

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ,
СТАТИСТИКА

УДК: 004.89:004.4

Обзорная статья

DOI: 10.35330/1991-6639-2024-26-6-175-187

EDN: IKXIFN

Методы организации вопросно-ответных систем

Г. В. Дорохина¹, А. Г. Полоус^{1,2}

¹Институт проблем искусственного интеллекта
283048, Россия, г. Донецк, ул. Артема, 118б

²Донецкий государственный университет
283001, Россия, г. Донецк, ул. Университетская, 24

Аннотация. На сегодняшний день актуальны проблемы человеко-машинного взаимодействия с помощью текста на естественном языке, в том числе с применением вопросно-ответных систем. Цель работы – анализ методов организации вопросно-ответных систем – достигнута путем рассмотрения типов вопросов и видов вопросно-ответных систем, описания методов построения вопросно-ответных систем. Встраивание вопросно-ответных систем в цифровые платформы позволит без существенного расширения персонала улучшать взаимодействие с клиентами, способствовать более эффективному решению их проблем, улучшать сервис. Это одновременно будет способствовать и росту дохода организаций-поставщиков, и повышению качества жизни потребителей товаров и услуг.

Ключевые слова: вопросно-ответная система, обработка естественно-языковых текстов, база данных, база знаний, онтология, машинное обучение, предобученные языковые модели

Поступила 25.11.2024, одобрена после рецензирования 09.12.2024, принята к публикации 11.12.2024

Для цитирования. Дорохина Г. В., Полоус А. Г. Методы организации вопросно-ответных систем // Известия Кабардино-Балкарского научного центра РАН. 2024. Т. 26. № 6. С. 175–187. DOI: 10.35330/1991-6639-2024-26-6-175-187

MSC: 68T40; 93C85

Review article

Methods of organizing question-answering systems

G.V. Dorokhina¹, A.G. Polous^{1,2}

¹Institute of Artificial Intelligence Problems
283048, Russia, Donetsk, 118b Artema street

²Donetsk State University
283001, Russia, Donetsk, 24 University street

Abstract. Today the issue of human-machine interaction via natural language text, including the use of question answering systems, is relevant. The purpose of the work is to analyze the methods of developing question answering systems. It was achieved by considering the types of questions and types of question answering systems, describing the methods of building question answering systems. Embedding question answering systems into digital platforms will allow improving customer interaction without significantly expanding staff, contributing to more effective solutions to their problems, and improving service. This will simultaneously contribute to the growth of income of supplier organizations and improve the life quality of goods and services consumers.

Keywords: question-answering system, natural language processing, database, knowledge base, ontology, machine learning, pre-trained language models

For citation. Dorokhina G.V., Polous A.G. Methods of organizing question-answering systems. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2024. Vol. 26. No. 6. Pp. 175–187. DOI: 10.35330/1991-6639-2024-26-6-175-187

ВВЕДЕНИЕ

Текст на естественном языке является ориентированным на человека средством хранения и передачи данных, информации, знаний. Вопросно-ответная система – это программный модуль, позволяющий человеку вести с машиной диалог на естественном языке, при котором пользователь задает вопросы программной системе, а программная система предоставляет ответы, формируемые в виде осмысленных предложений [1]. Работы по вопросно-ответным системам на естественном языке ведутся более пятидесяти лет. Большие массивы информации в виде текстов и текстовых документов на естественном языке привели к созданию поисковых систем, предоставляющих данные по текстовому запросу. Обычно эти данные представлены в виде ссылок на информационные ресурсы. Качество поиска значительно возросло благодаря распространению семантически структурированного содержимого. Появились вопросно-ответные системы на основе онтологий – формального описания структуры знаний в предметных областях. Автоматизация поиска и обработки данных с помощью вопросов и запросов на естественном языке расширяет возможности человека, упрощает и ускоряет работу с информацией, что указывает на актуальность темы исследований.

К настоящему времени разработан широкий спектр вопросно-ответных систем, которые существенно отличаются друг от друга по решаемым задачам, используемым методам и источникам данных. Достаточно много внимания уделяют вопросно-ответным системам, использующим семантический web, онтологии и графы знаний.

Цель работы состоит в анализе методов организации вопросно-ответных систем. Для ее достижения решены следующие задачи: 1) рассмотрены типы вопросов и виды вопросно-ответных систем; 2) описаны методы построения вопросно-ответных систем.

Практическая значимость работы. Встраивание вопросно-ответных систем в цифровые платформы [2] позволит без существенного расширения персонала улучшать взаимодействие с клиентами, способствовать более эффективному решению их проблем, улучшать сервис. Это одновременно будет способствовать и росту дохода организаций-поставщиков, и повышению качества жизни потребителей товаров и услуг. Примеры сфер, в которых эффективно применение вопросно-ответных систем: онлайн-службы доставки продуктов, покупки одежды и аксессуаров; интернет-банкинг; образование [3].

