

УДК 002:004.9

## **ЗАДАЧА ИЗВЛЕЧЕНИЯ ЗНАНИЙ ИЗ ЭЛЕКТРОННЫХ ФОНДОВ НА ОСНОВЕ ЛОГИКО-СЕМАНТИЧЕСКОЙ СЕТИ ВОПРОС–ОТВЕТ–РЕАКЦИЯ**

**Филозова Ирина Анатольевна<sup>1</sup>, Добрынин Владимир Николаевич<sup>2</sup>**

<sup>1</sup>Старший преподаватель, научный сотрудник;  
ГБОУ ВПО «Международный Университет природы, общества и человека «Дубна»,  
Институт системного анализа и управления;  
141980, Московская обл., г. Дубна, ул. Университетская, 19;  
Объединенный институт ядерных исследований,  
Лаборатория информационных технологий;  
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;  
e-mail: fia@jinr.ru.

<sup>2</sup>Кандидат технических наук, профессор Института системного анализа и управления;  
ГБОУ ВПО «Международный Университет природы, общества и человека «Дубна»,  
Институт системного анализа и управления;  
141980, Московская обл., г. Дубна, ул. Университетская, 19;  
e-mail: arbatsolo@yandex.ru.

*В статье излагается подход к структуризации информационных ресурсов электронных информационных фондов на основе логико-семантической сети (ЛСС) «Вопрос-Ответ-Реакция». Обсуждается возможность создания семантической поисковой системы на ее основе, и применимости такой системы для эффективного изучения информационных ресурсов научно-технического и образовательного содержания. В статье обсуждается возможность применимости семантической поисковой системы на основе ЛСС.*

Ключевые слова: научный текст, информационный поиск, пертинентность, навигация, электронные библиотеки, вопросно-ответные системы, семантические связи, логико-семантическая сеть «Вопрос-Ответ-Реакция».

## **PROBLEMS OF KNOWLEDGE EXTRACTION FROM DIGITAL FUNDS ON THE BASE OF THE LOGIC-SEMANTIC NETWORK QUESTION–ANSWER-REACTION**

**Filozova Irina<sup>1</sup>, Dobrynin Vladimir<sup>2</sup>**

<sup>1</sup>Senior teacher, scientist;  
Dubna International University of Nature, Society, and Man,  
Institute of system analysis and management;  
141980, Dubna, Moscow reg., Universitetskaya str., 19;  
Join Institute for nuclear Research,  
Laboratory of Information Technologies;  
141980, Dubna, Moscow reg., Joliot-Curie str., 6;  
e-mail: fia@jinr.ru.

<sup>2</sup> Candidate of Science in Engineering, professor of Institute of system analysis and management;  
Dubna International University of Nature, Society, and Man,  
Institute of system analysis and management;  
141980, Dubna, Moscow reg., Universitetskaya str., 19;  
e-mail: arbatsolo@yandex.ru.

*The paper presents an approach to structuring information resources of digital information funds based on the logical-semantic network (LSN) «Question-Answer-Response». The possibility of creation a semantic search engine based on LSN, and the applicability of such a system for effective learning of information resources of scientific and educational content are discussed.*

**Keywords:** scientific text, information search, pertinence, navigation, digital libraries, question-answer systems, semantic relations, logic-semantic network «Question-Answer-Reaction».

## **Введение**

В научной и образовательной сфере продукцией деятельности являются публикации в научном стиле с установленными семантическими связями на другие публикации. Эти связи имеют различную природу: связи, образованные цитированием; связи научного вывода; связи использования; связи мнений и профессиональных оценок; связи между персоналиями и организациями [1], [2]. В данной работе исследуются связи, отражающие логику изложения авторской мысли в рамках данной публикации, темы, предметной области. Структурирование и поддержка этих связей позволяет обеспечить новые возможности для исследования корпуса документов электронных библиотек (ЭБ), тем самым повышая их качество. Комплексное семантическое структурирование фонда ЭБ приводит к появлению многоуровневой сетевой структуры, которая может быть положена в основу навигационного механизма ЭБ. В данной работе излагается описание технологии структурирования научных текстов на основе логико-семантической сети (ЛСС) «Вопрос-Ответ-Реакция» для извлечения знаний из электронных фондов.

## **1. Проблемная ситуация**

Одна из важных прикладных проблем эффективного поиска информации (в условиях жестких ограничений) – это проблема поиска новых, уточненных, углубленных знаний в точках бифуркации социотехнических систем. Информационные фонды (в том числе и электронные библиотеки), отражающие накопленные знания (теоретические, прикладные, прагматические), являются источниками генерации новых идей и формирования постановки и решения широкого спектра задач. К примеру, этими задачами могут быть: исследование, экспертиза, инженерная задача, конструкторская задача и пр. Извлечение необходимых для решения поставленной задачи знаний (в узком смысле) в некотором массиве информации может быть осуществлен двумя способами: 1) посредством информационного запроса в рамках заданного информационно-поискового языка; 2) посредством вопроса на языке предметной области (усеченный естественный язык). В среде специалистов-профессионалов в некоторой предметной области коммуникация наиболее эффективна, когда она происходит в вопросно-ответной форме. Большинство существующих сегодня информационных поисковых систем (ПС) и семантических ПС (вопросно-ответных) предоставляют возможности исследователю получать множество ответов, среди которых, возможно, есть желаемое. Таким образом, поиск – это итерационная процедура, предполагающая уточнение запроса. Специалист в некоторой предметной области, осуществляющий поиск, имеет определенное представление о том, что из полученных результатов может являться ответом на его вопрос. Однако, большинство классических информационно-поисковых и вопросно-ответных систем являются одноконтурными или двухконтурными (позволяющими осуществить поиск в найденном). Специалисту же удобнее уточнять запрос/вопрос, не выходя из системы. При этом важно, насколько результаты поиска пертинентны (пертинентность – прагматическая метрика информационного поиска), т.е. удовлетворяют информационные потребности пользователя независимо от того, как полно и как точно эта информационная потребность выражена с помощью информационного запроса или вопроса. Следовательно, создание эффективных механизмов поиска ответов на вопросы в электронных информационных фондах, т.е. извлечение знаний (в узком смысле) – актуальная проблема.

