

УДК 004.81

Озера знаний для когнитивных систем: интеграция графовых баз и онтологий

Шендаров Артем Валерьевич^(*) shendarovav@student.bmstu.ru
Николаев Кирилл Сергеевич nikolaevkirill237@gmail.com

МГТУ им. Н.Э. Баумана, Москва, Россия

Аннотация. Современное когнитивное моделирование требует все большего количества разнотипных данных для работы и обучения, что вызывает трудности в их сортировке и получении доступа к ним. Озера знаний являются одним из вариантов решения данной проблемы.

Ключевые слова: когнитивное моделирование, озера знаний, OWL, базы данных, когнитивные архитектуры, графовые модели

Knowledge lakes for cognitive systems: integration of graph databases and ontologies

Shendarov Artem Valeryevich^(*) shendarovav@student.bmstu.ru
Nikolaev Kirill Sergeevich nikolaevkirill237@gmail.com

BMSTU, Moscow, Russia

Abstract. Modern cognitive modeling requires an increasing amount of diverse data for work and training, which makes it difficult to sort and access them. Knowledge lakes are one of the solutions to this problem.

Keywords: cognitive modeling, knowledge lakes, OWL, databases, cognitive architectures, graph models

Введение. Во время быстрого развития искусственного интеллекта когнитивные системы сталкиваются с вызовами, связанными с обработкой невероятно больших объемов разно типовых данных [1]. Эти данные могут включать информацию из баз данных, тексты из документов или веб-ресурсов, данные от устройств IoT или цифровые данные и формулы. Централизованное хранение знаний в форме озер знаний в наше время становится необходимостью, поскольку нынешние методы хранения уже не справляются с обработкой такого количества данных, что приводит к снижению эффективности обучения нейронных сетей. Озера знаний представляют собой универсальную платформу, которая объединяет все типы данных в едином пространстве, позволяющее когнитивным системам получать систематическую, структурированную информацию, включающую различные типы данных [2]. Данная тема особенно актуальна для задач, требующих междисциплинарного подхода, например, в медицине, где нужно интегрировать клинические данные с научными публи-

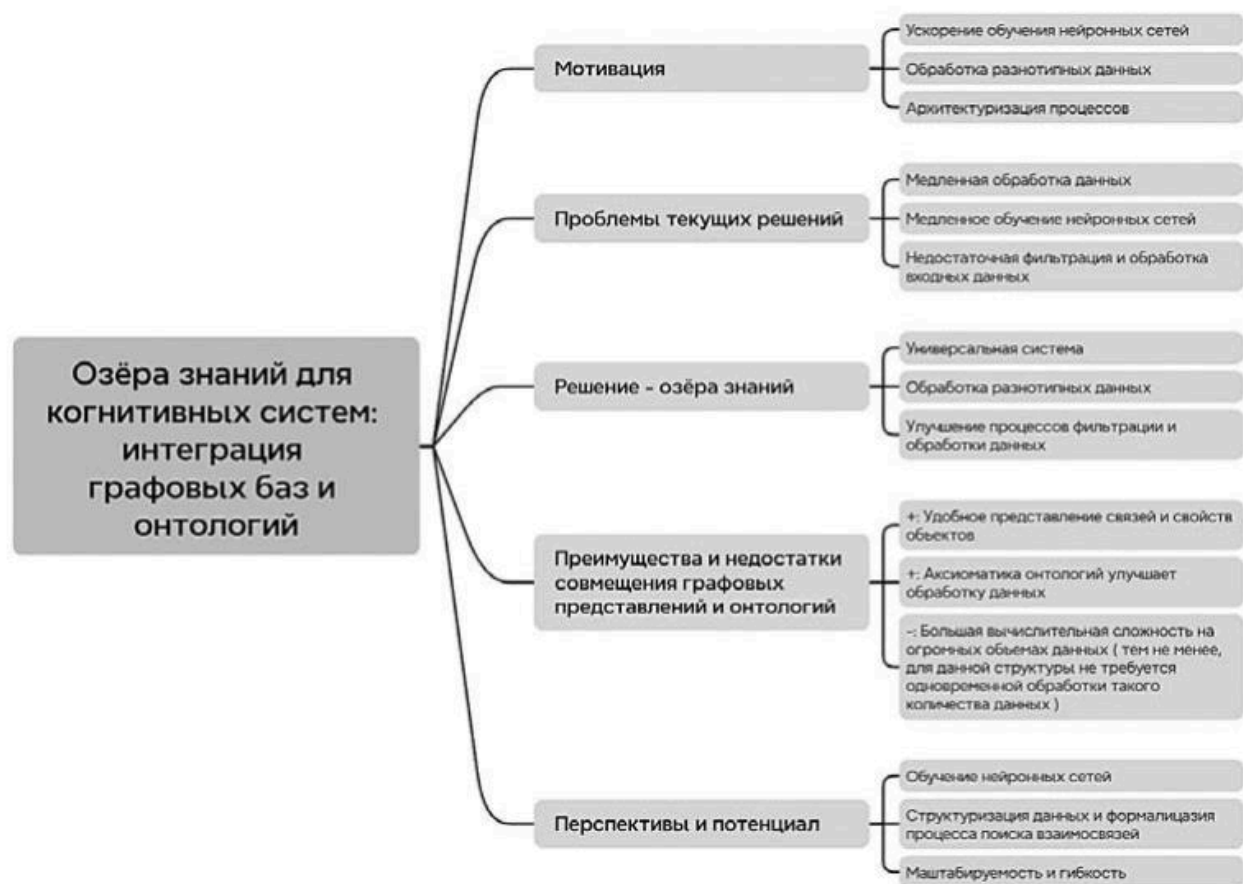
кациями, или в робототехнике, где сенсорные данные сочетаются с правилами поведения.

