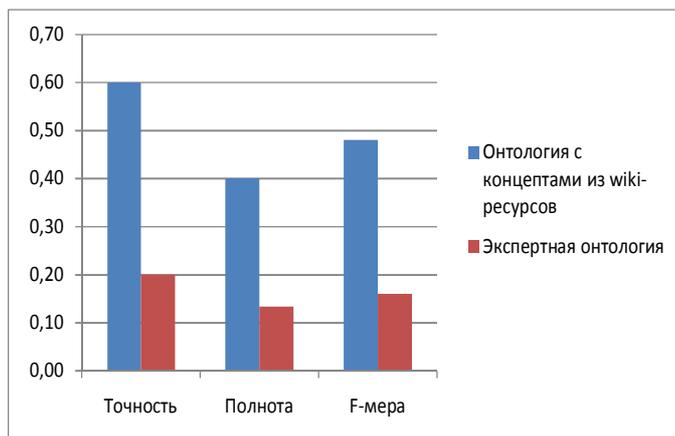


Рис. 11. Сравнение результатов экспериментов онтологических моделей поиска, использующие различные онтологии

В **заключении** приведены основные результаты исследований, полученные в диссертационной работе:

1. Предложена новая структурно-функциональная модель онтологии информационной поддержки проектирования АС, отличающаяся многоуровневой структурой, учитывающая этапы жизненного цикла проектируемого изделия и текущий контекст проекта.

2. Предложена онтологическая модель профиля проектировщика, учитывающая опыт взаимодействия субъекта проектирования с электронным архивом на концептуальном уровне.

3. Предложен алгоритм формирования контекстно-ориентированных проектных запросов к электронному архиву технических документов с целью повышения качества информационной поддержки проектировщика.

4. Разработан комплекс программ как подсистема информационной поддержки пользователя электронного архива технической документации, которая в полной мере реализует все описанные теоретические положения и позволяет сократить время поиска технических документов примерно на 40-50% с увеличением точности поиска на 30%.

5. Проведены вычислительные эксперименты, основными результатами которых являются:

- Онтологическая модель поиска, учитывающая информационную потребность проектировщика, является эффективным средством при решении задачи информационной поддержки проектирования АС. Онтологическая модель по сравнению с традиционными моделями формирования контекстно-ориентированных запросов показала более качественный результат с запросами, которые семантически нечетко выражают предметную область и с запросами, которые содержат большое количество терминов.

- Автоматизированное построение концептуальной сети с использованием внешних профессиональных wiki-ресурсов позволяет уменьшить время создания онтологии и улучшить качество поиска технических документов в ЭА проектной организации на 30%.

## **Список публикаций**

### **Публикации в изданиях, рекомендованных ВАК:**

1. Субхангулов Р.А. Разработка инструментария для интеллектуального анализа технической документации / А.М. Наместников, Р.А. Субхангулов, А.А. Филиппов // Известия Самарского научного центра Российской академии наук, 2011. – № 4(13). – С. 984–990.

2. Субхангулов Р.А. Разработка инструмента инженерии онтологии в

- интеллектуальном проектном репозитории / А.М. Наместников, **Р.А. Субхангулов** // Автоматизация процессов управления – 2012. – №2. – С.38–43.
3. Субхангулов Р.А. Онтологически-ориентированный метод поиска проектных документов / **Р.А. Субхангулов** // Автоматизация процессов управления – 2012. – №4. – С.83 – 89.
4. Субхангулов Р.А. Формирование информационных запросов к электронному архиву на основе концептуального индекса / А.М. Наместников, **Р.А. Субхангулов** // Радиотехника – 2014. – №7. – С. 126–129.
5. Субхангулов Р.А. Онтологический поиск технических документов на основе модели интеллектуального агента / **Р.А. Субхангулов** // Автоматизация процессов управления – 2014. – №4. – С. 85–91.
6. Субхангулов Р.А. Онтологическая модель контекстного поиска электронных документов в архиве проектной организации / А.М. Наместников, **Р.А. Субхангулов** // Радиотехника – 2015. – №6. – С. 73–78.

#### **Публикации в прочих изданиях:**

1. Субхангулов Р.А. Предметно-ориентированный инструментарий разработки онтологии / А.М. Наместников, **Р.А. Субхангулов** // Гибридные и синергетические интеллектуальные системы: теория и практика : материалы 1-го международного симпозиума. – Калининград : БФУ им. И. Канта, 2012. – С. 197–204.
2. Субхангулов Р.А. Система кластеризации и полнотекстового поиска проектных документов на основе прикладной онтологии / А.М. Наместников, **Р.А. Субхангулов**, А.А. Филиппов // Сборник научных трудов тринадцатой национальной конф. по искусственному интеллекту с международным участием. – Белгород: БГТУ, 2012. – Т.2. – С. 104–111.
3. Субхангулов Р.А. Онтологически- ориентированная система кластеризации и полнотекстового поиска проектных документов / А.М. Наместников, **Р.А. Субхангулов**, А.А. Филиппов // Открытые семантические технологии проектирования интеллектуальных систем: материалы III Междунар. научн. техн.

конф. – Минск: БГУИР, 2013. – С. 219–224.

4. Субхангулов Р.А. Применение нечетких моделей в задачах кластеризации и информационного поиска текстовых проектных документов / А.М. Наместников, **Р.А. Субхангулов**, А.А. Филиппов // Интегрированные модели и мягкие вычисления в искусственном интеллекте: сборник научных трудов VII-й Международной научно-практической конференции. – М.: Физматлит, 2013. – Т.3. – С. 1278–1289.

5. Субхангулов Р.А. Оптимизация поисковых запросов на основе байесовских классификаторов / А.М. Наместников, **Р.А. Субхангулов** // Гибридные и синергетические интеллектуальные системы: теория и практика: материалы 2-го международного симпозиума. – Калининград: БФУ, 2014. – С.246–251.

6. Субхангулов Р.А. Онтологически-ориентированная система интеллектуального анализа данных технических документов / А.А. Филиппов, **Р.А. Субхангулов** // Сборник научных трудов четырнадцатой национальной конф. по искусственному интеллекту с международным участием. – Казань: РИЦ «Школа», 2014 – Т.3. – С. 228–236.

### **Свидетельства о государственной регистрации программы для ЭВМ:**

1. Субхангулов Р.А. Предметно-ориентированный редактор онтологий / А.М. Наместников, **Р.А. Субхангулов** // Свидетельство о государственной регистрации программы для ЭВМ №2012617587.