ТИПЫ ВОПРОСОВ И ВИДЫ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ

Вопрос является формой мысли человека, которая нацелена на восполнение и уточнение известной ему информации [4]. Он выполняет познавательную и коммуникативную функции. Вопрос содержит в себе информацию, называемую базисом. Содержание будущего ответа является предметом вопроса, а область возможных ответов – объемом вопроса. Вопросы делят на восполняющие (что, где, когда и т.п.) и уточняющие (правда ли, что ...). Вопросы бывают сложными и простыми. Простые содержатся в простых предложениях. Сложные образуются из двух и более простых вопросов с помощью логических союзов. Различают три вида сложных суждений: соединительные, разделительные, смешанные. Вопросы по существу называют релевантными, а вопросы, не относящиеся к делу, – нерелевантными. Закрытые вопросы допускают возможность исчерпывающего ответа. Тогда как открытые предполагают продолжение диалога, дальнейшее уточнение и углубление

знания. Определенные вопросы требуют однозначного ответа, а не определенные допускают неполный или неясный ответ. Прямые вопросы явно требуют установления неизвестного, а наводящие содержат такое требование в скрытой форме. Общие относятся к интересующему предмету в целом, а частные – к отдельным сторонам предмета вопроса [4].

Вопросно-ответные системы более удобны для пользователя по сравнению с поисковыми системами по ключевым словам в случаях, когда нужен краткий ответ на вопрос, а не список ссылок на документы, в которых этот ответ можно найти. Вопросно-ответные системы существенно различаются сложностью и предоставляемыми возможностями. Типичные классы вопросов, которые задают вопросно-ответным системам [5]: запрос по ключевым словам или на получение определения; вопросы относительно фактов («что» – вопрос, утверждение или опровержение); вопросы на понимание причинно-следственных связей (почему, как); вопросы, связанные с временными и пространственными рассуждениями; вопросы, связанные с рассуждениями на основе здравого смысла; интерактивные диалоги.

Вопросно-ответные системы могут быть организованы как системы с фактами или как системы со знаниями [6]. Системы с фактами хранят ответы на вопросы в базе данных. Для определения соответствия между вопросом пользователя и предопределенным в системе вопросом, а также между вопросом и ответом используют алгоритмы машинного обучения. Системы со знаниями в качестве источников используют тексты, данные интернета и базы знаний. Поиск ответа в текстовом корпусе зачастую выполняют с помощью линейных моделей или нейронных сетей, посредством которых моделируют контекст и семантическую связь между вопросом и текстом. Базы знаний содержат формализованные модели, которые описывают концептуальную структуру отдельных предметных областей и мира в целом, методы доступа к элементам моделей и способы выражения этих элементов в терминах естественного языка. Стандартом описания знаний в настоящее время являются онтологии. Вопросно-ответные системы на основе онтологий выбирают концепты, упомянутые в вопросе, и по связанным с ними данным либо данным, которые можно из них вывести, синтезируют ответ.

В работе [7] вопросно-ответные системы, организованные как интерактивные диалоги, предложено различать по «направлению», которое определяет ведущего в диалоге. В прямой вопросно-ответной системе человек задает вопросы, машина отвечает на них, а в инвертированной ведущим диалога является машина.

Закрытые вопросно-ответные системы работают в определенной предметной области, открытые – не зависят от предметной области. Есть вопросно-ответные системы, работающие с частными базами знаний. Различны и источники знаний, которые вопросно-ответные системы используют для получения ответа: структурированные – преобразуют запрос на естественном языке в запрос к базе данных; полуструктурированные – источниками выступают структурированные документы (анкеты, медицинские записи, платежные документы); тексты; семантические источники (онтологии, базы знаний) [5].

У вопросно-ответных систем присутствуют внутренние проблемы, характерные для классических поисковых систем, например: адаптируемость и работа с неоднозначностью. Для решения этих проблем применяют: лингвистический подход, статистический подход и подход, основанный на сопоставлении образцов.

Различается технологическая основа вопросно-ответных систем. В их основе может лежать: веб-классификация, извлечение сущностей [8, 9] и извлечение отношений [10–12], правила.

Появление генеративных языковых моделей (GPT от OpenAI, PaLM и Gemini от Google, Copilot от Microsoft, LLaMA от Meta AI, YandexGPT от Яндекс, GigaChat от Сбер, MTS AI

Chat от МТС) повлекло разделение вопросно-ответных систем на экстрактивные и генеративные [13]. Экстрактивные вопросно-ответные системы дают «короткий ответ на вопрос, часто в виде цитаты из поданного на вход набора документов» [13]. Генеративные системы синтезируют ответ из фрагментов текстов, выбранных в качестве соответствующих вопросу, или элементов базы знаний. Некоторые генеративные вопросно-ответные системы построены на основе больших предобученных генеративных языковых моделей, которые считаются способными к «сложным рассуждениям» и «позволяют отвечать более развернуто на более сложные вопросы» [13].

ОРГАНИЗАЦИЯ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ СО СТРУКТУРИРОВАННЫМИ ИСТОЧНИКАМИ

Известными представителями закрытых вопросно-ответных систем, которые используют структурированные источники данных, являются BASEBALL (1961) и LUNAR (1973) [5]. Они являются естественно-языковыми интерфейсами к экспертным системам по заданной предметной области.

Первые системы этого класса преобразуют вопросы на естественном языке в запросы к базам данных, используя правила и сравнение с образцом. Например, правило: «Если в вопросе за словом «столица» следует название страны, то система должна вывести название столицы, которая соответствует стране» позволит ответить на следующие вопросы и запросы:

Какой город является столицей Италии?