Материалы и методы. Анализ существующих подходов к хранению знаний показывает, что традиционные базы данных, основанные на SQL и NoSQL-системах [3], эффективны только для хранения больших объемов структурированных данных, но они ограничены в способности обрабатывать сложные разнотипные данные, что очень важно для когнитивных систем [4]. Например, в реляционных базах связи представляются через таблицы и ключи, но это не позволяет легко моделировать связи между ними, такие как иерархии или неявные ассоциации [5]. Онтологии на базе стандарта OWL (Web Ontology Language) предлагают более продвинутый подход, позволяющий формализовать знания с использованием классов, свойств и логических аксиом. Однако OWL-онтологии сталкиваются с проблемами высокой сложности: при работе с миллионами строк данных вычислительная сложность растет экспоненциально, что не позволяет в полной мере использовать их для работы с нейронными сетями. Графовые базы данных, включая Neo4j и RDF-ориентированные системы вроде Apache Jena, превосходят другие варианты в представлении связей: данные моделируются как графы с узлами и ребрами, что хорошо подходит для задач, связанных с сетевыми структурами, такими как социальные сети или биологические взаимодействия [6]. Тем не менее, эти базы часто нуждаются в унифицированных моделях для когнитивных задач, поскольку без семантического слоя они не могут автоматически воспринимать смысл связей [7].

Основная задача исследования — разработка хранилища знаний, которое будет совмещать преимущества графовых баз данных и онтологий для создания озер знаний. Это хранилище должно не только хранить данные, но и обеспечивать их обработку, чтобы когнитивные модели могли быстро извлекать систематизированную информацию для обучения и принятия решений [8]. Ключевые требования к системе включают наличие поиска, позволяющего находить данные и по смыслу, и по ключевым словам.