## **2. Семантическая поисковая система на основе логико-семантической сети «Вопрос-Ответ-Реакция»**

Работа специалиста-профессионала с фондами предполагает наличие системы каталогизации и классификации материала. В рамках заданной проблемной темы предлагается технология формирования и поддержки «каталожной» службы, которая обеспечивает эффективный поиск ответов на вопросы. Стержнем такой «каталожной» службы является упорядоченное открытое множество логико-семантических сетей (ЛСС) «Вопрос-Ответ-Реакция» [3], [4].

Познавательная функция вопроса связана с восполнением, уточнением и конкретизацией ранее полученных общих представлений о предметах и явлениях действительности. Грамматической формой выражения запроса мысли в языке выступает вопросительное предложение.

**Вопрос** – это выраженный в форме вопросительного предложения и реализуемый в виде ответа запрос мысли, направленный на развитие: уточнение, дополнение или обобщение знаний.

Вопросительными называются предложения, в которых специальными языковыми средствами выражается стремление говорящего узнать что-либо или удостовериться в чем-либо. Вопросительные предложения, таким образом, информируют о том, что хочет узнать говорящий.

В процессе познания вопросы не возникают самостоятельно, без всяких на то оснований. Любой вопрос всегда опирается на уже известное **знание**, которое выступает его базисом и выполняет роль предпосылки вопроса.

Обычно выделяют следующие типы вопросов:

1. *Ли-вопросы (да-нет-вопросы)*, возможным ответом на которые обычно служит слово «да» или «нет». Примеры такого рода вопросов: «Является ли вода жидкостью?», «Существует ли Бог?».
2. *Какой (кто, сколько)* – вопросы, отвечая на которые необходимо указать какой-то конкретный предмет или ряд предметов (одушевленных или неодушевленных) или же запрашиваемую количественную характеристику. Например: «Какие реки впадают в Черное море?», «Сколько притоков у Волги?».
3. *Что такое (почему, зачем, отчего, для чего, как)* – вопросы, ответы на которые представляют собой объяснения тех или иных явлений: «Что такое простое число?», «Почему началась Первая мировая война?» и т. п.

Существуют различные по структуре и функциям вопросы. Среди них:

**Правильно и неправильно** поставленные вопросы. В отличие от суждения вопрос не содержит ни утверждения, ни отрицания, поэтому выраженную в нем информацию не оценивают в терминах истины и лжи. Вместе с тем вопрос явно или неявно включает либо опирается на определенное исходное, или базисное знание, выступающее предпосылкой вопроса. Качество базисного знания существенно влияет на логический статус вопроса, определяя правильность и неправильность постановки вопроса. **Правильно поставленным** считается вопрос, предпосылка которого представляет собой истинное и непротиворечивое знание.

**Неправильно поставленным** считается вопрос с ложным или противоречивым базисом. Примером неправильно поставленного вопроса может быть следующий вопрос: «Какой размер обуви у снежного человека?». Предпосылка этого вопроса включает отнюдь не бесспорную идею о существовании снежного человека. О таком вопросе говорят как о неправильном или некорректно поставленном – прежде чем выяснять размер обуви снежного человека, следует установить факт существования снежного человека. В том случае, когда неправильно поставленный вопрос умышленно используется с целью запутать отвечающего, такой вопрос квалифицируют как улавливающий или «провокационный».

Вопросы могут быть закрытыми и открытыми.

**Открытые** вопросы предполагают возможность неограниченного количества ответов.

**Закрытые** вопросы – это вопросы, на которые возможно дать ограниченное количество ответов. Такие вопросы чаще всего используют во время социологических опросов. Например, вопрос «Как Вы относитесь к деятельности Государственной налоговой администрации РФ?» является открытым, потому что человек может высказывать свое собственное мнение по этому поводу.

Этот же вопрос можно сформулировать в закрытой форме, если добавить к нему перечень возможных ответов. Например, «Как Вы относитесь к деятельности Государственной налоговой администрации РФ?» (Считаю ее деятельность удовлетворительной; считаю ее деятельность неудовлетворительной; не могу ответить).

По логической структуре и познавательной функции вопросы подразделяются на два основных типа:

- ✓ **уточняющие** или ли-вопросы;
- ✓ **восполняющие** или что-вопросы.

**Уточняющим** называется вопрос, направленный на выявление истинности выраженного в нем суждения. Например: «Верно ли, что Луна вращается вокруг Земли?». Грамматический признак уточняющих вопросов – наличие в предложении частицы «ли»: «Верно ли, что...»; «Является ли ...»; «Так ли, что...»; «Действительно ли, что...» и другие синонимные выражения.

**Восполняющим** называется вопрос, направленный на выявление новых свойств у исследуемых явлений: «Кто изобрел телефон?». Грамматический признак восполняющих вопросов – наличие в предложении вопросительных слов: кто? что? где? когда? как? и других, с помощью которых стремятся получить дополнительную информацию о том, что собой представляет исследуемый объект.

## 2.1. Логико-семантическая сеть «Вопрос-Ответ-Реакция»

Прототип семантической поисковой системы на основе ЛСС, общие положения системы «ВОПРОС-ОТВЕТ» описаны в работах [3], [4]. Логико-семантическая сеть – множество вопросов, ответов и связей между ними, образующих целостную систему «Вопрос-Ответ-Реакция». Целостность ЛСС определяется следующими свойствами:

- ✓ множество «Вопрос-Ответ-Реакция» относится к определенной теме предметной области;
- ✓ множество «Вопрос-Ответ-Реакция» иерархически упорядоченно по принципу «от общего к частному»;
- ✓ на нечетном уровне иерархии расположены вопросы, на четном уровне – ответы и реакции;
- ✓ вопросы  $i$ -го уровня иерархии связаны только и только с ответами  $i+1$ -го уровня;
- ✓ вопросы  $i+1$ -го уровня могут быть связаны с ответами  $i$ -го уровня;
- ✓ вопрос  $i$ -го уровня семантически связан с ответами  $i+1$ -го уровня, если удовлетворяет определенному условию ‘А’ или ‘В’. В случае удовлетворения условию ‘А’ имеет место конечная вершина, а в случае удовлетворения условию ‘В’ из данного ответа следуют вопросы  $i+2$ -го уровня;
- ✓ на  $i=1$ -м уровне находятся вопросы, которые раскрываются множеством ответов  $i=2$ -го уровня, частично или полностью охватывающим тему предметной области;
- ✓ на  $i=3$ -м уровне находятся вопросы, которые дополняют и уточняют ответы  $i=2$ -го уровня.