Выведи столицу Италии.

Что является столицей Италии?

Вы не могли бы назвать столицу Италии?

Такое поверхностное сопоставление с образцом часто приводит к неудаче, но также является неожиданно эффективной техникой при использовании в вопросах терминов, специфичных для предметной области, к которой относится источник знаний [5].

Недостатки систем этого класса связаны с их привязкой к конкретной базе данных. Их трудно модифицировать для использования с разными базами данных или перенести на другую предметную область. Настройка таких систем на предметную область требует существенных затрат времени и труда на синтез грамматик, которые отображают (mapping) вопросы на структуру базы данных. Их создают эксперты, используя явно сформулированные и неявные знания предметной области.

В следующем поколении систем использован формально-семантический подход: вместо жесткой привязки к структуре базы данных используют промежуточный язык представления, выражающий значение вопроса пользователя в терминах высокоуровневых концепций, независимо от структуры базы данных [5]. Это позволяет отделить языковые процессы, не зависящие от предметной области, от процесса отображения (mapping) вопроса на базу данных, который зависит от предметной области [5].

Развитием этого метода стал формально-семантический подход, в рамках которого отображают (mapping) предложения английского языка на выражения формальной семантической теории [5]. При таком подходе интерфейсы на естественном языке не зависят от предметной области. Настройка на предметную область состоит в настройке отображения выражений формальной семантической теории и терминов предметной области на структуру базы данных. При обработке запроса на естественном языке его сначала переводят в промежуточное логическое представление, а затем в SQL. Представителем такого подхода является система MASQUE / SQL (1993). В ней настройка на предметную область происходит полуавтоматически с помощью встроенного редактора, который помогает пользователю, используя иерархию is-a, описать типы сущностей, относящихся к базе данных, а затем указать ожидаемые слова вопроса и определить их значение в терминах логики предикатов, привязанной к таблицам (table) и отображениям (view) базы данных.

Система PRECISE (2003) сопоставляет вопрос с SQL-запросом путем определения классов семантически трактуемых вопросов. Вопрос преобразуют во множество пар атрибут/значение и маркеры отношения, указывающие на принадлежность либо к атрибутам, либо к значениям. Каждый атрибут в базе данных связан с определенным значением (что?, где? и т.д.), специальный словарь задает синонимы. Полученные в результате сопоставления элементы базы данных собираются в SQL-запрос. Если найдено более одного запроса, то правильный выбирает пользователь. Подход требует, чтобы лексические маркеры атрибутов и значений были различными. Вопросы с неизвестными словами являются семантически не интерпретируемыми – не обрабатываются. В экспериментах система правильно обработала 80% вопросов [5].

Более сложные запросы к базам данных по вопросу на естественном языке строят с помощью метода концептуальной разработки (Conceptual Authoring, 2007). Его применяли к большому медицинскому хранилищу. Логическое представление запроса строят с помощью естественно-языкового интерфейса, в котором вместо ввода текста все операции редактирования определяются в базовом логическом представлении, управляемом предопределенной онтологией. Это исключает проблемы с интерпретацией вопроса, но остается проблема настройки на предметную область [5].

Уменьшить сложность создания базовых грамматик для различных предметных областей призвана система C-PHRASE (2010). Ее графический веб-интерфейс позволяет построить семантическую грамматику с помощью серии именований, адаптаций и определенных операций. В этой системе запросы представлены в виде выражений в расширенной версии кортежного исчисления Кодда. Они могут быть напрямую сопоставлены с SQL-запросами или логическими выражениями первого порядка. Возможно также использование предикатов более высокого порядка [5].

ОРГАНИЗАЦИЯ ОТКРЫТЫХ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ ПО ТЕКСТАМ ДОКУМЕНТОВ

Большинство открытых вопросно-ответных систем по вопросу устанавливают тип ожидаемого ответа, например: имя (человека, организации), количество (выражающее сумму денег, расстояние, размер), дата. Исходя из типа ожидаемого ответа классы вопросов помещены в таксономиях. Разные типы вопросов обрабатывают в соответствии с разными стратегиями.

Некоторые открытые вопросно-ответные системы применяют поверхностное расширение на основе ключевых слов для поиска в документах интересных предложений. При этом документы отбирают по наличию слов, которые ссылаются на объекты того же типа, что и тип ожидаемого ответа. При поиске предложений их ранжируют, основываясь на синтаксических признаках, таких как порядок слов или сходство с запросом. Для поиска ответов, являющихся переформулировкой вопроса, могут использовать шаблоны. Для определения типа вопроса и сопоставления объектов с ожидаемым типом ответа системы могут обращаться к семантической паутине, например, к лексическим ресурсам WordNet, онтологии SUMO и другим. При построении ответа возможна более сложная синтаксическая, семантическая и контекстуальная обработка, включающая: распознавание именованных сущностей и отношений; разрешение кореференций; синтаксические чередования; устранение лексической, синтаксической и семантической неоднозначности; логические выводы и пространственно-временные рассуждения [5].