Результаты и обсуждения. Для решения поставленной задачи предполагается метод, основанный на интеграции графовых баз данных, таких как Neo4j, и онтологий OWL (см. рисунок). Графовые базы будут отвечать за хранение данных, где данные представляются в виде графа: узлы отвечают за сущности, а ребра — за отношения между ними, обеспечивая быстрый поиск по ним. Онтологии OWL дополняют это, определяя свойства и правила: например, аксиомы позволяют выводить новые связи, такие как классы в языках программирования. Схема взаимодействия между компонентами реализуется через маппинг: данные из графа экспортируются в OWL для логического анализа, а результаты возвращаются обратно для улучшения графа. Для семантического поиска и извлечения знаний используются комбинированные алгоритмы на базе языков запросов SPARQL, для онтологий, позволяющий обрабатывать сложные логические запросы и Cypher для графов, оптимизированный для работы с паттернами связей. Обновление знаний

в реальном времени достигается через специализированные API, которые позволяют подключать внешние источники, такие как веб-ресурсы или потоки данных от сенсоров объекта, с автоматическим парсингом и интеграцией: например, с помощью методов NLP (natural language processing) для извлечения данных из текстов. Дополнительно внедряются механизмы проверки данных для обеспечения точности, такие как проверка на противоречия в онтологиях, и оптимизации производительности, включая индексацию графов для ускорения обработки запросов.



Интеграция графовых баз и онтологий

Заключение. Идея озер знаний, основанная на совмещении графовых баз данных и онтологий OWL, является невероятно перспективным направлением для развития когнитивных систем. Предложенный подход предполагает создание централизованного хранилища знаний, которое объединяет различные виды данных и позволяющее их структурировать, обеспечивая их доступность для интеллектуальных систем. Теоретическая модель, сочетающая возможности графовых структур для представления сложных связей и онтологий для их интерпретации, это все открывает новые возможности для решения различных задач — от обработки текстов на различных языках до моделирования сложных систем, таких как медицинская диагностика или управление автономными устройствами. Потенциал озер знаний заключается в их способности поддер-

живать обучение ИИ, обеспечивая гибкость и масштабируемость при работе с разнотипными данными. Перспективы дальнейших исследований включают разработку теоретических основ масштабирования озер знаний для обработки больших данных, интеграцию с нейронными сетями, а также изучение методов автоматического обогащения знаний из открытых источников, таких как веб-ресурсы и социальные сети. Данная концепция закладывает основу для разработок, направленных на создание подобных озер знаний для работы с нейронными сетями.

Список источников

- [1] Ляпунцова Е.В. Когнитивная наука и искусственный интеллект. *Русский инженер. II Всероссий. конгресс с междунар. уч.: сб. тез.* Москва, Изд-во МГТУ им. Н.Э. Баумана, 2024, с. 282–284.
- [2] Цибизова Т.Ю., Панилов П.А. *Когнитивное моделирование и системный анализ высокоинтеллектуальных систем.* Москва, Изд-во МГТУ им. Н.Э. Баумана, 2025, 116 с.
- [3] Позняк А.А., Горячкин Б.С. Оценка времени выполнения SQL-запросов в озерах данных. *Наукосфера*, 2024, № 5–1, с. 60–70.
- [4] Гаврилова Т.А., Хорошевский В.Ф. *Базы знаний интеллектуальных систем.* Санкт-Петербург, Питер, 2000, 384 с.
- [5] Брешенков А.В., Гудзенко Д.Ю., Терехова Н.Ю. *Методы проектирования реляционных баз данных на основе информации табличного вида.* Москва, Изд-во МГОУ, 2012, 183 с.
- [6] Маслов В.П., Чеботарев П.Ю. Графовые модели в задачах управления знаниями. *Труды СПИИРАН*, 2015, № 41, с. 123–140.
- [7] Кудрявцев Д.В., Гаврилова Т.А. *Инженерия знаний: модели и методы.* Москва, ЛЕНАНД, 2018, 352 с.
- [8] Ветошкин А.А., Миронова А.Р., Зенгер А.С., Сухобоков А.А., Горячкин Б.С. Обработка запросов в системе управления озером данных, реализованном на основе универсальной модели данных. *Проблемы информационной безопасности. Компьютерные системы*, 2025, № 2 (64), с. 96–111.