Таким образом, ЛСС «Вопрос-Ответ-Реакция» можно представить в виде направленного графа (рис. 1).

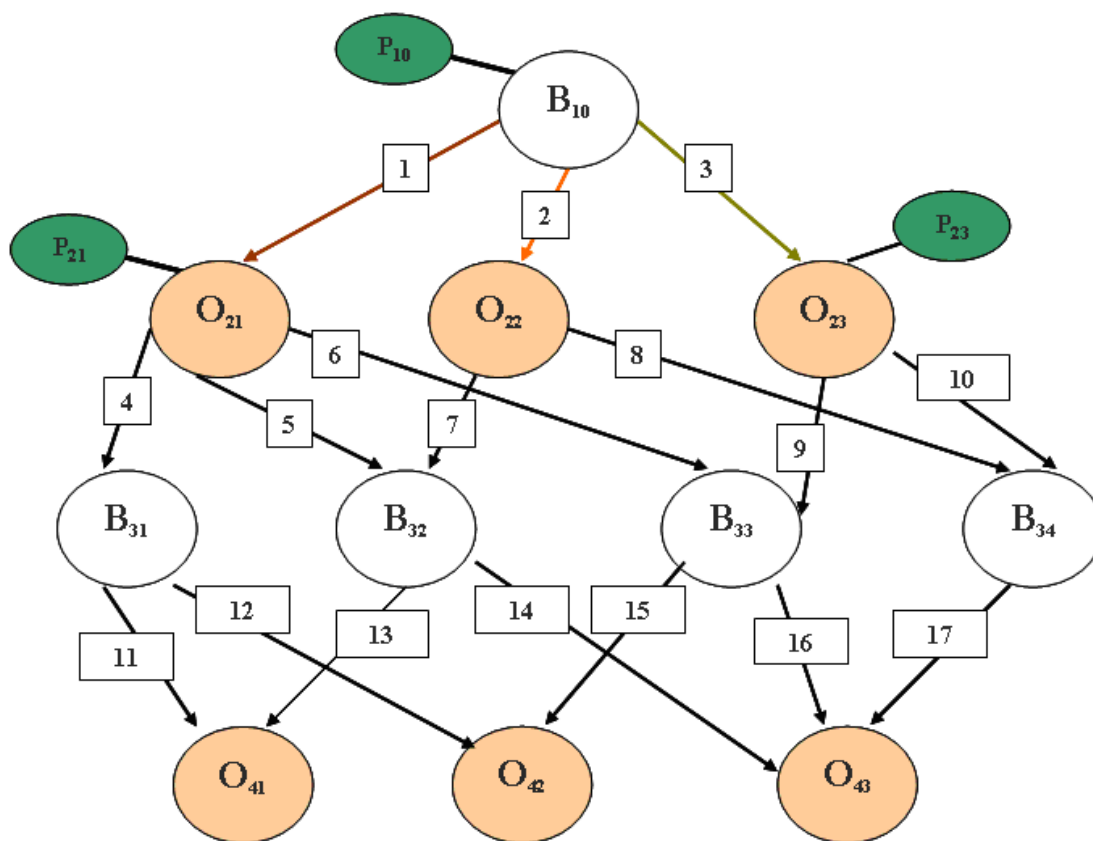


Рис. 1. Граф ЛСС «Вопрос-Ответ-Реакция»

Суть излагаемого подхода состоит в том, что любая задача или научно-технический текст могут быть представлены в виде логической последовательности вопросов и ответов, которая дополняется полезной информацией (реакцией).

Итак,

**Вопрос** – это выраженный в форме вопросительного предложения запрос, направленный на развитие (уточнение) или дополнение знаний.

**Ответ** – это реализация познавательной функции вопроса в форме вновь полученного суждения. При этом по содержанию и структуре ответ должен строиться в соответствии с поставленным вопросом. Лишь в этом случае ответ расценивается как релевантный, т.е. как ответ по существу поставленного вопроса.

**Реакция** – это смысловое описание вопроса и ответа, характеризующее предпосылки вопроса и область поиска ответа [1], [2]. Поясним значение *реакций*. Реакция позволяет учитывать и использовать дополнительные знания о предметной области. В процессе познания вопросы не возникают ниоткуда, они всегда опираются на уже известное знание, выступающее их базисом и выполняющее роль предпосылки вопроса. Познавательная функция вопроса реализуется в форме ответа на поставленный вопрос. Поиск ответа предполагает обращение к конкретной области теоретических или эмпирических знаний, которую называют областью поиска ответа. Полученное в ответе знание, расширяя либо уточняя исходную информацию, может служить базисом для постановки новых, более глубоких вопросов о предмете исследования [3]. Постановка вопроса и поиск информации для конструирования ответа составляют вопросно-ответную логическую форму развития знаний. Она всегда была направляющим началом в развитии естествознания и техники. Ввод реакций помогает пользователю понять, получил ли он релевантный и pertinentный ответ на свой вопрос. В качестве реакций может выступать дополнительная информация по теме вопроса и ответа, иллюстрации, изображения, таблицы, ссылки на сайты, словари, рубрикаторы, каталоги и т. д. Такими реакциями может сопровождаться как вопрос, так и ответ, что позволит пользователю лучше и быстрее сориентироваться в предметной области.

Типы реакций:

**1. Реакции вопроса** – это описание области предпосылки вопроса (для осознания обстоятельств и причин возникновения вопроса и дальнейшего установления смыслового соответствия с областью ответа). Реакция вопроса характеризует область вопроса – смысловое пространство, из которого аргументируется возникновение вопроса (хотя самой аргументации нет).

**2. Реакции ответа** – это описание области ответа (для осознания смысла вопроса и смысловой связи с ответом). Реакция ответа – смысловое пространство, имеющее связь с пространством вопроса, из которого следует ответ. Каркасом связи вопроса и ответа является ЛСС.