Вопросно-ответные системы, использующие в качестве источника текст, включают два этапа: 1) «определение семантического типа сущности, которую ищет вопрос» и 2) «определение дополнительных ограничений на объект ответа». В качестве ограничений используют ключевые слова, которые должны присутствовать (с точностью до синонимов или морфологических вариантов) в кандидате на ответ, и синтаксические или семантические

связи между сущностью-кандидатом на ответ и другими сущностями в вопросе. Такие системы обычно имеют встроенную иерархию типов вопросов на основе типов ожидаемых ответов, используют методы распознавания именованных сущностей и извлечения информации. Причем считается, что более 80 % вопросов в качестве ответа ожидает именованную сущность [5]. Рассмотрим особенности реализации этого метода на примерах систем LASSO (1999), FALCON (2000), DIMAP (2001).

Система LASSO может строить иерархию типов вопросов и ответов из обучающей выборки. По вопросу она автоматически определяет: тип вопроса (что, почему, кто, как, где); тип ответа (человек, расположение и т. д.); фокус вопроса – что нужно найти (многие вопросы неявно содержат ответ); ключевые слова. Распознавание именованных сущностей полезно, когда в ответе не может содержаться ключевое слово из вопроса, например: фокус вопроса «день недели».

Система FALCON в вопросе распознает именованные сущности и отображает на них семантические категории ответов. Идентифицированный таким образом тип ожидаемого ответа отображают на таксономию ответов, в которой категории верхнего уровня связаны с несколькими классами WordNet. То есть для фокуса вопроса и ключевых слов в иерархии WordNet выбирают наиболее общую категорию и ищут ответы, включая всех представителей синонимического ряда (synset) этой категории. Фокус вопроса – существительные, непосредственно связанные с вопросным словом, – также относят к ключевым словам.

Система DIMAP работает с шестью типами вопросов: время, местоположение, кто, что, размер и сколько. Она извлекает из документов «триплеты семантических отношений», которые хранит в базе данных и использует для ответа на вопрос. Триплет семантического отношения состоит из объекта дискурса, семантического отношения, характеризующего роль объекта в предложении, и управляющего слова, с которым объект находится в семантическом отношении. Для каждого предложения строится в среднем 9,8 триплетов. Аналогичную обработку выполняют для вопросов, генерируя в среднем 3,3 триплета на предложение, в том числе один триплет на каждый вопрос содержит несвязанную переменную, соответствующую типу вопроса.

ОРГАНИЗАЦИЯ ОТКРЫТЫХ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ ПО ТЕКСТАМ В СЕТИ ИНТЕРНЕТ

Вопросно-ответные системы на основе веб-классификации извлекают ответы на вопросы относительно фактов путем обращения к хранилищу документов в интернете. Они состоят из трех компонентов: 1) механизм формулирования запросов – преобразует запрос на естественном языке в запросы извлечения информации (IR-запросы); 2) поисковая система в интернете; 3) модуль извлечения ответов из документов, полученных от поисковой системы. В системах этого типа так же, как в системах поиска в текстах, для классификации типов ответа прибегают к WordNet или разметке именованных сущностей [5].

Система Mulder (2001) является вопросно-ответной системой относительно фактов в сети интернет, использующей результаты нескольких запросов к поисковому движку Google. Чтобы правильно сформировать поисковые запросы, система выполняет следующие действия. В вопросе она анализирует тип объекта глагола с помощью WordNet (например: числовой, именной, временной). Далее модуль формулирования запросов преобразует вопрос в набор запросов по ключевым словам, используя различные стратегии: извлечение наиболее важных ключевых слов, цитирование неполных предложений (именных групп), спряжение глагола или его расширение посредством WordNet. Ответ извлекают из возвращаемых Google фрагментов или резюме. При этом группируют похожие ответы и выбирают лучший методом голосования [5].

Система AskJeeves ищет вопрос в собственной базе данных и возвращает список близких вопросов, на которые известен ответ. Пользователь выбирает наиболее подходящий

вариант из списка, и его направляют на веб-страницы, где можно найти ответ. Данная система полагается на редакторов-людей, которые шаблонам вопросов ставят в соответствие страницы авторитетных сайтов.

Другой метод основан статистическом или семантическом сходстве. Например, FAQ Finder (1997) – это вопросно-ответная система на естественном языке, у которой источником данных выступают файлы часто задаваемых вопросов. Для сопоставления вопросов с ответами она использует две метрики: статистическое сходство и семантическое сходство. Статистическое сходство определяют с учетом предварительной статистической обработки большого количества специально подготовленных объемных документов предметной области. Семантическое сходство оценивают путем поиска в WordNet связей между вопросом пользователя и ответом. Данный метод не находит ответы на вопросы, слова которых явно не присутствуют в базе знаний. Для решения проблемы находят семантические эквиваленты, которые являются перефразировкой вопроса пользователя и на которые уже есть ответ в социальных сетях вопросов и ответов.

START (SynTactic Analysis using Reversible Transformations, 1993) – сетевая вопросно-ответная система на естественном языке. Отвечает на вопросы о фактах, дает определения из словарей по четырем предметным областям: география, искусство, наука и справочная информация, история и культура. Данные о фактах хранит в специальной «универсальной базе» Omnibase, имеющей модель «объект – свойство – значение». Каждому объекту этой базы сопоставлен источник данных. Источниками знаний являются локальное хранилище (база знаний) и интернет [14].

