

ПОИСК НА ОСНОВЕ ЛОГИКО-СЕМАНТИЧЕСКОЙ СЕТИ «ВОПРОС-ОТВЕТ-РЕАКЦИЯ» В НАУЧНЫХ ИНФОРМАЦИОННЫХ ФОНДАХ

Добрынин Владимир Николаевич¹, Филозова Ирина Анатольевна²

¹Кандидат технических наук, профессор Института системного анализа и управления;
ГОУ 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.

² Научный сотрудник;
Объединенный институт ядерных исследований;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
e-mail: fia@jinr.ru.

В статье излагается описание технологии семантического поиска в электронных библиотеках на основе логико-семантической сети «Вопрос-Ответ-Реакция» (ЛСС ВОР), адекватной смыслу текста. Такая поисковая система позволит пользователю получить ответ на сформулированный на естественном языке вопрос. Излагается имеющийся опыт и наработки в данной области. Приводится краткий обзор известных вопросно-ответных поисковых систем.

Ключевые слова: информационный поиск, вопросно-ответные системы, логико-семантическая сеть Вопрос-Ответ-Реакция, электронные библиотеки, семантический учебник.

THE SEARCH BASED ON THE LOGICAL SEMANTIC NETWORK «QUESTION-ANSWER-REACTION»

Dobrynin Vladimir¹, Filozova Irina²

¹ Candidate of Science in Engineering, professor;
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.

² Scientist;
Join institute for nuclear research;
141980, Moscow reg., Dubna, Joliot-Curie, 6;
e-mail: fia@jinr.ru.

The technology of semantic search in digital libraries based on the framework of Logical Semantic Network (LSN) «Question-Response-Reaction» is described. Such a system allows one to get an adequate response on the question, formulated in human language.

Keywords: informational search, technology, Question-Answer Systems, Logical Semantic Network (LSN) «Question-Response-Reaction», digital libraries, semantic book.

Введение

Информационные потребности пользователей научных электронных библиотек (ЭБ) определяются задачами, которые они решают в процессе своей профессиональной деятельности. Этими задачами могут быть: исследование, экспертиза, инженерная задача, конструкторская задача, научное руководство и пр. Коммуникация специалистов-профессионалов в данной предметной области эффек-

тивна, когда она происходит по принципу вопрос – ответ. Когда пользователь решает некоторую задачу, то, обращаясь к информационной системе, он хочет получить ответ на свой вопрос. Но, приступая к решению новой для себя задачи, пользователь может неточно и/или неполно сформулировать свой вопрос. Это естественно, т.к. полная ясность наступает, когда задача уже решена. Тогда возникает типичная проблема: ответ есть в системе, но он не найден, т.к. вопрос сформулирован неточно. Но если пользователь сможет уточнять свой запрос в процессе поиска, он получит новую формулировку вопроса, что приблизит его к осознанию исследуемой проблемы.

Работа специалиста-профессионала с фондами предполагает наличие системы каталогизации и классификации материала. В рамках заданной проблемной темы предмета предлагается технология формирования и поддержки «каталожной» службы, которая обеспечивает эффективный поиск ответов на вопросы. Стержнем такой «каталожной» службы является упорядоченное открытое множество множеств логико-семантических сетей 2Вопрос-ответ-реакция» (ЛСС «ВОР»). Множество исходных документов фонда представляются как корпус, ориентированный не только на пользовательские вопросно-ответные потребности, но и на возможность его использования для решения лингвистических задач, связанных с языковыми особенностями документов фонда. Результаты решения лингвистических задач могут служить базой для семантической дифференциации ЛСС ВОР.

Семантические поисковые системы

Стандартные поисковые системы выдают список ссылок на найденные страницы, содержащие слова запроса. Навигацией в массиве найденных ссылок, анализом страниц и поиском необходимой информации пользователь вынужден заниматься самостоятельно.

