

МЕТОД ДЛЯ ПРОВЕРКИ ЗАПРОСОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ К СЕМАНТИЧЕСКОЙ МОДЕЛИ

Кушнеревич П.М.

Белорусский государственный университет информатики и радиоэлектроники

г. Минск, Республика Беларусь

Герман Ю.О. – канд. тех. наук, доцент

Работа посвящена разработке метода для проверки запросов на естественном языке к семантической модели. Предложенный метод включает лексический, синтаксический и семантический анализ текста для формирования запросов. Описаны основные шаги алгоритма для обработки естественно-языковых запросов, включая предварительную обработку, лексический и синтаксический анализ, семантическую интерпретацию и исполнение запроса. Приведены примеры использования метода в различных приложениях, в том числе в электронной коммерции и чат-ботах. Работа представляет интерес для специалистов в области искусственного интеллекта в части обработки естественного языка.

В современном мире существует необходимость взаимодействия человека с компьютерными системами на естественном языке. Для этого требуется разработка формального аппарата для проверки корректности запросов к семантической модели.

В рамках данной работы предложен метод проверки запросов на естественном языке для взаимодействия с семантическими моделями на основе искусственного интеллекта. Метод включает лексический, синтаксический и семантический анализ текста для формирования запросов.

Для реализации данного подхода предложен следующий алгоритм.

Шаг 1. Предварительная обработка запроса. На данном этапе происходит удаление стоп-слов, то есть слов, которые не несут смысловой нагрузки, таких как "и", "или", "но" и т.д. Кроме того, выполняется стемминг – процесс нахождения основы слова, чтобы можно было свести различные формы слова к единой форме. Например, для запроса "Какие магазины в Москве продают яблоки?" будут удалены стоп-слова "какие", "в" и "продают", а слово "магазины" будет приведено к основе "магазин". Для выполнения этого шага используются инструменты обработки естественного языка, такие как NLTK (Natural Language Toolkit) и SpaCy.

Шаг 2. Лексический и синтаксический анализ. Лексический анализ заключается в разбиении запроса на лексемы – минимальные смысловые единицы, такие как слова и знаки препинания [1].

Синтаксический анализ заключается в определении структуры запроса и его синтаксических свойств. На данном этапе происходит построение дерева синтаксического разбора, которое показывает зависимости между словами в запросе и их ролью в предложении. Для выполнения этого шага используются инструменты обработки естественного языка, такие как Stanford Parser и SyntaxNet.

Шаг 3. Семантическая интерпретация. На этом этапе происходит определение смысла запроса, то есть его семантики. Для этого используется семантическая модель, которая содержит знания о мире и его объектах [2]. Для выполнения этого шага используются инструменты обработки естественного языка, такие как ConceptNet и WordNet.

Шаг 4. Исполнение запроса. На данном этапе выполняется поиск соответствующих объектов в базе знаний и возвращается результат пользователю.

Примеры использования данного подхода демонстрируют его широкие возможности и гибкость в различных сферах. Одной из них является электронная коммерция.

В рамках данной сферы метод может быть использован для проверки запросов на естественном языке, которые позволяют потенциальным клиентам находить необходимые товары и услуги с большей эффективностью.

Например, пользователь может задать запрос на естественном языке, такой как "найти черные кроссовки размером 42", и система сможет автоматически обработать этот запрос, произвести поиск в базе данных и вывести пользователю результаты, соответствующие его запросу.

Другим направлением использования данного метода являются чат-боты. Они могут помочь пользователям быстро и удобно получать необходимую информацию.

59-я научная конференция аспирантов, магистрантов и студентов БГУИР

Например, чат-бот для покупок может помочь пользователю найти товар, основываясь на запросе на естественном языке, таком как "хочу купить куртку", и предложить пользователю соответствующие варианты.

В целом предложенный формальный аппарат имеет широкий потенциал применения в различных областях, связанных с взаимодействием человека с компьютерной системой.

Список использованных источников:

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. *Attention is All You Need* // *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*. Long Beach, CA: Curran Associates, 2017. P. 112-113.
2. Hirschberg, J. & Manning, C.D. *Advances in Natural Language Processing* // *Science*. 2015. Vol. 349, Number 6245. P. 261-266.