ВОПРОСНО-ОТВЕТНЫЕ СИСТЕМЫ НА ОСНОВЕ ПРЕДОБУЧЕННЫХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Влияние на все задачи обработки естественно-языковых текстов оказали генеративные предобученные языковые модели. Ограничением в их использовании является лимит по числу входных токенов, что не позволяет работать с большим количеством документов или с документом большого размера [13]. Для его преодоления в работе [13] рассмотрена индексация документов с последующим поисковым запросом и генерацией ответа на базе open source фреймворков Haystack, LlamaIndex. Рассмотрим схему ее работы. Вопрос поступает на вход «ретривера», который выбирает из хранилища индексированных документов наиболее релевантные вопросу фрагменты текста. Из этих фрагментов текста генерируют ответ. Предпочтительным способом статистической индексации документов считают индексацию с использованием меры TF-IDF или ее вариации BM25 [15]. Статистическая индексация имеет существенные недостатки, связанные с невозможностью учитывать порядок слов, контекст, замену слов синонимами. Эту проблему решают векторные семантические представления как способ в векторной форме описать семантический контекст слова с помощью так называемого эмбединга. Представления текста в виде эмбедингов строят большие языковые модели, такие как BERT и GPT. Индекс может быть построен либо как набор последовательных эмбедингов, соответствующих последовательным частям текста, либо как древовидная иерархическая структура, являющаяся представлением последовательной восходящей суммаризации текста. Существуют фреймворки с открытым исходным кодом, которые можно применять для создания вопросно-ответных систем на основе генеративных предобученных языковых моделей, например LlamaIndex и Haystack. При работе с ними можно использовать оптимизированные хранилища, такие как Weaviate, Pinecone, FAISS.

Организация вопросно-ответных систем на основе модели BERT, применяемых в известных виртуальных ассистентах, рассмотрена в работе [16]. Реализация мультимодальных визуальных вопросно-ответных систем приведена в работе [17].

ВОПРОСНО-ОТВЕТНЫЕ СИСТЕМЫ НА ОСНОВЕ ОНТОЛОГИЙ

Вопросно-ответные системы на основе онтологий детально рассмотрены в работе [18]. Охарактеризуем наиболее известные из них.

MOSES отвечает на вопросы на естественном языке, используя базу знаний, которая основана на фактах и построена на семантически структурированном содержимом Семантической паутины.

AQUA – закрытая доменная вопросно-ответная система, которая объединяет связанные с доменом документы и знания баз данных с помощью онтологии. При поиске ответа взаимодействует с человеком для уточнения входящего запроса.

JAVELIN – система для открытых доменов, в которой предусмотрено дополнение данными. JAVELIN имеет звездообразную архитектуру. В ней все подзадачи, связанные с ответом на вопрос (анализ вопроса, поиск информации, извлечение ответа и поиск ответа), рассматриваются как узлы, соединенные с центральным узлом. Получив вопрос от пользователя, центральный узел генерирует стратегию поиска ответа по классу вопроса.

PowerAQUA – это вопросно-ответная система на основе нескольких онтологий, которая ориентирована на запросы к нескольким ресурсам Семантического Веба и принимает пользовательские запросы на естественном языке.

AquaLog – это портативная вопросно-ответная система, основанная на онтологиях с ограниченным доменом. Источниками данных являются онтология и WordNet. Система принимает вопросы на естественном языке и возвращает ответы из семантических баз знаний, совместимых с онтологией. Она способна настраиваться на «жаргон» пользователя. Обучение происходит с использованием онтологических рассуждений для формирования общих шаблонов. По вопросу на естественном языке лингвистический компонент системы строит запросы в формате триплетов (Query Triples), а служба сходства отношений преобразует их в форму онто-триплетов (OntoTriples).

PANTO моделирует портативный интерфейс для преобразования данных на естественном языке в запрос SPARQL. Он основан на модели триплетов, в которой дерево разбора строят для модели данных с использованием стандартного парсера Stanford. К запросам на естественном языке применяют логические правила в форме отрицания, сравнительной и превосходной степени. Для отображения используют метрические алгоритмы WordNet и String. Дерево синтаксического анализа формирует промежуточное представление в виде Query Triples. Далее это представление преобразует в форму OntoTriples, которые представлены как сущности в онтологии. В конечном итоге OntoTriples интерпретируются как форма SPARQL. Таким образом, эта система помогает преодолеть разрыв между реальными пользователями и онтологией с помощью семантической сети, основанной на логической модели.

FREyA (Feedback, Refinement и Extended Vocabulary Aggregation) – вопросно-ответная система с возможностью уточнения вопроса и расширенной агрегацией словарного запаса. Сочетает синтаксический анализ со знаниями в онтологии для быстрого развертывания новой системы. В этой модели концепции онтологии изначально идентифицируются и проверяются. По вопросу пользователя строят дерево синтаксического разбора с помощью StanfordParser. Сопоставление пользовательского запроса с концепцией онтологии реализуется двумя способами: автоматически и с помощью пользователя. Уточнение вопроса у пользователя производят с помощью закодированных в онтологии знаний. В результате определяют тип ответа и генерируют запрос SPARQL. Модель ранжирования используется по типу сходства строк. Тип ответа этой системы представлен в виде графа. Для визуализации полученного графа используют JIT-библиотеку.