В настоящее время семантические поисковые системы активно создаются и совершенствуются. Но, зачастую, как результат пользователь получает перечень ссылок на ресурсы, где можно найти ответ. Такие системы дают лучший результат, чем обычные поисковые машины, т.к. понижается уровень информационного шума за счет исключения спама и рекламы, другой лишней информации. Но, тем не менее, это не ответ на вопрос пользователя, а список ресурсов, где он может найти ответ.

Поэтому поиск ответов на вопросы пользователей, заданных на естественном языке, – актуальная задача. В этом направлении уже есть некоторые успехи, и такие поисковые системы существуют и развиваются.

2.1. Краткий обзор информационно-поисковых систем Вопрос-Ответ

Вопросно-поисковая система (QA-система) – информационно-поисковая интеллектуальная справочная система с естественно языковым интерфейсом. Первые QA-системы появились в 1960-е гг. и использовались как естественно-языковые оболочки для экспертных систем.

Среди современных национальных разработок следует упомянуть – Nigma (<http://www.nigma.ru>), AskNet (<http://www.asknet.ru>), Генон (<http://www.genon.ru/>), среди зарубежных – Lexxe (<http://www.lexxe.com/>), Start (<http://start.csail.mit.edu/>), Hakia (<http://hakia.com>), Powerset (<http://www.powerset.com>). Ниже приводится краткое описание перечисленных систем.

Интеллектуальная поисковая система Nigma.ru – это первая кластеризующая поисковая система в Рунете. Конечная цель проекта – создание программного обеспечения, позволяющего анализировать проиндексированные документы и выдавать конкретную информацию на запрос пользователя, а не ссылки на другие сайты. Созданы такие сервисы, как Nigma-математика, Nigma-химия и Nigma-музыка. Разработки ведутся при участии Стэнфордского университета.

Другой пример – система AskNet обеспечивает поиск ответов на запросы пользователей на русском и английском языках. Не требуется переходить по ссылкам, так как все ответы содержатся непосредственно на странице результатов поиска. Поисковая машина AskNet Global Search осуществляет поиск информации с использованием интернет-сервисов существующих поисковых систем и интернет-энциклопедий по запросам удаленных пользователей. Работа Интернет-сервиса AskNet Global Search основана на базовых алгоритмах технологии AskNet Search, адаптированных к особенностям их применения для реализации метапоисковых процедур на сервере.

Система Генон является симбиозом вопросно-ответных и поисковых систем. В Геноне реализована модель накопления и хранения проверенной информации вместе с вопросами, на которые эта

информация отвечает. Добавлять свои знания в Геноне может каждый человек. Ответы и вопросы создаются Авторами, зарегистрировавшимися на Геноне, и проверяются Редакторами на предмет соответствия правилам написания вопросов и ответов (<http://www.genon.ru/rules.aspx>). Таким образом, базу Генона формируют вопросы, на которые есть однозначные, полные и актуальные ответы, не содержащие какого-либо информационного мусора и прямой рекламы товаров или услуг. Помимо интереса, мотивацией для Авторам и Редакторов участвовать в развитии проекта служит схема получения прибыли, зависящая от количества посещений ответов и от их качества.

Поисковая машина Lexxe использует полностью автоматическую технологию поиска без участия редактирования пользователем. Большинство ответов приходят из неструктурированных текстов и веб-страниц в Интернете. В Lexxe применяется вычислительная лингвистика, что позволяет получить более релевантные результаты, чем от обычных поисковых систем. Lexxe достигает этой цели путем анализа и извлечения значения из поискового запроса. Lexxe распознает, когда запрос является вопросом, а затем пытается найти ответ в Сети, извлекая потенциальные ответы с веб-страниц. Если запрос не является вопросом, будет производиться поиск по ключевым словам.

Поисковая система Start была разработана группой InfoLab в Массачусетском технологическом институте информатики и лаборатории искусственного интеллекта в 1993 г., является универсальной системой. В настоящее время может ответить на миллионы вопросов на английском языке о местах (например, города, страны, озерах, координаты, погода, карты, демографических, политических и экономических систем), фильмах (например, названия, актеры, режиссеры), персоналиях (например, даты рождения, биографии), терминах и др.