Данный подход позволяет заменить неопределенности, связанные с информационным поиском (когда неясно точно, какая информация ищется и с какой целью) на более продуктивную технологию, ориентированную на пространство смыслов [5]. Как уже было сказано выше, для профессионала, осуществляющего поиск, всегда есть осознанность той информации, которая является или может являться ответом на его вопрос. В этом случае ему проще сформировать вопрос и иметь возможность от полученного ответа с помощью специализированного навигатора либо уточнить вопрос, либо его углублять, получая соответствующие связки вопрос-ответ. Тем самым, пользователь от имеющихся знаний может получить расширенные знания, углубленные знания, уточненные знания или новые знания. При этом за счет реакции пользователь может контролировать согласованность смыслового собственного понимания вопросов и ответов и понимания вопросов и ответов, заложенных в семантической поисковой системе. Поскольку система открытая, пользователь в процессе взаимодействия может уточнять и расширять саму ЛСС. Т.е. пользователь при активном развитии системы становится автором смыслового пространства ЛСС. В этом состоит адаптация системы.

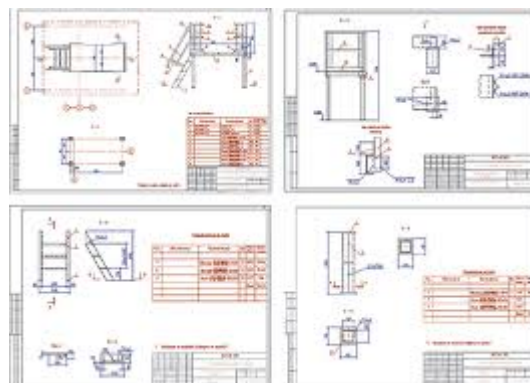
Проиллюстрируем роль реакции на примере. Рассмотрим логическую связку «*Вопрос-Ответ-Реакция*», сформированную на материале учебного пособия по базам данных:

**Вопрос 1.** Что представляет собой проектирование БД?

**Реакция 1 Вопроса 1.** Проектирование – деятельность человека или организации(ий) по созданию проекта, то есть прототипа, прообраза предполагаемого или возможного объекта, состояния; комплекта документации, предназначенной для создания определенного объекта, его эксплуатации, ремонта и ликвидации, а также для проверки или воспроизведения промежуточных и конечных решений, на основе которых был разработан данный объект (см. <http://ru.wikipedia.org/wiki/>).

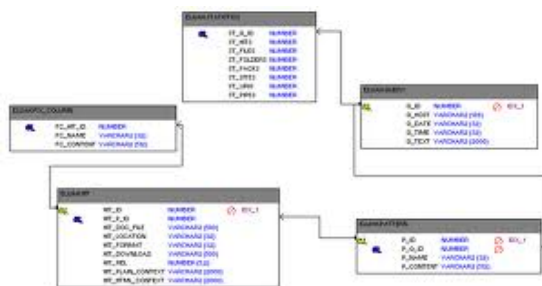
**Реакция 2 Вопроса 1.** От специфического для машиностроения, строительства и других отраслей науки и техники понятия «проект» (англ. design) в значении «проектная документация» следует отличать используемое в области деятельности управление проектами в контексте менеджмента понятие «проект» (англ. project, от лат. Projectus – брошенный вперед, выступающий) в значении «некоторая задача с определенными исходными данными и требуемыми результатами (целями), обуславливающими способ ее решения», «программа», «комплекс работ» и т. п. (см. <http://ru.wikipedia.org/wiki/>).

**Реакция 3 Вопроса 1.** Иллюстрации. Проектная документация.



**Ответ 1 на Вопрос 1.** Проектирование баз данных – процесс создания схемы базы данных и определения необходимых ограничений целостности в соответствии с требованиями пользователей.

Реакция 1 Ответа 1 на Вопрос 1. Иллюстрация схемы БД.



В качестве реакций может выступать текстовая информация со ссылками на первоисточники, иллюстрации (рисунки, графики, таблицы, слайд-шоу, видео-ролики и пр.) и/или их сочетание. Реакции помогают пользователю осознать смысловое поле вопроса и полученных на него ответов, тем самым позволяют улучшить прагматическую метрику информационного поиска – pertinence.

Основными направлениями исследования для разработки метода и механизма эффективного поиска множества релевантных ответов на вопрос являются:

(1) Разработка технологии формирования и поддержки «каталожной» службы информационного фонда, обеспечивающей эффективный поиск ответов на вопросы, на основе ЛСС «Вопрос-Ответ-Реакция».

(2) Создание инструментария (ПО) – АРМ аналитика для структурирования информационного фонда, предназначенный для создания и редактирования множества ЛСС.

Основой метода является способ описания научно-технической и образовательной информации множеством логико-семантических сетей «вопрос-ответ-реакция». Основой механизма поиска является способ движения по ЛСС, управляемый пользователем посредством выбора в ЛСС узлов – вопросов или ответов – на основе онтологической модели вопроса пользователя (см. рис. 1).

## 2.2. Формальное представление предметной области

Предмет исследования любой научно-практической области знаний может быть представлен проблемным полем – перечнем проблемных вопросов. Проблемные вопросы в свою очередь могут быть представлены в виде иерархического дерева по принципу «от общего к частному». Для некоторых вопросов уже существуют возможные альтернативные ответы и способы их реализаций (реакции). Для понимания вопроса также необходима определенная реакция. Ответы могут порождать в свою очередь вопросы (рис. 1). Таким образом, проблемный вопрос соотносится к определенной теме предметной области и раскрывается семантической структурой вопрос-ответ-реакция, которая в общем случае является открытой (т.е. пополняемой, изменяемой во времени). Т.е. знания, накопленные в предметной области, могут быть представлены открытым множеством логико-семантических сетей, упорядоченных по предметным темам. Аналогично, некоторая задача предметной области может быть сформулирована в форме вопроса. А решение этой задачи может быть представлено серией вопросов и ответов с соответствующими описаниями предобласти вопроса и предпосылки ответа (реакциями). Выявление в вопросе таких смыслов, как тема вопроса, содержание вопроса, объем вопроса позволяет найти релевантные ЛСС, в которых могут содержаться как ответы, так и объяснения (реакции). Под предметной областью будем понимать область научно-практической деятельности человека, характеризующаяся объектом и предметом исследования. Предметом исследования являются проблемы и задачи, связанные с объектом. Теории, методы, инструменты, опыт специалистов, научные и эмпирические знания и метазнания – ресурс, который используется специалистами для исследования проблем, продуцирования новых знаний, разработки теорий и технологий, решения научных и практических задач. Будем считать, что накопленные знания предметной области представлены в научных текстах – отчетах, монографиях, статьях, учебных материалах, информационных фондах, справочниках, словарях и т.д. Весь объем информации представляется множеством тематических разделов, каждый из которых отражает определенный аспект знания предметной области. Каждой теме можно поставить в соответствие ЛСС «Вопрос-ответ-реакция». В этом случае знания предметной области формально можно представить следующим образом.