Querix – использует диалоги уточнений в случае неоднозначности вопроса. Включает в себя пользовательский интерфейс, менеджер онтологий, анализатор запросов, центр сопоставления, генератор запросов, компонент диалога и уровень доступа к онтологии; использует парсер Stanford для построения дерева синтаксического разбора вопроса. Запросы на естественном языке преобразует SPARQL-запрос, а с помощью WordNet определяет синонимы. Querix не использует семантические методы, основанные на логике.

ORAKEL используют для вычисления потенциальных ответов на запросы пользователя. Эта система формирует вопросы на основе слов пользователя и знаний, представленных в форме логики. Она переводит вопрос в форму запроса и отправляет его в модель обобщения для получения ответа. Затем механизм вывода оценивает запрос и возвращает ответ пользователю. ORAKEL настраивается через взаимодействие с пользователем в программе Frame Mapper, которая сопоставляет лингвистические структуры с отношениями в онтологии.

Система QACID основана на сборе запросов из заданного домена. Их анализируют и группируют в кластеры, далее вручную аннотируют с помощью запросов SPARQL. Каждый запрос рассматривается как мешок слов, сопоставляющий слова в естественно-языковых запросах с базой знаний с помощью метрик строковых расстояний. Генератор SPARQL заменяет онтологию экземплярами, сопоставленными для исходного вопроса на естественном языке. Он специфичен для домена – производительность зависит от типов вопросов, собранных в домене.

Систему SMART используют для управления информацией в семантической сети. Она включает в себя автоматический инструмент рассуждения и интеграцию с инструментами запроса. Система SMART способна выполнять запросы к семантическим базам данных с использованием языка запросов дескриптивной логики (DL), который отображается в форме запроса SPARQL. Одной из особенностей SMART является возможность проверки семантических запросов и преобразование запросов DL в формат SPARQL. Полученные результаты отображаются в формате RDF (триплетов «субъект – отношение – объект»), где каждый объект тройки представлен в виде унифицированного идентификатора ресурсов (URI). SMART хранит онтологии в файловой системе.

SWSE – система семантического веб-поиска, поддерживающая SPARQL с представлением RDF. Структура индекса состоит из полного индекса на четверках с функцией поиска по ключевым словам, основанной на инвертированной форме индекса. Компоненты обработки запросов распределены по нескольким машинам.

Система Cus является открытой и обладает большой базой знаний. Она существует уже около 30 лет и включает в себя множество онтологий. Кроме того, Cus имеет интерфейс на английском языке и может отвечать на вопросы.

ЗАКЛЮЧЕНИЕ

В данной работе изложены основные методы построения вопросно-ответных систем. Их анализ показал, что на практике применяют широкий спектр методов. Методы, реализованные в ранних вопросно-ответных системах, отвечали на вопросы относительно фактов в заданной предметной области и были построены как интерфейсы к базам данных. Задача получения человеком информации из технической системы с помощью вопроса на естественном языке сыграла не последнюю роль в появлении таких технологий, как семантический веб, машинное обучение, базы знаний на основе онтологий. А сами эти технологии позволили существенно ускорить, упростить и удешевить разработку вопросно-ответных систем в разных предметных областях.

СПИСОК ЛИТЕРАТУРЫ

1. *Лапишин В. А.* Вопросно-ответные системы: развитие и перспективы // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2012. № 6. С. 1–9. EDN: PDWJEV
2. *Пономарев В. В., Туманов В. Е.* Модель цифрового интерактивного документа в системном анализе предметной онтологии // Системный анализ в проектировании и управлении: сборник научных трудов XXV международной научной и учебно-практической конференции, 13–14 октября 2021 г.: в 3 частях. Ч. 3. DOI: 10.18720/SPBPU/2/id21-361
3. *Муромцев Д. И.* Модели и методы индивидуализации электронного обучения в контексте онтологического подхода // Онтология проектирования. 2020. Т. 10. № 1(35). С. 34–49. DOI: 10.18287/2223-9537-2020-10-1-34-49
4. *Веревищев И. И.* Логика: краткий теоретический курс: учебное пособие для студентов гуманитарных факультетов. Ульяновск: УлГТУ, 2009. 101 с.
5. *Lopez V. Is., Uren V., Sabou M., Motta E.* Question answering fit for the Semantic Web? // Article in Semantic Web. DOI: 10.3233/SW-2011-0041. UK: Milton Keynes, 2011. Pp. 14–47.
6. *Тюрина Д. А., Пальмов С. В.* Применение нейронных сетей в обработке естественного языка // Журнал прикладных исследований. 2023. № 7. С. 158–162. DOI: 10.47576/2949-1878_2023_7_158
7. *Черноморова Т. С., Воробьев С. П.* Классификация и принципы построения систем вопросно-ответного поиска // Бюллетень науки и практики. 2020. Т. 6. № 8. С. 145–156. DOI: 10.33619/2414-2948/57/12
8. *Дородных Н. О., Юрин А. Ю.* Подход к автоматизированному наполнению графов знаний сущностями на основе анализа таблиц // Онтология проектирования. 2022. Т. 12. № 3(45). С. 336–352. DOI: 10.18287/2223-9537-2022-12-3-336-352
9. *Видия А. В., Дородных Н. О., Юрин А. Ю.* Подход к созданию онтологий на основе электронных таблиц с произвольной структурой // Онтология проектирования. 2021. Т. 11. № 2(40). С. 212–226. DOI: 10.18287/2223-9537-2021-11-2-212-226
10. *Ломов П. А., Никонорова М. Л., Шишаев М. Г.* Извлечение отношений из NER-размеченных предложений для обучения онтологии // Труды Кольского научного центра РАН. Серия: Технические науки. 2022. Т. 13. № 2. С. 23–30. DOI: 10.37614/2949-1215.2022.13.2.002
11. *Тихобаева О. Ю., Бручес Е. П., Батура Т. В.* Извлечение семантических отношений из текстов научных статей // Вестник НГУ. Серия: Информационные технологии. Т. 20. № 3. С. 65–76. DOI: 10.25205/1818-7900-2022-20-3-65-76
12. *Мезенцева А. А., Бручес Е. П., Батура Т. В.* Методы и подходы к автоматическому связыванию сущностей на русском языке // Труды ИСП РАН. 2022. Т. 34. № 4. С. 187–200. DOI: 10.15514/ISPRAS-2022-34(4)-13
13. *Голиков А. А., Акимов Д. А., Романовский М. С., Траценков С. В.* Аспекты создания корпоративной вопросно-ответной системы с использованием генеративных предобученных языковых моделей // Litera. 2023. № 12. С. 190–205. DOI: 10.25136/2409-8698.2023.12.69353
14. *Деревянко Д. В., Пальчунов Д. Е.* Формальные методы разработки вопросно-ответной системы на естественном языке // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2014. Т. 12, № 3. С. 34–47. EDN: RKCZKL
15. *Милкова М. А., Неволин И. В., Пигорев Д. П.* Современные методы извлечения ключевой информации из нормативных документов // Экономическая наука современной России. 2021. № 2(93). DOI: 10.33293/1609-1442-2021-2(93)-101-114