Поисковая машина Nakia производит поиск во всех сегментах, включая веб-новости, блоги, галереи. Новости, блоги, галереи обрабатываются на основе собственной технологии, называемой QDEXing. Веб, видео и изображения обрабатываются Nakia с помощью алгоритма SemanticRank.

Проведение глубокого исследования этих систем затруднено из-за того, что не для всех перечисленных систем представлена в открытом доступе необходимая для полномасштабного исследования информация. Поэтому был проведен экспресс-анализ в виде следующего эксперимента: в системы вводился один и тот же вопрос, на который заведомо известен однозначный правильный ответ. В данном случае это был вопрос «Где был открыт 105-й элемент периодической таблицы?» В англоязычные системы вводился тот же вопрос на английском языке. Оценивались показатели: общее количество полученных ответов на первой итерации поиска (Σ), количество ответов после фильтрации при наличии в системе соответствующих фильтров (ΣF), количество релевантных ответов (ΣR). Результаты представлены в таблице 1.

Таблица 1.

Система	Σ	ΣF	ΣR	Примечание
Powerset	6	–	0	
Lexxe	100	–	0	
Start	0	–	0	
Nakia	0	–	0	
AskNet	24	–	7	
Nigma	25 000	9800	7	Из первых 100
Генон	10	–	1	

Как видно из представленной таблицы, наиболее высокие показатели у AskNet.

Качество поиска в описанных системах, несомненно, выше, чем в традиционных поисковых машинах. Но в них не учитывается, что пользователь, задающий вопрос, может не быть профессионалом в данной предметной области, поэтому вопрос может быть сформулирован им некорректно. Следствие – другой ответ. Зачастую пользователю трудно понять, является ли полученный им ответ релевантным поставленному им вопросу.

Эти системы являются универсальными и позволяют задавать любые вопросы по всем областям знаний. В основе их работы лежат лингвистические механизмы – морфологический и синтаксический анализ. Все вопросно-ответные машины ориентированы на получение релевантных ответов на вопросы в широких тематических предметных полях. Теоретические основы таких машин имеют общие и частные подходы. Но они не могут служить основой для создания «каталожной» основы корпуса научных документов.

Подход, излагаемый в данной статье, предполагает создание социо-технической системы {*Эксперт + Инструмент*}. Основная идея – создать дополнительный инструментарий для специализированного фонда документов, содержащий научные тексты, протоколы, проекты и т.д.

Семантическая поисковая система на основе логико-семантической сети «Вопрос-Ответ-Реакция»

Рассматриваемый в данной статье подход является экспериментальным и основан на качественном анализе научных текстов.

Информационный поиск документов по запросу всегда подразумевает определенную степень осознанности пользователя: прежде чем составить информационный запрос, пользователь либо осознанно представляет себе, на какой вопрос он хочет получить ответ, или он не может сформулировать вопрос и представляет себе, какая ему необходима информация, чтобы удовлетворить информационную потребность. В последней ситуации он реализует поиск методом проб и ошибок, в процессе которого может либо ясно осознать на какой вопрос искал ответ, либо не получить желаемого. Для профессионалов всегда есть осознанность той информации, которая является ответом на его вопрос. В этом случае ему проще сформулировать вопрос и иметь возможность от полученного ответа с помощью специализированного навигатора либо уточнять вопрос, либо его углублять, получая соответствующие связки (вопрос-ответ). Тем самым, пользователь от имеющихся знаний может получить расширенные знания, углублённые знания, уточнённые знания или новые знания. При этом за счёт 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.

Общие положения, которые легли в основу данной работы:

- структурированная, слабо структурированная информация предметной области может быть представлена в виде логико-семантических сетей «вопрос-ответ-реакция»;
- логико-семантическая сеть отражает определённую 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 релевантности ответов на заданный вопрос.

Предполагается, что:

- логико-семантические сети «вопрос-ответ-реакция» (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 ВОР включает следующие стадии:

1. Разработка теоретических положений технологии поиска ответов на вопросы для специализированных научных корпусов.