Введем следующие условные обозначения:

*ПрОб* – наименование предметной области,

$T_{m_i}$  – наименование  $i$ -й темы предметной области,

$LCC_{ij}$  – наименование  $j$ -й ЛСС  $i$ -й темы.

Тогда, предметная область выражается как:

$$ПрОб = \bigcup_{i=1}^n T_{m_i}; T_{m_i} = \bigcup_{j=1}^{m_j} LCC_{ij},$$

причём  $\bigcap T_{m_i} T_{m_j} \neq 0$  для  $i \neq j$ ,

$\bigcap LCC_{ij} LCC_{kr} \neq 0$  для  $i \neq k, u j \neq r$ .

Таким образом, предметную область можно описать множеством ЛСС, объединенных в тематические классы. ЛСС предметной области могут служить основой для создания поисковых информационных систем. Механизм поиска информации в предметной области на основе ЛСС может обеспечивать следующие режимы: движение по ЛСС, управляемое пользователем (режим навигации); поиск информации по заявке (вопросу).

## 2.3. Формирование ЛСС Вопрос-Ответ-Реакция

### 2.3.1. Анализ научного текста

Прежде чем приступить к созданию ЛСС для данного ресурса (единицы хранения электронного фонда) необходимо убедиться, что он представляет собой научный текст, т.е. текст в научном стиле речи, в котором можно выделить несколько разновидностей:

а) академический (собственно научный) – наиболее строгий, точный; диссертации, монографии, статьи научных журналов, инструкции, ГОСТы, энциклопедии;

б) научно-популярный (научные статьи в газетах, научно-популярных журналах, научно-популярные книги; публичные выступления в средствах массовой информации на научные темы, выступления ученых, специалистов перед массовой аудиторией);

в) научно-учебный (учебная литература по разным предметам для разных типов учебных заведений; справочники, пособия).

Вне зависимости от жанра таким текстам присущи основные признаки: логичность изложения, смысловая точность (однозначность), обобщенность, объективность.

Логичность изложения научного текста предполагает, что все части в нем связаны по смыслу и располагаются строго последовательно; выводы вытекают из положений и фактов, излагаемых в тексте.

Смысловая точность (однозначность) достигается четкостью формулировок, использованием слов в их прямом значении, широким употреблением терминов и специальной лексики.

Обобщенность выражается в использовании абстрактных понятий, слов с отвлеченным значением (например: сущность, время, сила, количество, качество, закон, число), формул, символов, условных обозначений, графиков, таблиц, диаграмм, схем, чертежей.

Научный стиль имеет лексические, морфологические, синтаксические, структурные особенности, а также характерное использование средств выразительности. Так, в научном тексте недопустимо личное, субъективное мнение, не принято использовать местоимение *я* и глаголы в первом лице единственного числа. Предпочтительны неопределенно-личные (считают, что...), безличные (известно, что...), определенно-личные (рассмотрим проблему...) предложения. В научном тексте отсутствуют образные средства: эпитеты, метафоры, художественные сравнения, поэтические символы, гиперболы. В научной речи распространены сложные предложения, особенно сложноподчиненные, т.к. сложноподчиненные предложения с придаточными изъяснительными выражают обобщение,



раскрывают явление, закономерность. В зависимости от жанра структура научного текста может варьироваться, но всегда кроме основного содержания он включает и другие обязательные элементы. Обычно научный текст включает:

- заглавие,
- сведения об авторах,
- аннотация,
- ключевые слова,
- основной текст,
- библиографические ссылки,
- библиографический список.

Особенности научного стиля речи представлены в таблице 1.

Таблица 1. Характеристики научного стиля речи

<b>Основной тип речи</b>	Рассуждение и описание
<b>Лексика</b>	<ol style="list-style-type: none"> <li>1. наличие общих и специальных терминов (узкоспециальная лексика);</li> <li>2. однозначность трактовки терминов;</li> <li>3. однозначность слов;</li> <li>4. частая повторяемость ключевых слов;</li> <li>5. употребление наречий (<i>сначала, прежде всего, потом, затем, далее</i>), вводных слов (<i>во-первых, во-вторых, в-третьих, наконец, итак, следовательно, наоборот</i>) и союзов (<i>так как, потому что, чтобы, поэтому</i>) для обозначения последовательности развития мысли;</li> <li>6. недопустимость вневитературной лексики, жаргона и сленга.</li> </ol>
<b>Состав слова</b>	<ol style="list-style-type: none"> <li>1. интернациональные корни, приставки, суффиксы;</li> <li>2. суффиксы, придающие отвлеченное значение.</li> </ol>
<b>Морфология</b>	<ol style="list-style-type: none"> <li>1. преобладание существительных;</li> <li>2. частое использование абстрактных отглагольных существительных;</li> <li>3. неупотребительность местоимений я, ты и глаголов 1-го и 2-го лица единственного числа;</li> <li>4. неупотребительность восклицательных частиц и междометий.</li> </ol>
<b>Синтаксис</b>	<ol style="list-style-type: none"> <li>1. прямой порядок слов (предпочтителен);</li> <li>2. широкое использование словосочетаний сущ. + сущ. в род. п.;</li> <li>3. преобладание неопределенно-личных и безличных предложений;</li> <li>4. редкое использование неполных предложений;</li> <li>5. обилие сложных предложений (особенно, сложноподчиненные предложения с придаточными изъяснительными);</li> <li>6. частое употребление причастных и деепричастных оборотов.</li> </ol>
<b>Использование средств выразительности</b>	<ol style="list-style-type: none"> <li>1. отсутствие художественных образных средств;</li> <li>2. усилительные частицы, местоимения, наречия: <i>лишь, абсолютно, только</i>;</li> <li>3. прилагательные типа: колоссальное, наивыгоднейшее, величайший, труднейшая, громадный;</li> <li>4. проблемные вопросы.</li> </ol>
<b>Структура</b>	<p>Типовая структура:</p> <ul style="list-style-type: none"> <li>- заглавие,</li> <li>- сведения об авторах,</li> <li>- аннотация,</li> <li>- ключевые слова,</li> <li>- основной текст,</li> <li>- библиографические ссылки,</li> <li>- библиографический список.</li> </ul>