16. Гуляев П. А., Елистратова Е. А., Коновалов В. П. и др. Отслеживание состояния целеориентированного диалога на основе БЕРТ // Труды МФТИ. 2021. Т. 13. № 3(51). С. 48–61. DOI: 10.53815/20726759_2021_13_3_48
17. Фаворская М. Н., Андреев В. В. Адаптивное блочное тензорное разложение в визуальных вопросно-ответных системах // Программные продукты и системы. 2021. Т. 34. № 1. С. 164–171. DOI: 10.15827/0236-235X.133.164-171
18. Bali V., Verma A. A study on components, benchmark criteria and techniques used in ontology-based question answering systems // IJISAE, 2022, 10(1s). India: Chandigarh, 2022. Pp. 9–17.

REFERENCES

1. Lapshin V.A. Question-answering systems: development and prospects. *Automatic Documentation and Mathematical Linguistics*. 2012. No. 6. Pp. 1–9. EDN: PDWJEV. (In Russian)
2. Ponomarev V.V., Tumanov V.E. Model of a digital interactive document in the systems analysis of subject ontology. *Sistemnyy analiz v proyektirovanii i upravlenii* [Systems analysis in design and management]: collection of scientific papers of the XXV International scientific and educational-practical conference, October 13–14, 2021, in 3 parts. Part 3. DOI: 10.18720/SPBPU/2/id21-361. (In Russian)
3. Muromtsev D.I. Models and methods of individualization of e-learning in the context of the ontological approach. *Ontology of Designing*. 2020. Vol. 10. No. 1(35). Pp. 34–49. DOI: 10.18287/2223-9537-2020-10-1-34-49. (In Russian)
4. Verevichev I.I. *Logika: kratkiy teoreticheskiy kurs: uchebnoye posobiye dlya studentov gumanitarnykh fakul'tetov* [Logic: a short theoretical course: a tutorial for students of humanities faculties]. Ulyanovsk: UIGTU, 2009. 101 p. (In Russian)
5. Lopez V.Is, Uren V., Sabou M., Motta E. Question answering fit for the Semantic Web? *Article in Semantic Web*. DOI: 10.3233/SW-2011-0041. UK: Milton Keynes, 2011. Pp. 14–47.
6. Tyurina D.A., Palmov S.V. Application of neural networks in natural language processing. *Journal of Applied Research*. 2023. No. 7. Pp. 158–162. DOI: 10.47576/2949-1878_2023_7_158. (In Russian)
7. Chernomorova T.S., Vorobyov S.P. Classification and principles of constructing question-answer search systems. *Bulletin of Science and Practice*. 2020. Vol. 6. No. 8. Pp. 145–156. DOI: 10.33619/2414-2948/57/12. (In Russian)
8. Dorodnykh N.O., Yurin A.Yu. An approach to automated filling of knowledge graphs with entities based on table analysis. *Ontology of Designing*. 2022. Vol. 12. No. 3(45). Pp. 336–352. DOI: 10.18287/2223-9537-2022-12-3-336-352. (In Russian)
9. Vidiya A.V., Dorodnykh N.O., Yurin A.Yu. An approach to creating ontologies based on spreadsheets with an arbitrary structure. *Ontology of Designing*. 2021. Vol. 11. No. 2(40). Pp. 212–226. DOI: 10.18287/2223-9537-2021-11-2-212-226. (In Russian)
10. Lomov P.A., Nikonorova M.L., Shishaev M.G. Extracting relations from NER-labeled sentences for ontology training. *Trudy Kol'skogo nauchnogo tsentra RAN. Seriya: Tekhnicheskiye nauki* [Proceedings of the Kola Science Center of the Russian Academy of Sciences. Series: Technical sciences]. 2022. Vol. 13. No. 2. Pp. 23–30. DOI: 10.37614/2949-1215.2022.13.2.002. (In Russian)