2. Разработка автоматизированной технологии формирования и поддержки специализированных научных корпусов.
3. Разработка структурно-функциональной модели семантической-поисковой системы на основе ЛСС ВОР.
4. Реализация прототипов компонент системы.
5. Разработка технического задания (ТЗ).

Общие положения системы «ВОПРОС-ОТВЕТ»

Вопросы возникают там, где есть познавательная неопределенность. Вопрос не является суждением, но в основе его всегда лежит суждение или совокупность суждений. Раздел логики, изучающий вопросы, называется эротетическая или интеррогативная логика, в которой «единицей мысли» выступает комплекс вопроса и ответа – логическая связка вопрос-ответ.

Развитие научных и практических знаний протекает как переход от ранее установленных суждений к новым, более точным и более богатым по содержанию суждениям. Этот переход представляет собой последовательность следующих этапов:

1. постановка вопроса;
2. поиски новой информации;
3. формирование ответа на поставленный вопрос.

Роль вопроса как средства информационного поиска исключительно велика. Без вопроса не может быть познания. Это связующее звено между познанным и непознанным. В форме вопроса осуществляется постановка новых проблем в науке, с помощью вопросов получают новую информацию в социальной, производственной и бытовой сфере. 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 или дополнение знаний.

В процессе познания вопросы не возникают сами по себе. Любой вопрос всегда опирается на уже известное знание, выступающее его базисом и выполняющее роль предпосылки вопроса (*datum questionis*).

Познавательная функция вопроса реализуется в форме ответа на поставленный вопрос. Ответ представляет собой новое суждение, уточняющее или дополняющее в соответствии с поставленным вопросом прежнее знание. 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. 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.

Формулирование вопроса связано с его познавательной функцией, направленной на получение уточняющей информации или расширенного знания, или углубленного знания, или нового знания об объекте исследования. Вопрос, представленный в форме текста, включает ключевые слова и словосочетания, относящиеся к предмету исследования. Между ключевыми словами и словосочетаниями зафиксированы (формой предложения вопроса) определённые отношения. Множество ключевых слов вопроса и отношений между ними образует так называемую онтологическую модель вопроса (ОМВ). Процесс формирования вопроса должен опираться на внутреннюю логику его организации, которая отражается в ОМВ. Это обстоятельство является основой для технологии формирования вопроса.

Процесс поиска ответа на вопрос направлен на поиск информации в области предмета исследования или расширенной области знаний об объекте исследования. Ответ, представленный в форме текста, содержит ключевые слова и отношения между ними. Множество ключевых слов и отношений образуют онтологическую модель ответа (ОМО). Онтологическая модель ответа и система правил,

устанавливающая целостность системы «вопрос-ответ» – основа для технологии поиска ответа на вопрос.

Процесс установления соответствия вопроса и ответа направлен на выявления возможных несоответствий в ответе. В этом случае необходимо расширить либо область поиска ответа, либо область предпосылки вопроса или предмета исследования. Это должно привести к формированию вопроса либо его уточнению, либо расширению.

Любой возможный ответ на тот или иной вопрос должен удовлетворять стандартным «постулатам Хэмблина» [1]:

1. ответ на вопрос должен быть сформулирован в виде высказывания или предложения;
2. возможные ответы на вопрос образуют исчерпывающее множество взаимно исключающих альтернатив;
3. значение вопроса известно, если и только если известно, что может считаться ответом на этот вопрос. Иными словами, сущность вопроса можно отождествить с множеством возможных ответов на него [3].

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, выполняющий свое основное назначение – уточнить неясную или неопределенную и доставить новую информацию.

Если в качестве ответа приводят хотя и истинные, но содержательно не связанные с вопросом суждения, то их расценивают как ответы не по существу вопроса и обычно исключают из рассмотрения. Появление таких ответов в дискуссии – либо результат заблуждения, когда отвечающий не уловил смысл вопроса, но пытается отвечать на него, либо сознательное стремление уйти от невыгодного ответа на поставленный вопрос.