### 2.3.2. Методика анализа научных текстов для формирования ЛСС «Вопрос-Ответ-Реакция»

Излагаемая методика применима только к научным текстам (п. 2.3.1.). Для того, чтобы построить ЛСС «Вопрос-Ответ-Реакция» документа (например, научной статьи), необходимо проанализировать его. Для этого документ исследуется экспертом (аналитиком) с точки зрения (см. рис. 2.):

- 1) смыслового соответствия заглавия и содержания;
- 2) набора фильтров:

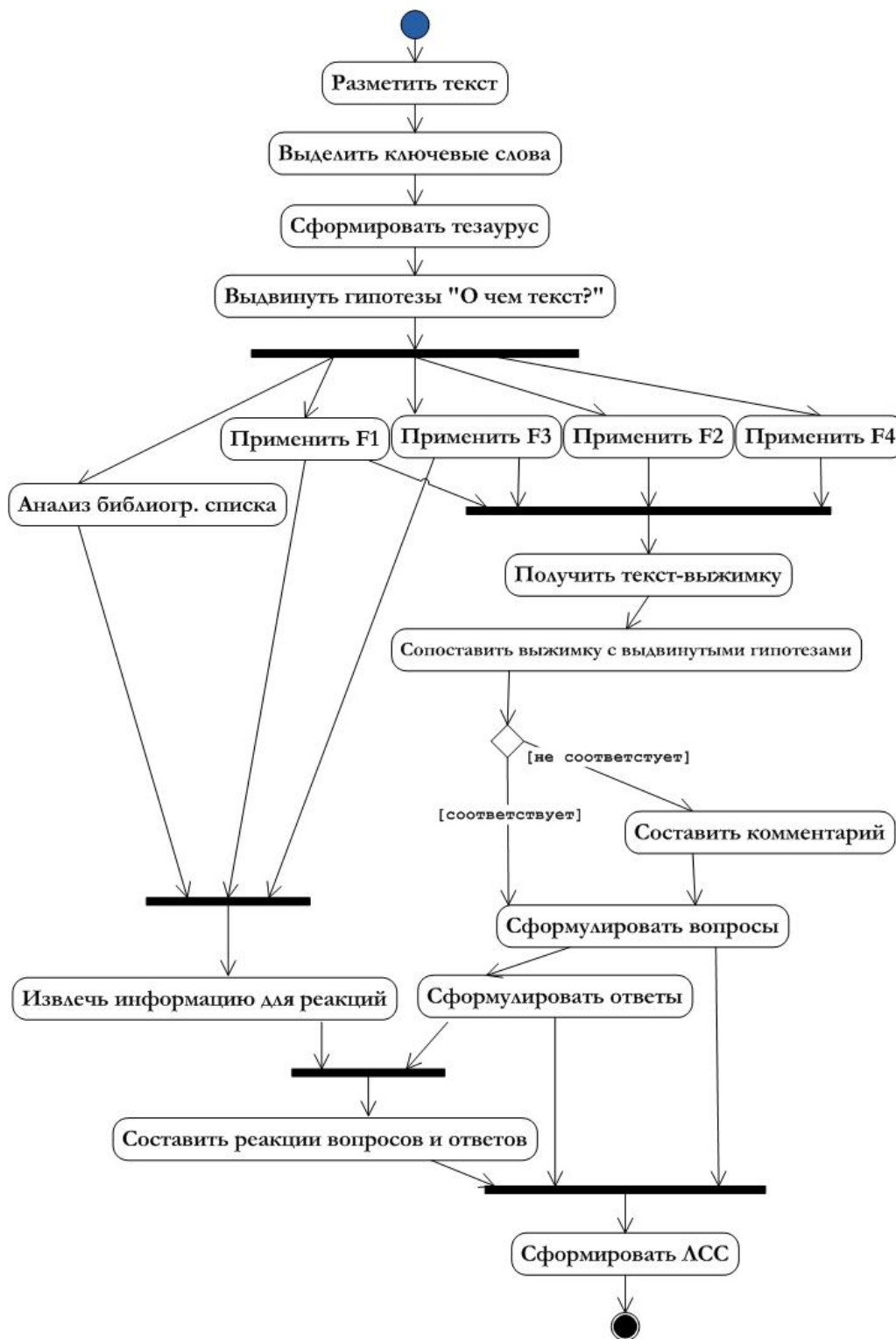


Рис. 2. Рабочий процесс аналитика

**Фильтр 1 (F1)** – Общая часть. F1 включает анализ проблемы, ее историю, обзор, актуальность.

**Фильтр 2 (F2)** – Авторские понятия. F2 включает вводимые авторами новые термины, общеупотребляемые термины с авторской интерпретацией, сужающие семантику.

**Фильтр 3 (F3)** – Примеры и иллюстрации. Предназначены для пояснения сложных мест в тексте, позволяют сократить размер текста при строгих ограничениях по объему.

**Фильтр 4 (F4)** – Идея автора. Описывает и раскрывает основную авторскую идею.

3) Формирования базовых вопросов, на которые отвечает текст.

На полученном таким образом материале далее строится ЛСС нижнего уровня (ЛСС документа):

1. Из названия научного текста извлекаются ключевые слова.

2. Формируется тезаурус.

3. Выдвигаются гипотезы, о чем идет речь в тексте: какие проблемы обсуждаются в анализируемом тексте.

4. Текст разбивается на несколько частей (информационных блоков), к которым применяются фильтры F1, F2, F3, F4.

5. Для полученных информационных блоков формулируется основная мысль – набор предложений (утверждений). Таким образом, мы получаем сжатое изложение (выжимку) текста.

Сопоставив полученную выжимку с выдвинутыми гипотезами, мы получаем возможность делать выводы о том, насколько заглавие текста соответствует его реальному содержанию. После этого этапа можно приступить к формированию ЛСС нижнего уровня:

Формулирование вопросов к выделенным информационным блокам.

Выделения ответов из анализируемого текста и ссылок на них.

Формирование реакций вопросов и ответов. Для ЛСС научного текста (нижнего уровня) реакции вопросов и ответов формируются из информационных блоков по фильтру *F1. Общая часть*, а также по библиографическим ссылкам.

