

**Глотов Илья Анатольевич,**  
магистрант  
МГТУ им. Н. Э. Баумана

**Рудаков Игорь Владимирович,**  
к.т.н., доцент  
МГТУ им. Н. Э. Баумана

**Волкова Лилия Леонидовна,**  
старший преподаватель  
МГТУ им. Н. Э. Баумана

## **МОДИФИЦИРОВАННЫЙ МЕТОД ФОРМИРОВАНИЯ ОТВЕТОВ НА ВОПРОСЫ НА ОГРАНИЧЕННОМ ЕСТЕСТВЕННОМ ЯЗЫКЕ ДЛЯ РОБОТА Ф-2 НА ОСНОВЕ АНАЛИЗА ДЕРЕВЬЕВ СИНТАКСИЧЕСКИХ ЗАВИСИМОСТЕЙ**

**Аннотация.** Разработан модифицированный метод формирования ответов на вопросы на ограниченном естественном языке для робота Ф-2 на основе анализа деревьев синтаксических зависимостей, сформулированы основные его особенности.

**Ключевые слова:** Вопрос-ответные системы, диалоговые системы, человеко-машинное взаимодействие, синтаксико-семантический анализ, робот Ф-2.

**Введение.** Одним из важнейших направлений автоматической обработки естественно-языковых данных является разработка и совершенствование интеллектуальных диалоговых систем и их упрощенных версий – чат-ботов. Эти системы стали все чаще применяться в коммерческих проектах, где они используются в общении с клиентами для помощи в покупке товаров, технической поддержки, навигации по сайтам и т. д. [1]. Диалоговые системы используются в качестве интеллектуальных модулей коммуникации социальных роботов [2].

**Постановка задачи.** Предлагаемый модифицированный метод является гибридным: он сочетает элементы экспертной системы (правила формирования ответов для каждого класса вопросов), поиска по аннотированному тексту (использование семантически аннотированных деревьев зависимостей) и метапоисковой системы.

Входные данные: текст вопроса на ограниченном русском естественном языке; база знаний робота Ф-2.

Выходные данные: текст ответа на вопрос на русском языке.

Метод ответов на вопросы имеет следующие особенности и ограничения.

1. Метод поддерживает десять классов вопросов: подтверждение, дополнение, пример, свойство, квантификация, определение, сравнение, координация действий, переспрос, пристрастие.
2. Метод применяется для робота Ф-2 и работает с его базой знаний.
3. Метод поддерживает ограниченный естественный язык – русский.
4. Классификация вопросов выполняется нейронной сетью.
5. Метод поддерживает кэширование пар «вопрос–ответ».
6. Метод учитывает контекст диалога с разрешением анафорических ссылок.
7. Производится оценка уверенности ответа.
8. В методе предусмотрен резервный поиск.
9. Ответ на вопрос не верифицируется.
10. Не производится проверка источника.
11. Не производится проверка достоверности факта, извлеченного из базы знаний.
12. Не производится оценка субъективности факта, извлеченного из базы знаний.
13. Ответ на вопрос ограничен множеством фактов в базе знаний.
14. Не производится проверка культурности и уместности вопроса и ответа на вопрос.



Разработанный модифицированный метод не является экспертной системой. Этот метод позволит отвечать на вопросы исключительно по имеющейся базе фактов, но с учетом контекста. Если некоторые ответы могут быть неточными с точки зрения гипотетической экспертной системы, но могут быть допустимы с точки зрения беседы на отвлеченные темы, подобные ответы будут считаться приемлемыми [3].

Для робота Ф-2 важно, чтобы время формирования ответов на вопросы было как можно меньше, чтобы взаимодействие с собеседником было естественным. Приемлемым значением является 5–10 секунд, допускаются нечастые отклонения до 15 секунд, которые робот может компенсировать воспроизведением поведенческого кадра «задумчивость».

**Ключевые этапы метода.** Ключевыми этапами метода являются: классификация вопроса, переформулирование вопроса, построение семантически аннотированного дерева синтаксических зависимостей для переформулированного вопроса, формирование SQL-запросов к базе знаний, исполнение SQL-запросов, при отсутствии результатов – резервный поиск, агрегация результатов и формирование ответа.

**Разработанные модификации.** Исходный метод поддерживал семь классов вопросов. В модифицированном методе добавлена поддержка трех дополнительных классов: переспрос, пристрастие, координация действий.

В исходном методе классификация вопросов выполнялась по ключевым словам. В модифицированном методе классификация выполняется нейронной сетью. Для представления вопроса в виде числового вектора используется комбинация двух типов признаков.