Логическая зависимость между вопросом и ответом означает, что качество ответа во многом определяется качеством вопроса. На расплывчатый и двусмысленный вопрос трудно получить ясный ответ. Для получения точного и определенного ответа необходимо сформулировать точный и определенный вопрос.

Под точностью и определенностью в данном случае имеется в виду логическая, т.е. понятийно-структурная характеристика вопроса. Она выражается в точности употребляемых понятий и вопросительных слов, а также в рациональном использовании сложных вопросов.

Двусмысленные понятия нередко используются в улавливающих или «провокационных» вопросах, в которых содержится скрытая информация. Неопределенность в ответах может быть результатом неясности используемых при постановке вопроса понятий.

Итак,

1. качество ответа зависит от качества вопроса;
2. под качеством вопроса и ответа следует понимать точность и определенность;
3. под точностью и определенностью имеется в виду понятийно-структурная характеристика вопроса и ответа;
4. понятийно-структурная характеристика вопроса выражается в точности употребляемых понятий и вопросительных слов, а также в рациональном использовании сложных вопросов;
5. понятийно-структурная характеристика ответа выражается в точности употребляемых понятий, а также в рациональном использовании сложных ответов.

Формальная структура вопроса и ответа

Будем считать, что логическая структура вопроса включает следующие составляющие:

- тему вопроса (ТВП),
- содержание вопроса (СВП),
- объём вопроса (ОВП).

Под темой вопроса будем понимать предпосылки вопроса (предпосылки вопроса – знания об объекте и предмете исследования, к которым относится вопрос).

Под содержанием вопроса будем понимать предметы, явления, процессы, технологии, инструменты, теории, относящиеся к объекту и предмету исследования, связи между ними посредством фиксации общих и специфических признаков. Будем считать, что содержание вопроса отражается в ключевых терминах и отношениях между ними, т.е. в онтологической модели вопроса.

Под объёмом вопроса будем понимать множество адекватных ответов, удовлетворяющих заданному уровню качества.

Таким образом структура вопроса (СТВП) представляется как:

$$(1) \quad \text{СТВП} = \text{СТВП}(\text{ТВП}, \text{СВП}, \text{ОВП}).$$

Будем считать, что логическая структура ответа включает следующие составляющие:

- тему ответа (ТОТ);
- содержание ответа (СОТ);
- объём ответа (ООТ).

Под темой ответа будем понимать область поиска (область поиска – знания об объекте и предмете исследования, к которым относится вопрос)

Под содержанием ответа будем понимать предметы, явления, процессы, технологии, инструменты, теории, относящиеся к объекту и предмету исследования, связи между ними посредством фиксации общих и специфических признаков. Будем считать, что содержание ответа отражается в ключевых терминах и отношениях между ними, т.е. в онтологической модели ответа.

Под объёмом ответа будем понимать множество адекватных вопросу ответов, удовлетворяющие заданному уровню качества. Таким образом, структура ответа (СТОТ) представляется как:

$$(2) \quad \text{СТОТ} = \text{СТОТ}(\text{ТОТ}, \text{СОТ}, \text{ООТ}).$$

Формальная связь вопроса и ответа

Будем считать, что вопрос и ответ образуют целостную систему, если удовлетворяются следующие условия:

- А. Тема вопроса совпадает с темой ответа, т.е. $\text{ТВП} = \text{ТОТ}$ (знак «=» означает совпадение, согласованность, адекватность тем);
- В. Содержание ответа не больше содержания вопроса (число ключевых терминов в вопросе не меньше числа ключевых терминов в ответе и пересечение множества терминов вопроса и множества терминов ответа не пустое);
- С. Объём вопроса не меньше объёма ответа (множество ответов вопроса на предпосылках вопроса больше, чем множество ответов области поиска).

Из сказанного следуют ситуации:

1. Если содержание вопроса совпадает с содержанием ответа и объём вопроса совпадает с объёмом ответа, тогда ответ и вопрос удовлетворяют качеству и образуют полную систему.
2. Если содержание ответа является частью содержания вопроса и объём ответа соответствует части объёма вопроса, тогда ответ частично и качественно соответствует вопросу, и они образуют неполную систему.
3. Если содержание ответа не соответствует содержанию вопроса, тогда вопрос и ответ не образуют систему и не удовлетворяют качеству.

