Использование ИИ для разработки вопросно-ответной системы на казахском языке

Муканова Асель Сериковна
Сотрудник
Евразийский национальный университет имени Л. Н. Гумилёва, Нур-Султан
E-mail: asiserikovna@gmail.com
Соавторы: Бурибаева А.К.

В настоящее время объем информации в сети Интернет растет очень быстро и найти нужную информацию становится все труднее. Поиск по ключевым словам не дает результатов, адекватных смыслу искомой информации. Поэтому использование искусственного интеллекта в разработке интеллектуальной вопросно-ответной системы на казахском языке, основанных на представлении, обработке и извлечении знаний является весьма актуальной задачей, поскольку именно в такой системе устанавливаются лингвистические и семантические связи между текстами вопроса и ответа.

Первые прототипы вопросно- ответных систем появились в 60-х годах прошлого века. Самые известные из них BASEBALL, LUNAR, ELIZA, START. Другие интересные системы, претендующие на статус открытых, являются системы Cyc, OpenEphyra и PIQUANT. В следующих работах [4, 3, 5, 1] описаны современные вопросно-ответные системы на основе онтологической модели предметной области. Также исследователи семантической паутины (Semantic Web) предложили использовать технологии семантического поиска для английского языка, таких как OntoSearch, Semantic Wikis [6], Semantic Portals [7], мультиагентная P2P семантическая система маршрутов (запросов) на основе онтологии, вопросно-ответные системы на основе онтологии [2].

Искусственный интеллект играет ключевую роль в разработке и улучшении вопросно-ответных систем. Интеграция искусственного интеллекта (ИИ) в вопросно-ответные системы привлекла значительное внимание в различных междисциплинарных областях.

Для реализации интеллектуальной вопросно-ответной системы на казахском языке будут использованы следующие методы: методы построения онтологических моделей предметных областей, методы морфологического анализа и синтеза словоформ, методы синтаксического и семантического анализа и синтеза предложений, методы машинного обучения.

Одним из основных параметров анализа текста для понимания смысла предложения является определение свойств слов в тексте и связей между ними. Мы использовали онтологическую модель для представления знаний предметной области. С помощью онтологической модели можно отражать понятия предметной области и семантические отношения между ними. Преобразовывая слова текста в эти понятия и отношения можно сделать разные запросы и получить ответы соответствующие по смыслу.

На рисунке 1 представлен фрагмент запросов.

Кроме онтологической модели предметной модели создан набор данных на казахском в формате SquAD и текстовой корпус которые будут использоваться при разработке вопросно ответной системы на казахском языке. Следующим этапом работы является создание алгоритма морфологического и синтаксического анализа текста с использованием машинного обучения.

Построенная онтологическая модель, набор данных на казахском в формате SquAD и текстовой корпус могут служить научной основой для многих научных исследований и информационных систем на казахском языке. Они будут востребованы при создании программ машинного перевода с казахского и на казахский язык, систем принятия решений и других интеллектуальных систем. Компьютеризация казахского языка способствует его

сохранению, развитию и интеграции в современные технологические и социокультурные контексты.

Тезисы представлены сотрудником Международного университета Астана

Данное исследование финансировалось Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант № AP1957792, 2023-2025)

Источники и литература

1) Kalaivani S., Duraiswamy K. Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment. Journal of Computer Science, Vol. 8, No. 9, 2012
2) Kotis, K.."Semantic Web Search: perspectives and key technologies", Encyclopedia of Data Warehousing and Mining, 2nd Edition: IDEA GROUP - REFERENCE, 2008
3) Maria Vargas-Vera, Miltiadis D. Lytras. AQUA: A Closed-Domain Question Answering System. IS Management 27(3), 2010. - pp. 217-225
4) Pengfei Zhang ; Chunhua Wu ; Cong Wang ; Xiaohong Huang. Personalized Question Answering System Based on Ontology and Semantic Web. 2006 IEEE International Conference on Industrial Informatics, 2006.- pp. 1046-1051
5) Qinglin Guo, Ming Zhang. Question answering system based on ontology and semantic web. Proceedings of the 3rd international conference on Rough sets and knowledge technology, 2008. - pp. 652-659
6) Völkel, M., Krötzsch, M., Vrandecic, D., Haller, H., and Studer, R. 2006. Semantic Wikipedia. In Proceedings of the 15th International Conference on World Wide Web. WWW '06. ACM Press, New York, NY, pp. 585-594
7) Zhang, L., Yu, Y., & Yang, Y. An enhanced Model for Searching in Semantic Portals. WWW 2005. In Proceedings of the 14th international conference on World Wide Web. WWW '05. ACM Press, New York, NY, pp. 453-462

Иллюстрации

Виды связи	Зависимое слово dep	Главное слово head	Subject S	Predicate P	Object O	
KabEs	Adj	head	head	? (hasProperty)	dep	қызыл алма (red apple)
KabEs	Num	head	head	? (hasNumber)	dep	бес студент (five students)
KabEs	Num	head	head	? (hasOrder)	dep	алтыншы орын (sixth place)
MatEs	dep	head	dep	head	?	Серіктің қызы Асель (Serik's daughter **Assel**)
Kiysu	dep	head	dep	?	head	Арман студент (Arman **is a** student)
Kiysu	dep	V	dep	head	?	Мен оқимын кітап (I read **the book**)
MengEt	dep	V	?	head	dep	Оқушылар мектепке келді (Students came to school)

Рис. : 1. Фрагмент правил преобразования слов в тексте в понятия и взаимосвязи базы знаний.