11. Tikhobaeva O.Yu., Bruches E.P., Batura T.V. Extracting semantic relations from the texts of scientific articles. *Vestnik NGU. Seriya: Informatsionnyye tekhnologii* [Bulletin of NSU. Series: Information technologies]. Vol. 20. No. 3. Pp. 65–76. DOI 10.25205/1818-7900-2022-20-3-65-76. (In Russian)
12. Mezentseva A.A., Bruches E.P., Batura T.V. Methods and approaches to automatic linking of entities in Russian. *Proceedings of The Institute for System Programming of the RAS*. 2022. Vol. 34. No. 4. Pp. 187–200. DOI: 10.15514/ISPRAS-2022-34(4)-13. (In Russian)
13. Golikov A.A., Akimov D.A., Romanovsky M.S., Trashchenkov S.V. Aspects of creating a corporate question-answering system using generative pre-trained language models. *Litera*. 2023. No. 12. Pp. 190–205. DOI: 10.25136/2409-8698.2023.12.69353. (In Russian)
14. Derevyanko D.V., Palchunov D.E. Formal methods for developing a question-answering system in natural language. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informatsionnyye tekhnologii* [Bulletin of the Novosibirsk State University. Series: Information Technology]. 2014. Vol. 12. No. 3. Pp. 34–47. EDN: RKCZKL. (In Russian)
15. Milkova M.A., Nevolin I.V., Pigorev D.P. Modern methods for extracting key information from regulatory documents. *Economics of Contemporary Russia*. 2021. No. 2(93). DOI: 10.33293/1609-1442-2021-2(93)-101-114. (In Russian)
16. Gulyaev P.A., Elistratova E.A., Konovalov V.P. et al. Tracking the state of a goal-oriented dialogue based on BERT. *Proceedings of Moscow Institute of Physics and Technology*. 2021. Vol. 13. No. 3(51). Pp. 48–61. DOI: 10.53815/20726759_2021_13_3_48. (In Russian)
17. Favorskaya M.N., Andreev V.V. Adaptive block tensor decomposition in visual question-answering systems. *Software & Systems*. 2021. Vol. 34. No. 1. Pp. 164–171. DOI: 10.15827/0236-235X.133.164-171. (In Russian)
18. Bali V., Verma A. A study on components, benchmark criteria and techniques used in ontology-based question answering systems. *IJISAE*, 2022. Vol. 10. No. 1s(2002). India: Chandigarh, 2022. Pp. 9–17.

Вклад авторов:

А. Г. Полоус – подбор материала по тематике исследований, типизация вопросно-ответных систем и описание вопросно-ответных систем на основе онтологий.

Г. В. Дорохина – систематизация материала, изложение методов построения вопросно-ответных систем, оформление работы.

Authors' contribution:

A.G. Polous – selection of material on the research topic, typification of question-answering systems and description of question-answering systems based on ontologies.

G.V. Dorokhina – systematization of material, presentation of methods for constructing question-answering systems, design of the work.

Финансирование. Исследование осуществлено в рамках фундаментальных исследований ФГБНУ «Институт проблем искусственного интеллекта» по теме «Исследование и разработка методов обработки данных и естественно-языковых текстов с применением онтологий», Рег. № НИОКТР 123092600030-4.

Funding. The study was carried out within the framework of fundamental research of the Federal State Budgetary Scientific Institution Institute of Artificial Intelligence Problems on the topic Research and development of methods for processing data and natural language texts using ontologies, Reg. No. NIOKTR 123092600030-4.

Информация об авторах

Дорохина Галина Владимировна, науч. сотр., отдел системного анализа и интеллектуальных интерфейсов, Институт проблем искусственного интеллекта;

283048, Россия, г. Донецк, ул. Артема, 118б;

SPIN-код: 6522-7758

Полоус Андрей Геннадиевич, техник-программист, отдел системного анализа и интеллектуальных интерфейсов, Институт проблем искусственного интеллекта;

283048, Россия, г. Донецк, ул. Артема, 118б;

студент физико-технического факультета, Донецкий государственный университет;

283001, Россия, г. Донецк, ул. Университетская, 24;

SPIN-код: 6966-7988

Information about the authors

Galina V. Dorokhina, Researcher, Department of Systems Analysis and Intelligent Interfaces, Institute of Artificial Intelligence Problems;

283048, Russia, Donetsk, 118b Artema street;

SPIN-code: 6522-7758

Andrey G. Polous, Software Engineer, Department of Systems Analysis and Intelligent Interfaces, Institute of Artificial Intelligence Problems;

283048, Russia, Donetsk, 118b Artema street;

student, Faculty of Physics and Technology, Donetsk State University;

283001, Russia, Donetsk, 24 University street;

SPIN-code: 6966-7988