Методика анализа научных текстов

Данная методика применима только к научным текстам. Документ исследуется экспертом с точки зрения

1. Смыслового соответствия названия и содержания.
2. Набора фильтров:
 - 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:

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 частей текста, которые попали в фильтр **Идея автора** формулируется основная мысль – несколько предложений. Таким образом, мы получаем сжатое изложение (выжимку) текста.

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

1. Формулирование вопросов к выделенным информационным блокам.
2. Выделение ответов из анализируемого текста и ссылок на них.
3. 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.
4. Формирование графа Вопрос-Ответ-Реакция.

Логико-семантическая сеть «Вопрос-Ответ-Реакция»

Прототип семантической поисковой системы на основе ЛСС описан в работе [2]. Логико-семантическая сеть – множество вопросов, ответов и связей между ними, образующее целостную систему [1]. Целостность ЛСС определяется следующими свойствами:

- множество «Вопрос-Ответ-Реакция» относится к определённой теме предметной области;
- множество «Вопрос-Ответ-Реакция» иерархически упорядоченно по принципу «от общего к частному»;
- на нечётном уровне иерархии расположены вопросы, на чётном 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 определённому условию 'А' или 'В'. В случае удовлетворения условию 'А', например, имеет место конеч-

ная 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 $i+3$ -ем уровне, находятся вопросы, которые дополняют и уточняют ответы $i+2$ -го уровня.

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

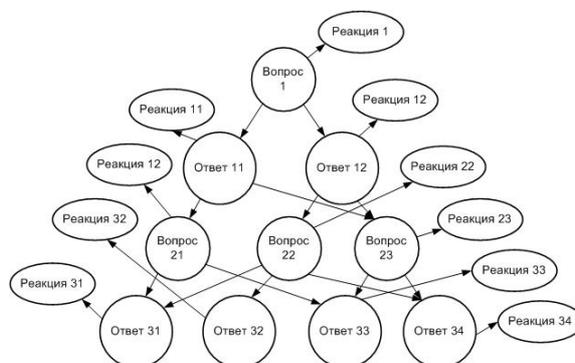


Рис. 1. Граф ЛСС «Вопрос-Ответ-Реакция»

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 [1].

Ввод реакций помогает пользователю понять, получил ли он релевантный ответ на свой вопрос. В качестве реакций может выступать дополнительная информация по теме вопроса и ответа, 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 предобласти вопроса (для осознания обстоятельств и причин возникновения вопроса и дальнейшего установления смыслового соответствия с областью ответа).
2. Реакции ответа – это описание области ответа (для осознания смысла вопроса и смысловой связи с ответом).

Процесс постановки вопроса, поиска ответа на него и формирования реакций вопроса и ответа – сложный многоступенчатый процесс, описанный подробно в работе [4].

В целом, логико-семантическая сеть «вопрос-ответ-реакция» полностью отвечает постулатам Хэмблина. Реакция для вопроса это описание области предпосылки вопроса. Реакция ответа – это описание области поиска ответа

Данная методика (см. п. 3.2, 3.3.) была применена для построения ЛСС научной статьи: Белая В. В., Семчуков П. Д., Стеценко М. С. РАЗРАБОТКА ПРОГРАММНОЙ ОБОЛОЧКИ ДЛЯ МУЛЬТИМЕДИЙНОГО ОБРАЗОВАТЕЛЬНОГО ПРОДУКТА // Электронный журнал «Системный анализ в науке и образовании», Международный университет природы, общества и человека «Дубна», кафедра САУ. – 2009, вып. 2. – <http://www.sanse.ru/archive/11>. Не представляется возможным представить полное описание процесса вместе с разметкой текста указанной статьи. Поэтому в Приложении 1

приведен только полученный граф ЛСС. Отражена сквозная нумерация вопросов и ответов, выполненная экспертом, работавшим с текстом. По свойствам ЛСС (п. 3.3) она может быть преобразована к виду (ij) , где i – номер уровня, j – номер вопроса/ответа данного уровня.