Формирование графа «Вопрос-Ответ-Реакция».

### **3. Организация информационного поиска на основе ЛСС в электронных фондах**

Рассматриваемый подход основан на качественном анализе научных текстов (п. 2.3.2).

Итак:

- структурированная, слабо структурированная информация предметной области может быть представлена в виде логико-семантических сетей «вопрос-ответ-реакция»;
- логико-семантическая сеть отражает определенную тему предметной области;
- предметная область представляется множеством тем;
- предметная задача может быть представлена в форме вопроса (или совокупности вопросов);
- решение задачи может быть представлено в форме ответа или совокупности ответов;
- способ решения задачи может быть представлен унифицированным механизмом поиска реакций на множестве логико-семантических сетей;
- качество решения задачи представляется как уровень пертинентности ответов на заданный вопрос.

Предполагается, что:

- ЛСС являются основой для структурирования произвольных текстов научно-технической информации,
- ЛСС являются основой структурирования знаний предметной области,
- поиск релевантной информации по запросу может осуществляться на основе унифицированного механизма поиска на ЛСС,
- Таким образом, создание семантической поисковой системы на основе ЛСС *Вопрос-Ответ-Реакция* включает следующие стадии:
  - Разработка теоретических положений технологии поиска ответов на вопросы для специализированных научных корпусов.
  - Разработка автоматизированной технологии формирования и поддержки специализированных научных корпусов.
  - Разработка структурно-функциональной модели семантической поисковой системы на основе ЛСС ВОР.
  - Реализация прототипов компонент системы.

Семантические поисковые системы на основе ЛСС «Вопрос-Ответ-Реакция» могут иметь широкий спектр применимости, в том числе и в электронных библиотеках. Для реализации подобной системы в рамках конкретной ЭБ необходимо:

- сформировать набор ЛСС;
- реализовать механизм поиска информации в режиме ответа на вопрос;
- реализовать навигационный механизм для обеспечения движения вверх (от частного к общему) и вниз (от общего к частному) по ЛСС.

Построение множества ЛСС возможно на нескольких уровнях (слоях), охватывающих собственно документ, тему предметной области, предметную область в целом и тезаурус – см. рис. 3.

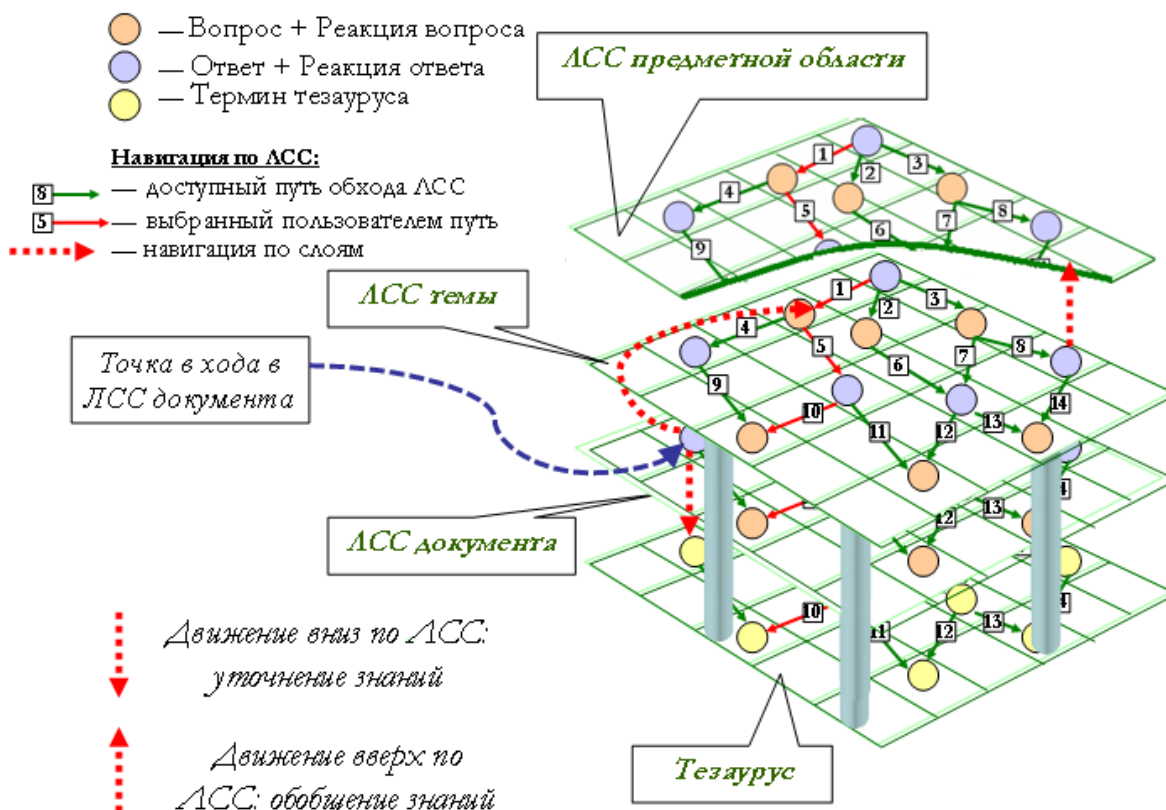


Рис. 3. Многослойный связанный набор графов ЛСС.

Первый слой содержит ЛСС конкретного информационного ресурса (документа). Узлы графа этого слоя ЛСС связаны с элементами тезауруса (нулевой слой). Следующим слоем является ЛСС, построенная на базе функционирующего в данной ЭБ тематического рубрикатора, затем следует слой ЛСС предметной области. Таким образом, мы получим многоуровневый связанный набор графов, который обеспечивает поиск и навигацию в горизонтальном и вертикальном направлении. Навигационный механизм здесь играет важную роль. Пользуясь им, пользователь получает возможность корректировать вопросы, на которые в системе не существует адекватных ответов. Движение по многоуровневой логико-семантической сети «вопрос-ответ-реакция» (рис. 3) обладает следующими свойствами:

1. Движение с  $i$ -го на  $i + k$ -й уровень сети углубляет знания.
2. Движение по горизонтали сети расширяет знания.
3. Движение вверх по сети обобщает знания.