1. Распределенные представления слов (Word2Vec) [4]. Каждое слово вопроса представляется вещественнозначным вектором, полученным из предобученной модели Word2Vec для русского языка. Вопрос представляется последовательностью векторов фиксированной длины (8 слов): короткие вопросы дополняются нулевыми векторами, длинные – усекаются.

2. Бинарные признаки. Формируется вектор из 40 бинарных признаков, отражающих наличие ключевых маркеров класса: вопросительных слов («что», «кто», «сколько», «какой», «зачем», «почему»), глагольных форм, числительных, сравнительных конструкций и т. д.

Классификатор реализован как полносвязная нейронная сеть (multilayer perceptron), обученная на размеченном наборе вопросов. На вход подается конкатенация матрицы Word2Vec-представлений и вектора бинарных признаков. На выходе – вероятности принадлежности к каждому из классов (softmax). Класс с наибольшей вероятностью выбирается как результат классификации. Модель обучена с использованием библиотеки Keras [5] и сохранена в формате HDF5. Предобработка и извлечение признаков выполняются при помощи библиотек rymorphy2 [6] (морфологический анализ) и gensim (Word2Vec).

Для снижения времени формирования ответов на повторяющиеся вопросы реализовано кеширование пар «вопрос–ответ». Кеширование реализовано в виде декоратора для каждого из методов формирования ответов. При поступлении вопроса сначала проверяется наличие ответа в кеше. Если ответ найден, он возвращается немедленно без обращения к базе данных. Если ответ не найден, выполняется полный цикл формирования ответа, после чего пара «вопрос–ответ» сохраняется в кеше.

При отсутствии результата основного SQL-поиска по леммам модифицированный метод использует каскадный резервный поиск (fallback) в той же базе знаний. Описание подхода следующее.

1. Основной поиск. Система формирует набор лемм из переформулированного вопроса и контекстного окна диалога, затем выполняет SQL-поиск фактов по семантическим валентностям.

2. Упорядочивание лемм. Перед fallback леммы сортируются по информативности: содержательные леммы обрабатываются раньше служебных.

3. Резервные шаги. Если основной поиск не дал результата, выполняются резервные стратегии: эвристика уникальной пары лемм и пошаговое ослабление условия поиска (drop-tail) до нахождения кандидатов.



4. Выбор факта. Из найденных кандидатов выбирается наиболее релевантный факт с учетом покрытия лемм и дополнительных сигналов сценариев для класса «определение».

В исходном методе каждый вопрос обрабатывается изолированно. В модифицированном методе реализован механизм учета контекста диалога с разрешением анафоры для корректной обработки последовательных реплик.

В модифицированном методе каждый ответ сопровождается численной оценкой уверенности – вещественным числом в диапазоне  $[0, 1]$ . Оценка уверенности вычисляется на основе следующих факторов.

1. Путь retrieval. Учитывается источник ответа в каскаде поиска: PRIMARY, FALLBACK или NONE.

2. Покрытие лемм. Уверенность зависит от доли лемм вопроса, подтвержденных найденным фактом.

3. Сигналы сценариев. Для класса «определение» дополнительно учитываются признаки из сценарных данных (например, proximity, best).

**Заключение.** Описаны основные особенности разработанного метода, представлены шесть модификаций: расширение набора классов вопросов до десяти, нейросетевая классификация вопросов, кеширование пар «вопрос–ответ», резервный поиск, учет контекста диалога с разрешением анафоры и оценка уверенности ответа

#### **Список литературы:**

1. Balakrishnan J. Conversational commerce: entering the next stage of AI – powered digital assistants // Annals of Operations Research. – 2021. – № 333. – С. 653 – 687. – DOI: 10.1007/s10479-021-04049-5.

2. Daqar A. The Role of Artificial Intelligence on Enhancing Customer Experience //International Review of Management and Marketing. – 2019. – С. 22 – 31.

3. Minsky M. A framework for representing knowledge // Winston, P.H. (ed.) The Psychology of Computer Vision. – New York: McGraw-Hill, 1975. – С. 211–277.

4. Efficient Estimation of Word Representations in Vector Space / Т. Миколов [и др.] // arXiv preprint arXiv:1301.3781. – 2013.

5. Документация Keras [Электронный ресурс]. – Режим доступа: <https://keras.io/> (дата обращения: 28.05.2026).

6. Документация библиотеки морфологического анализа pymorphy2 [Электронный ресурс]. – Режим доступа: <https://pymorphy2.readthedocs.io/en/stable/> (дата обращения: 28.05.2026)