Формальное представление предметной области

Любая научно-практическая область знаний включает предмет исследования, который может быть представлен проблемным полем (перечнем проблемных вопросов), являющийся основой для научной и практической деятельности. Проблемные вопросы 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 определённая реакция. Ответы могут порождать в свою очередь вопросы. Таким образом, проблемный вопрос соотносится к определённой теме предметной области и раскрывается семантической структурой вопрос-ответ-реакция, которая, вообще говоря, является открытой (т.е. пополняемой, изменяемой во времени). Другими словами, знания, накопленные в предметной области, могут быть представлены открытым множеством логико-семантических сетей, упорядоченные по предметным темам. Задача предметной области может быть сформулирована в форме вопроса. 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, характеризуемую 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 отчётах, монографиях, статьях, учебных материалах, информационных фондах, справочниках, словарях и т.д. Будем считать, что информацию можно представить 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.

Обозначим через

$ПрОб$ – наименование предметной области,

$Тм_i$ – наименование i -ой темы предметной области,

$ЛСС_{ij}$ – наименование j -ой ЛСС i -ой темы.

Тогда предметная область представляется как

$$ПрОб = \bigcup_{i=1}^n Тм_i; Тм_i = \bigcup_{j=1}^{m_j} ЛСС_{ij},$$

причём $\bigcap Тм_i Тм_j \neq 0$ для $i \neq j$,

$$\bigcap ЛСС_{iji} ЛСС_{kr} \neq 0 \text{ для } i \neq k, \text{ и } j \neq r.$$

Представленные в разделе теоретические основы логико-семантических сетей «вопрос-ответ-реакция» позволяют описать предметную область множеством ЛСС, объединённые в тематические классы;

1. ЛСС предметной области могут служить основой для создания поисковых информационных систем;
2. Механизм поиска информации в предметной области на основе ЛСС может обеспечивать следующие режимы:
 - движение по ЛСС, управляемое пользователем;
 - поиск информации по заявке (вопросу).

Организация поиска на основе ЛСС в электронных информационных фондах

Семантические поисковые системы на основе ЛСС «Вопрос-Ответ-Реакция» могут иметь широкий спектр применимости, в том числе и в массивах электронных документов, электронных библиотек (ЭБ). Также актуально использование ЛСС в образовании для создания семантических учебников и пособий, т.е. таких учебных материалов, в которых раскрытие содержания материала происходит через логику его изложения и организацию учебного процесса. Для реализации подобной системы в рамках конкретного информационного наполнения необходимо

1. Построить множество ЛСС.
2. Реализовать механизм поиска информации в режиме ответа на вопрос.
3. Разработать навигационный механизм как движения вверх (от частного к общему), так и вниз (от общего к частному) по ЛСС.

Построение множества ЛСС возможно на нескольких уровнях (слоях). Первый слой (самый нижний) содержит ЛСС конкретного информационного ресурса (документа). Следующим слоем может быть ЛСС, построенная на базе функционирующего в данной ЭБ тематического рубрикатора и т.д. 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. 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.

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.

Создание, наполнение и сопровождение такой системы требует большой и серьезной работы, как технологической, так и организационной. Создание каталожной службы является трудоёмким 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 будет предоставлена новая возможность – получать ответы на вопросы, заданные на естественном языке.

Имеющиеся наработки

Разработки ведутся в НИЦ Управления знаниями и распределёнными вычислениями университета «Дубна». К настоящему времени:

1. Предложена методика и 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». Формы интерфейса АРМ разработчика ЛСС поисково-информационной консультативной системы (ПИКС) представлены на рис. 2. Раздел *Темы* позволяет разработчику просматривать, находить и корректировать темы предметной области. Раздел *ЛСС* предоставляет разработчику возможность просматривать, редактировать и создавать ЛСС.
3. Разработана методика поиска ответа на вопрос для корпуса документов.
4. Разработано ПО (в частном случае) поисковой машины.
5. Осуществлена ручная апробация методик на примерах.
6. 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».