Классификация вопросов может осуществляться по реперным словам, которые определяют смысловую направленность вопроса. К таким словам (словосочетаниям) относятся: почему, зачем, отчего, где, когда, кто, как, что, какой (какими), в чем, с какой целью. Так, *ли*-вопросы направлены на уточнение знаний. Вопросы кто? что? где? когда? – на углубление знаний. Вопросы как? – на расширение знаний. Вопросы почему? зачем? отчего? направлены на обобщение знаний.

Задачи **навигации** – нахождение оптимального маршрута (траектории), определение местоположения, направления и значения скорости и других параметров движения объекта.

Задача навигации состоит в выделении эффективной части траектории ЛСС. При этом определено: начальная точка движения и уровень смыслового соответствия реакций вопросов системы с реакциями вопросов пользователя.

Под эффективностью понимается время, потраченное на выделение траектории ЛСС, удовлетворяющей заданным условиям.

Применяя средства визуализации, опирающиеся на эмоциональное восприятие и имеющийся опыт пользователя, можно добиться еще более эффективного использования ресурсов электронных библиотек. С точки зрения пользователя такая система позволяет в большинстве случаев найти ответ на поставленный вопрос. Пользователь задает вопрос и получает на него ответ с дополнительной информацией в виде реакций вопроса и ответа, которые помогают скорректировать вопрос, либо воспользоваться уточняющими или обобщающими вопросами. В режиме вопрос-ответ реализуется расчет меры близости вопроса, заданного пользователем, к уже существующим в ЛСС. Если такого вопроса нет, то он в дальнейшем может быть внесен в систему.

Создание, наполнение и сопровождение такой системы требует большой и серьезной работы как технологической, так и организационной. Создание каталожной службы является трудоемким ручным процессом. Поэтому для создания технологии формирования и поддержки каталога ЛСС требуется максимальная автоматизация, чтобы предоставить АРМ аналитикам, которые будут заниматься формированием ЛСС документов и предметных областей. При успешной реализации этой системы пользователям ЭБ будет предоставлена новая возможность – получать ответы на вопросы, заданные на естественном языке.

К настоящему моменту имеются следующие наработки в данном направлении. Совместно с НИЦ Управления знаниями и распределенными вычислениями университета «Дубна»:

1. Разработана методика и предложена технология формирования ЛСС документа.
2. Разработано ПО для ввода, редактирования, накопления ЛСС в БД: прототип автоматизированного рабочего места разработчика логико-семантических сетей «вопрос-ответ-реакция».
3. Разработана методика поиска ответа на вопрос для корпуса документов.
4. Разработано ПО (в частном случае) поисковой машины.

Осуществлена ручная апробация методик на примерах. Методики апробированы в учебном процессе для различных дисциплин (Корпусная лингвистика, Теоретические основы автоматизированного управления, Системное моделирование и т.д.). В результате совместно с технопарком г. Дубна в

рамках проекта «Разработка портала Содействие инновационной деятельности» была разработана ЛСС «Поиск инвестора».

## Заключение

Авторами предлагается создание «каталожной службы» и ее поддержки для фондов-корпусов, создание вопросно-ответного навигатора, обладающего особенностями, обеспечивающие такие качества, как: возможность в процессе поиска ответов уточнение, углубление понимания смысла вопроса, возможность в процессе поиска ответа уточнения, углубления, расширения знаний или, как следствие, получения нового знания. Ожидается, что производительность вопросно-ответной поисковой системы на основе ЛСС будет превосходить по показателю *скорость/качество* на 30% существующие поисковые системы. Для подтверждения данного показателя был предложен эксперимент, который заключается в следующем. Три контрольные группы пользователей осуществляют информационный поиск по заданной теме, чтобы получить ответы на сформулированные вопросы. По каждой группе вычисляется среднее время поиска, релевантность и пертинентность полученных результатов. Механизм поиска в группе «А» – информационные запросы. Группа «Б» выполняет ту же работу, используя в качестве механизма поиска вопросно-ответную систему AskNet. Группа «С» выполняет поиск посредством ЛСС-навигатора.

Авторами апробирована в учебном процессе на различных дисциплинах методика анализа научной информации, необходимого для построения множества ЛСС.

Поскольку система открытая, в общем случае для одной единицы хранения фонда может быть построено несколько ЛСС разными экспертами. В виду субъективных причин (особенностей восприятия аналитиком исследуемого текста) личные мнения и профессиональные оценки экспертов будут различаться в большей или меньшей степени, что приведет к различию сформированных ими ЛСС. Это может повлечь искажение авторской мысли. В таком случае есть вероятность, что конечный пользователь не сможет удовлетворить своей информационной потребности (не достигнет ожидаемого уровня пертинентности), пользуясь выбранным вопросно-ответным навигатором. Но он всегда может обратиться в системе к полнотекстовому документу (оригиналу).

Основной проблемой создания предлагаемой вопросно-ответной системы является необходимость максимальной автоматизации процесса создания и поддержки «каталожной службы» фонда.

## Список литературы

1. Filozova I.A., Korenkov V.V., Musulmanbekov G.. Towards Open Access Publishing at JINR. – Proceedings of XXII International Symposium on Nuclear Electronics and Computing (NEC'2009). – Varna, Bulgaria, Sept. 7-14, 2009. – Dubna: JINR, 2010. – p.124-128 – (JINR ; E10,11-2010-22).
2. Паринов С.И., Коголовский М.Р. Технология семантического структурирования контента научных электронных библиотек // Труды XIII Всероссийской научной конференции RCDL'2011 «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». – Воронеж: Воронежский университет, 2011. – [Электронный ресурс]. URL: <http://rcdl.ru/doc/2011/paper13.pdf>.
3. Добрынин В.Н., Лобачева М.В. Прототип семантической поисковой системы на основе логико-семантической сети «Вопрос-Ответ-Реакция» // Системный анализ в науке и образовании: сетевое научное издание. – 2009. – №2. – [Электронный ресурс]. URL: <http://www.sanse.ru/archive/11>.
4. Добрынин В.Н., Филозова И.А. Поиск на основе логико-семантической сети Вопрос-Ответ-Реакция в научной электронной библиотеке // Труды XII Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». – ISBN 978-5-98180-838-8. – Казань: Казанский университет, 2010 – С. 301-308.
5. Аверьянов Л. Я. Почему люди задают вопросы? – М.: «Социолог», 1993.