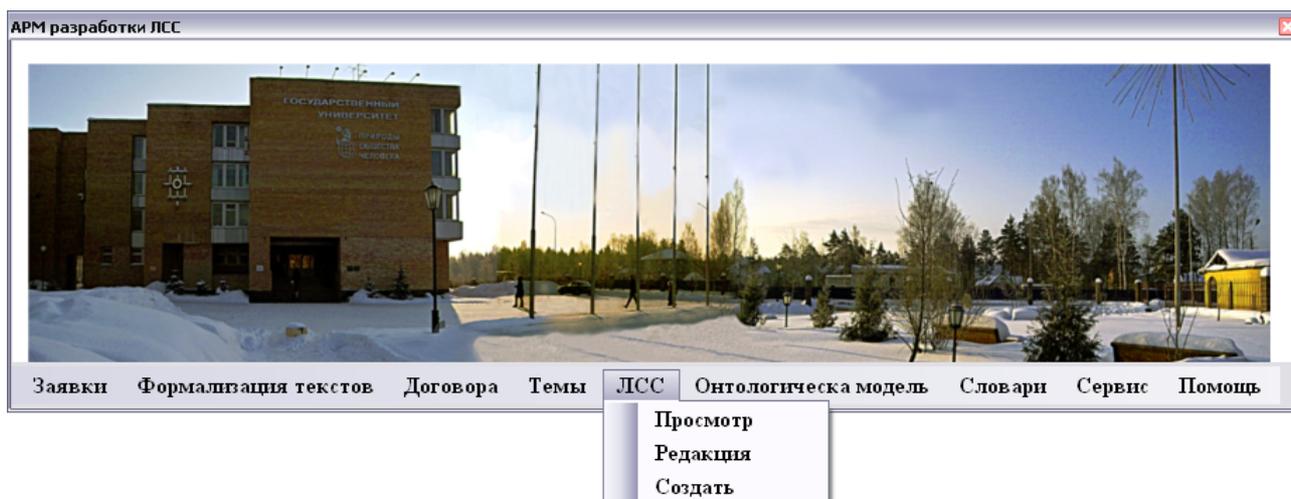


Рис. 2. Формы интерфейса АРМ разработчика ЛСС ПИКС

Заключение

Существующие проблемы и разработки QA-систем пересекаются с проблемами вопросно-ответных систем для фондов-корпусов научной информации. Авторами предлагается создание "каталожной службы" и её поддержки для фондов-корпусов, создание вопросно-ответного навигатора, обладающего особенностями, обеспечивающими такие качества, как возможность в процессе поиска ответов уточнение, углубление понимания смысла вопроса, возможность в процессе поиска ответа уточнения, углубления, расширения знаний или, как следствия, получения нового знания.

Основной проблемой создания предлагаемой вопросно-ответной системы является максимальная автоматизация процесса создания и поддержки "каталожной службы" фонда, возможность осуществления такого проекта.

Авторами апробированы в учебном процессе на различных дисциплинах методики анализа научной информации.

Состоятельность и актуальность излагаемого подхода на данном этапе исследований и разработок подтверждается экспресс-анализом существующего положения QA-систем.

Список литературы

1. Hamblin C. L. Questions // Australasian Journal of Philosophy. – 1958. – V. 36. – Pp. 159-158.
2. Аверьянов, Л. Я. Почему люди задают вопросы?. – М.: Социолог, 1993.
3. Белнап Н., Стил Т. Логика вопросов и ответов. – М.: Прогресс, 1981. – С .44.
4. Добрынин В.Н., Лобачева М.В. Прототип семантической поисковой системы на основе логико-семантической сети «ВОПРОС-ОТВЕТ-РЕАКЦИЯ» // Системный анализ в науке и образовании: электрон. науч. журнал. – Дубна, 2009. – №2. – [Электронный ресурс]. URL: <http://www.sanse.ru/archive/11> (дата обращения: 24.02.2010).

Приложение 1. Граф ЛСС «Вопрос-Ответ-Реакция» информационного ресурса

